



# **ICWMC 2014**

The Tenth International Conference on Wireless and Mobile Communications

ISBN: 978-1-61208-347-6

June 22 - 26, 2014

Seville, Spain

## **ICWMC 2014 Editors**

Dragana Krstic, University of Nis, Serbia

Mari Carmen Aguayo Torres, University of Malaga, Spain

# ICWMC 2014

## Foreword

The Tenth International Conference on Wireless and Mobile Communications (ICWMC 2014), held between June 22-26, 2014 - Seville, Spain, followed on the previous events on advanced wireless technologies, wireless networking, and wireless applications.

ICWMC 2014 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 2014 Technical Program Committee, as well as the numerous reviewers. The creation of such a 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 2014. 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 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICWMC 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of wireless and mobile communications.

We are convinced that the participants found the event useful and communications very open. We also hope the attendees enjoyed the charm of Seville, Spain.

### **ICWMC 2014 Chairs:**

#### **ICWMC Advisory Committee**

Dragana Krstic, University of Nis, Serbia

Magnus Jonsson, Halmstad University, Sweden

Mari Carmen Aguayo Torres, University of Malaga, Spain

Wolfgang Narzt, Johannes Kepler University Linz, Austria

Robert Bestak, CVUT in Prague, Czech Republic

Abdulrahman Yarali, Murray State University, USA

#### **ICWMC Industry/Research Chairs**

Christopher Nguyen, Intel Corp., USA

Mohamad Sayed Hassan, Orange Labs (Alten) / Paris - France

Wolfgang Aichmann, Nokia Siemens Networks, Germany

Ahmed Ibrahim, Intel Corporation, Egypt

Daniele Grasso, STMicroelectronics Srl, Italy

**ICWMC Publicity Chairs**

Siavash Rahimi, McGill University, Canada

Wu-Shiung Feng, Chang Gung University, Taiwan

Claudio Monteiro, IFTO, Brazil

Isaí Michel Lombera, University of California - Santa Barbara, USA

## **ICWMC 2014**

### **Committee**

#### **ICWMC Advisory Committee**

Dragana Krstic, University of Nis, Serbia  
Magnus Jonsson, Halmstad University, Sweden  
Mari Carmen Aguayo Torres, University of Malaga, Spain  
Wolfgang Narzt, Johannes Kepler University Linz, Austria  
Robert Bestak, CVUT in Prague, Czech Republic  
Abdulrahman Yarali, Murray State University, USA

#### **ICWMC Industry/Research Chairs**

Christopher Nguyen, Intel Corp., USA  
Mohamad Sayed Hassan, Orange Labs (Alten) / Paris - France  
Wolfgang Aichmann, Nokia Siemens Networks, Germany  
Ahmed Ibrahim, Intel Corporation, Egypt  
Daniele Grasso, STMicroelectronics Srl, Italy

#### **ICWMC Publicity Chairs**

Siavash Rahimi, McGill University, Canada  
Wu-Shiung Feng, Chang Gung University, Taiwan  
Claudio Monteiro, IFTO, Brazil  
Isaí Michel Lombera, University of California - Santa Barbara, USA

#### **ICWMC 2014 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  
Fumiyuki Adachi, Tohoku University, Japan  
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  
Alessandro Aldini, University of Urbino "Carlo Bo", Italy  
Erick Amador, Intel Mobile Communications, France  
Karine Amis, Institut Mines-Telecom/Telecom Bretagne, France

Lidiane Araújo, Federal University of Pernambuco, Brazil  
Jose Enrique Armendariz-Inigo, Universidad Publica de Navarra, Spain  
Radu Arsinte, Technical University of Cluj-Napoca, Romania  
Hakim Badis, University of Paris-Est Marne-la-Vallée, France  
Mohammad M. Banat, Jordan University of Science and Technology, Jordan  
Alessandro Bazzi, IEIIT-CNR, University of Bologna, Italy  
Norman C. Beaulieu, University of Alberta, Canada  
Carlos Becker Westphall, Federal University of Santa Catarina, Brazil  
Mouncef Benmimoune, Université du Québec, Canada  
Robert Bestak, Czech Technical University in Prague, Czech Republic  
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  
Vicente Casares-Giner, Universidad Politécnica de Valencia, Spain  
Pedro Castillejo Parrilla, Universidad Politécnica de Madrid, Spain  
Eddie Chan, Hong Kong University of Science and Technology, Hong Kong  
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  
Yung-Ting Chuang, National Chung Cheng University, Taiwan  
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  
Alban Duverdier, Centre National d'Etudes Spatiales(CNES), France  
Péter Ekler, Budapest University of Technology and Economics, Hungary  
Ghaïs El Zein, IETR - INSA Rennes, France  
Gianluigi Ferrari, Università di Parma, Italy  
Armando Ferro Vázquez, Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain  
Jane Louie Fresco Zamora, Nara Institute of Science and Technology, Japan  
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  
Ignacio González Alonso, University of Oviedo, Spain  
Javier Manuel Gozálvez Sempere, University Miguel Hernandez of Elche, Spain  
Tao Gu, RMIT University, Australia  
M. Amac Guvensan, Yildiz Technical University in Istanbul, Turkey  
Xiang Gui, Massey University, New Zealand  
Christian Hägerling, TU Dortmund University, Germany  
Gerhard Hancke, City University of Hong Kong, Hong Kong  
Mohamad Sayed Hassan, Orange Labs (Alten) / Paris, France  
Laurent Herault, CEA-Leti - Grenoble, France  
Unai Hernández-Jayo, Deusto Institute of Technology - DeustoTech., Spain  
Chih-Lin Hu, National Central University, Taiwan  
Yueh Min Huang, National Cheng Kung University, Taiwan  
Sandor Imre, Budapest University of Technology and Economics, Hungary  
Fumio Ishizaki, Nanzan University, Japan  
Muhammad Ismail, University of Waterloo, Canada  
Minoru Ito, Nara Institute of Science and Technology, Japan  
Yasunori Iwanami, Shikumi College Nagoya/Institute of Technology -Nagoya-shi, Japan  
Anurag Jain, HCL Technologies Ltd, India  
Tauseef Jamal, University Lusofona - Lisbon, Portugal  
Ali Jemali, 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 Abdrabou Kalil, Suez University, Egypt  
György Kálmán, ABB AS - Akershus, Norway  
Georgios Kambourakis, University of the Aegean, Greece  
Riad Kanan, University of Applied Sciences Western Switzerland (HES-SO), Switzerland  
Subrat Kar, Indian Institute of Technology Delhi - New Delhi, India  
Shigeru Kashihara, Nara Institute of Science and Technology, Japan  
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  
Ondrej Krejcar, University of Hradec Kralove, Czech Republic

Dragana Krstic, University of Nis, Serbia  
Alexey Lagunov, M.V. Lomonosov - The Northern (Arctic) Federal University, Russia  
Zhihua Lai, Ranplan Wireless Network Design Ltd., UK  
Alain Lambert, University of Paris Sud, France  
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, University of Erlangen-Nürnberg, Germany  
Valeria Loscrí, University of Calabria, Italy  
Jonathan Loo, Middlesex University - London, UK  
Christian Maciocco, Intel Corporation -Santa Clara, USA  
Christian Makaya, IBM Research, USA  
Pratyusa K. Manadhata, HP Labs, 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, QMIC - Qatar Mobility Innovations 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  
Augusto Morales, Technical University of Madrid, Spain  
Mohamed M. A. Moustafa, Egyptian Russian University, Egypt  
Lorenzo Mucchi, University of Florence, Italy  
Raja Kumar Murugesan, Taylor's University, Malaysia  
Nidal Nasser, Alfaisal University, Saudi Arabia  
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  
Christopher Nguyen, Intel Corp., USA  
Nhut Nguyen, University of Texas at Dallas, 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 Politécnic de Valencia, Spain  
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  
Jae-Young Pyun, Chosun University, Korea  
Ion Emilian Radoi, The University of Edinburgh, U.K.  
Hani Ragab Hassen, University of Greenwich, UK  
Yusnita Rahayu, Universiti Malaysia Pahang (UMP), Malaysia  
Muttukrishnan Rajarajan, City University London, UK  
Eric Renault, Télécom SudParis, France  
Christian Renner, University of Luebeck, Germany  
Teng Rui , NICT, Japan  
Brian M. Sadler, Army Research Laboratory, USA  
José Antonio Sánchez Fernández, Universidad Politécnic 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  
Mujdat Soy Turk, Marmara University, Turkey  
Kuo-Feng Ssu, National Cheng Kung University, Taiwan  
Razvan Stanica, INSA Lyon, France  
Álvaro Suárez Sarmiento, Universidad de Las Palmas de Gran Canari, Spain  
Yichuang Sun, University of Hertfordshire, UK  
Yasihisa Takizawa, Kansai University, Japan

Fatma Tansu Hocanin, Eastern Mediterranean University, Turkey  
Thomas Ußmüller, University of Erlangen-Nuremberg, Germany  
Václav Valenta, Ulm University, Germany  
Juha-Matti Vanhatupa, Tampere University of Technology, Finland  
Emmanouel (Manos) Varvarigos, University of Patras, Greece  
K. Vasudevan, Indian Institute of Technology - Kanpur, India  
Angeles Vazquez-Castro, Universitat Autònoma de Barcelona, Spain  
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  
Sabine Wittevrongel, Ghent University, Belgium  
Ouri Wolfson, University of Illinois at Chicago, USA  
Lawrence Wong, National University of Singapore, Singapore  
Qishi Wu, University of Memphis, USA  
Pei Xiao, University of Surrey, UK  
Mee Loong Yang, AUT University, New Zealand  
Muersel Yildiz, Technical University of Berlin - GT-ARC Labor, Germany  
Erkan Yüksel, Istanbul University, Turkey  
Sven Zacharias, University of Limerick, Ireland  
Sherali Zeadally, University of Kentucky, USA  
Yuanyuan Zeng, Wuhan University, China  
Hans-Jürgen Zepernick, Blekinge Institute of Technology, Sweden  
Yan Zhang, IMEC- NL, The Netherlands  
Wu Zhanji, Beijing University of Post and Telecommunications, China  
Jan Zizka, Mendel University in Brno, Czech Republic

## 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

Multi-agent Topology Approach for Distributed Monitoring in Wireless Sensor Networks <i>Haffaf Hafid, Bechar Rachid, and Congduc Pham</i>	1
Outage Performance and Derivation due to Adjacent Channel Interference <i>YoungKeun Yoon and YoungJun Chong</i>	10
Uplink Performance Evaluation of Broadband Systems which Adopt a Massive MU-MIMO Approach <i>Paulo Torres, Luis Charrua, and Antonio Gusmao</i>	14
Evaluative Study of Detect-Split-Forward Scheme over MIMO Relays <i>Suhaib Al-Basit, Samir Al-Ghadhban, and Salam Zummo</i>	20
On the SC/FDE Uplink Alternative to OFDM in a Massive MU-MIMO Context <i>Paulo Torres, Luis Charrua, and Antonio Gusmao</i>	26
Capacity Estimation of IEEE 802.15.4 Chains of Sensor Nodes <i>Evaggelos Chatzistavros and George Stamatelos</i>	32
Performance Enhancement of Heterogeneous Networks via Dense Clusterization and Higher Order Modulation <i>Ahmed Elhamy and Ahmed Ibrahim</i>	37
Using GSPNs for Performance Evaluation of Networks with Repeated Calls and Different Vacation Policies <i>Nawel Gharbi</i>	43
Level Crossing Rate of MRC Receiver Over $k$ - $\mu$ Multipath Fading Environment <i>Dragana Krstic, Ilija Temelkovski, Srdjan Maricic, Dragan Radenkovic, and Vladeta Milenkovic</i>	50
Statistical Analysis and Re-generation of the Radiated Disturbances from Multiple Noise Sources <i>Suna Choi and Jong-wha Kwon</i>	55
Modeling and Performance Evaluation of Scheduling Algorithms For Downlink LTE cellular Network <i>Bechir Nsiri, Nasreddine Mallouki, Mahmoud Ammar, Walid Hakimi, and Sofien Mhatli</i>	60
Channel Estimation for Downlink LTE System Based on LAGRANGE Polynomial Interpolation <i>Mallouki Nasreddine Mallouki, Nsiri Bechir Nsiri, Walid Hakimi Hakimi, and Mahmoud Ammar Ammar</i>	65
Reception for Layered STBC Architecture in WLAN Scenario <i>Piotr Remlein and Hubert Felcyn</i>	70
A Wideband Envelope Modulator Design for Envelope-Tracking SiGe Power Amplifier (ET-PA) for Broadband	76

<b>Wireless Applications</b> <i>Yan Li, Jerry Lopez, and Donald Y.C. Lie</i>	
Fully Functional Passive RFID Tag with Integrated Sensor for Item Level Tagging Based on Collective Communications and Organic Printed Electronics <i>Predrag Jakimovski, Ali Hadda, Michael Zangl, and Nina Oertel</i>	84
An End-to-End QoS Performance Evaluation of VoLTE in 4G E-UTRAN-based Wireless Networks <i>Myasar Tabany and Chris Guy</i>	90
Enhancing the Vector-Based Forwarding Routing Protocol for Underwater Wireless Sensor Networks: A Clustering Approach <i>Dina M. Ibrahim, Tarek E. Eltobely, Mahmoud M. Fahmy, and Elsayed A. Sallam</i>	98
On-Demand Data Collection in Sparse Underwater Acoustic Sensor Networks Using Mobile Elements <i>Jalaja M Janardanan Kartha and Lillykutty Jacob</i>	105
On Type II Hybrid-ARQ with Decode and Forward Relay using Non-Binary Rate-Compatible Punctured LDPC Code on MIMO SC-FDMA up-link <i>Tomotaka Hamada and Yasunori Iwanami</i>	112
Interference Management for D2D Communications Underlying Cellular Networks at Cell Edge <i>Bin Guo, Shaohui Sun, and Qiubin Gao</i>	118
Optimization in Backoff Technique for IEEE 802.15.4/ ZigBee Networks <i>Muneer Bani Yassein, Yaser Khamayseh, Maged Fakirah, and Ghassan Mohsen</i>	124
Circuit-Switched Voice over Scalable UMTS <i>Soumya Das, Yuheng Huang, Edwin Park, and Samir Soliman</i>	129
Non-frequency-selective I/Q Imbalance in Zero-IF Transceivers for Wide-Band mmW Links <i>Ainhoa Rezola, Juan Francisco Sevillano, Roc Berenguer, Martin Leyh, Moises Lorenzo, Aharon Vargas, and Igone Velez</i>	136
A Distributed Algorithm for In-Network Adaptive Estimation Using Incremental Aggregated Gradient <i>Wael Bazzi, Amir Rastegarnia, Azam Khalili, and Saeid Sanei</i>	142
Channel Inversion CoMP Technique in Cellular System A User-Selection Algorithm <i>Jesus Gomez Galvan, Mari Carmen Aguayo Torres, Francisco Blaquez Casado, and Francisco Javier Martin Vega</i>	147
Enhancement of A5/1 Stream Cipher Overcoming its Weakneses <i>Mahdi Madani and Salim Chitroub</i>	154

Enhancing Safety-Critical Message Dissemination in WAVE <i>Daniel Neville, James Brill, Joseph Juzl, Christopher Howell, and Victor Sanchez</i>	160
Investigating Bit Error Patterns for Radar Pulse Detection in IEEE 802.11 <i>Claudio Pisa, Andres Garcia Saavedra, and Douglas Leith</i>	164
Cross-Platform End-to-End Encryption of Contact Data for Mobile Platforms using the Example of Android <i>Markus Hofmarcher, Michael Strauss, and Wolfgang Narzt</i>	168
Problematic of QoS-based On-demand Routing Protocols to Improve Communication in Mobility Context <i>Tiguiane Yelemou and Tounde Mesmin Dandjinou</i>	174
A Channel-Aware Uplink Scheduling Algorithm for Mobile WiMAX Networks <i>Nada ElShennawy, Mahmoud Fahmy, Mohammed Elderini, and Moustafa Yousef</i>	180
Novel Pico-cell Range Expansion with Adaptive RACH Resource Allocation for Random Access of M2M Devices <i>Woo-Jong Jo, Chang-Yeong Oh, Yunmin Kim, and Tae-Jin Lee</i>	186
Insights into Cellular Networks Anatomy of Traffic Profiles <i>Wieslawa Wajda</i>	192
Classifying Users Based on Their Mobility Behaviour in LTE Networks <i>Bart Sas, Kathleen Spaey, and Chris Blondia</i>	198

# Multi-agent Topology Approach for Distributed Monitoring in Wireless Sensor Networks

Bechar Rachid, Haffaf Hafid

Department of Computer Science, University of Oran  
Oran, Algeria

E-mail: rachid\_bec@yahoo.fr, haffaf.hafid@univ-oran.dz

Congduc Pham

University of Pau, LIUPPA Laboratory  
Pau, France

E-mail: congduc.pham@univ-pau.fr

**Abstract** — In this paper, the multi-agent technology is applied in wireless sensor networks domain in order to adapt software architecture, and to optimize its performance in monitoring. We explore in particular the issues of topology control, especially some related work using multi-agent systems. In the third part, we will propose an agent-based algorithm for fault tolerance and topology control in a wireless sensor network. Our proposal consists of embedding an agent at each node that is responsible for selecting its parent or the next hop to the sink when transferring packets. The main contribution is the proposal of a new process of changing parent, which is based on the computation of a fault tolerance degree, calculated each time by the agent in cooperation with its neighboring nodes. Several parameters are exploited to calculate this metric, such as the number of hops, the energy and the quality of links. Simulation results show that this method of changing parent allows an enhanced lifetime, as well as network fault tolerance, when compared with the collection tree protocol.

**Key words:** *wireless sensor networks; multi-agent systems; monitoring; topology control.*

## I. INTRODUCTION

Wireless sensor networks (WSN) [1] require large amount of data to be transmitted with high reporting rates, leading to consume specific resources, such as bandwidth, storage, computation, and energy. Research in WSNs aims to meet the above constraints by introducing new design concepts, creating, improving existing protocols such that optimization in this field has been a topical issue of many works in the last decade.

Multi-agent systems (MAS) have a principle that can be easily adapted and integrated in complex systems due to their fully decentralized and “intelligent” approach [2]. They can be used to model phenomena where global behavior emerges from the local behavior of system entities and components. These components have the ability to percept, process, act and react in their environment.

The wireless sensor networks structure, distributed processing ability and complexity considerations especially when the number of nodes increases, conducts us to exploit recent developments made in multi-agent systems field to improve networks performances and simplify the design process in order to have reliable and fault tolerant sensors.

The multi-agent approaches for WSN are introduced in many levels and operating aspects. Some works propose software architectures for applications and services [3][4]. Others are interested in network organization and cooperation between nodes [5][6], clustering being the dominant approach in this area. Furthermore, many agent-based works treat routing problems in WSN with different applications [7]. Finally, monitoring and mobility are also subjects of several studies [7][8], where the use of bio-inspired principles seems interesting with multi-agent systems to solve mobility and scheduling tasks problems.

These works can be divided into two classes: the first class considers the sensor network as a multi-agent system, in this case, the application of agent technology consists of deploying the same agent to all nodes in order to have cooperation between them. This seems to be better adapted to wireless sensor nodes, and it will be considered in our approach. The second class provides multi-agent systems adapted to the logical structure of a WSN where different kinds of agent cooperate in the network.

Supervision or monitoring, which aims to avoid or detect failures, is a set of techniques used to increase the performance of a WSN, optimize its lifetime and ensure fault tolerance using all network parameters.

The main approach for load balancing [9] and fault tolerance in sensor networks is the maintenance of a topology that guarantees good conditions of transfer. Indeed, the use of multiple paths helps to balance the energy consumption of nodes by distributing the flow of packets on better possible paths [10]. This mechanism based on topology control contributes to reducing the delay and packet loss by reducing the number of hops between nodes and the sink. Thus, the quality of service will be improved.

Topology control consists of the use of network characteristics or parameters to generate and/or maintain a topology. Despite the fact that MAS technology is the only approach that allows nodes to take into account information from their environment, only few works on WSNs topology control include agent systems.

Our work targets to increase the lifetime of WSN nodes by introducing a new topology control approach and then compare it to the collection tree protocol (CTP). In our work, the main contribution, which is introducing MAS in WSN, is performed when using neighborhood information to give a metric called tolerance degree for each node. This

metric is calculated by the node itself in cooperation with its neighbors. When transferring packets, this degree will be used to choose qualified nodes for the transfer. We shall see that this approach increases the fault tolerance lifetime of a network, without affecting the quality of service. This method is implemented as an embedded multi-agent system.

Section II is devoted to a classification of the state of the art that uses multi-agent systems in wireless sensor networks. In Section III, a new proposal is introduced on distributed agent based topology control in WSN. Performances of this proposal will be studied in Section IV, and finally, we will give some concluding remarks and future work in the last section.

## II. USING MAS IN WSN

Using agent technology in a WSN consists in associating agents to nodes which cooperate between them in order to calculate topology parameters. Each agent has a proper code that can be executed by a node, with the ability for the agent to move between nodes for processing or searching information. According to Rijubrata [11], the multi-agent approach in WSN has many advantages: the easy network scalability, extensibility and adaptation tasks, energy efficiency, and progressive system.

In recent years, several research works are interested in the WSN distributed processing based on agent technology. A classification of these contributions is given below, according to their objectives and the level of integration in a WSN.

### A. Software architecture of applications and services

Biswas et al. [12] presents an interoperable multi-agent architecture through layers. The authors demonstrate the effectiveness of their method by comparing the client/server approach and the multi-agent systems in terms of execution time and energy consumption. The work is an extension of an existing model called interoperable agent model: the new MAS model includes eight agent types: Sensor Agent (SA), agent management system (AMS), Directory Facilitator (DF), which helps other agents to cooperation, agent communication channel (ACC), Controller synthesizer (CS), Data Manager (DM), Application Agent (AA) and Agent Interface (AI). These agents are distributed in a sensor network that may contain both wired and IP nodes. The communication between agents is provided by XML messages.

After their work on the MWAC model (Multi-Wireless-Agent Communication) and DIAMOND method [3], which is interested in embedded multi-agent systems, especially in wireless infrastructure, the authors present in [2] a discussion on embedded systems design specifics that use multi-agent systems.

Smarsly et al. [4] proposes a system design-based on migrant agents to define a dynamic operation in a WSN according to nodes requirements, this system was really

tested on a platform for thermal variations treatment in an experimental environment.

Rahal et al. [13] propose a formal model based on real-time temporal logic for multi-agent system specification and evaluation when it is integrated into a WSN environment. Reactive decisional agents are used to take advantage of their ability to cooperation, reaction to events, communication and concurrence.

### B. Organization of the network, clustering and cooperation

In order to facilitate the design and implementation of WSN, Wang et al. [5] propose a model as combination between mobile agents and MAS. The proposed architecture is hierarchically structured according to the roles played by each sensor in the network. These roles are assigned to the nodes using some elective algorithms, the MAS is used for collaboration and mobile agents for data exchange purpose. The objective is localization and classification of acoustic targets.

In [6], the optimization of energy consumption and reply time is based on MAS applied to a data collection algorithm used for monitoring emergent events where the WSN is divided into dynamic clusters. This is defined by the event importance which determines the size and the lifetime of a cluster. Mobile agents traverse the network through cluster heads; itinerary planning is determined by the residual energy and the packet loss degree in the path. Simulation results show that the multi-agent model has better performance in terms of energy consumption and replay time.

Logical clustering model adapted to multi-agent operation is presented by Jabeur et al.[7]. It divides the nodes into a four levels hierarchy: atomic level which is the node itself, micro level that represents a group of nodes managed by a cluster head, meso level which is the upper level grouping a number of clusters of the same area, and finally, a virtual cluster representing the entire network. This logical structure can change after an event. At each level, is assigned a type of agent cooperating with the higher level agent to accomplish the distributed network operations.

### C. Routing

Many works study the routing problem in WSN, the classical problem consists in routing data from source node to a destination node (sink). According to the application goal, multipath routing may be used in order to increase the reliability of data transmission i.e., fault tolerance. Liu et al.

[8] propose a new agent-based routing algorithm with quality of service in WSN. By participating in routing and maintenance of paths, agents are used to manage the topology changes and communication flow. The method is based on a Swarm Intelligence principle [33], which is inspired from the collective intelligence system of insects. In this case, latency, packet loss, and energy conservation are considered in general as quality of service factors.

The MAS has two agent types: Forward agent FA (to establish a connection with a neighbor in searching path) and Reverse agent RA (in response to build a path). As in [8], Dario et al. [14] propose a model called MAM (Markovian Agent Model) which is based on Swarm intelligence, but using a Markov model. The operation of agents is based not only on local transitions from a node itself, but the probable transitions of other nodes too (local transitions and induced transitions).

**D. Monitoring and Mobility**

A Bayesian model named BNGRAZ (Bayesian network algorithm grazing) is proposed by Matthew et al. [15] for managing mobility in WSNs. It is bio-inspired model that emulates the behavior of herbivores grazing pastures. The WSN in question contains some mobile nodes to adjust coverage and connectivity. The choice of itinerary taken by a mobile node is based on the probability of disconnection or inaccessibility, this probability is calculated using information provided by neighbors.

As in [15], Saamaja et al. [16] propose a similar principle but with the aim of optimizing the lifetime and satisfy requested quality of service using data collectors that form clusters by changing position, the movements are made according to objective rather than probability. A self-adaptation strategy for scheduling tasks in a WSN is presented in [17], where a mathematical model is proposed for dynamic allocation tasks. The algorithm has a collective intelligence functioning called PSO (Particle Swarm Optimization algorithm).

In [3], Jamont proposes that each node plays a specific role in its neighborhood. This role is determined by an embedded agent in the node itself. If the node is in the area of intersection of multiple clusters, it has the role of liaison or gateway; else it has the simple data capture role. Finally a representative node or a cluster head is elected in the cluster to manage communications. Mobility and node failures are well treated by this structure.

To optimize the task scheduling problem and data transmission in video WSN, the work presented by Huang et al. [18] is based on a set of intelligent procedures associated with agents by using ant colony algorithms, genetic algorithms, or mixed algorithms. Security problem is also pointed out by this recent trend through bio-inspired methods.

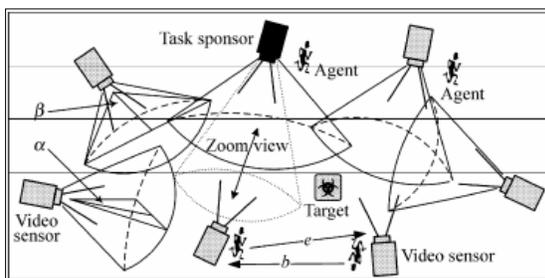


Figure 1. MAS for tasks allocation [18].

Each agent (representing a node) decides to participate or not in target detection according to the following factors:

- S: denotes the status set {busy, free} of the current node;
- E: depicts whether or not the current node has enough energy to accomplish the assignment;
- $\alpha$ : angle of vision of the camera on the x-axis (right to left);
- $\beta$ : angle of vision on the y-axis (up and down);
- q: determines the required quality of picture by the monitoring process.

The results show that the algorithms require less energy than AODV protocol.

**E. Topology control**

Topology control consists of using different parameters of the network in order to provide a well organization achieving some important tasks. These parameters could be radio range, state or role of the node, etc. The majority of works that use multi-agent systems in topology control are based on hierarchical structures with clusters, more adapted for MAS running on multiple levels [6]. First of all, we recall the principles and techniques used in this field and related work. According to [19], there exist three main techniques:

**E.1. Power Adjustment Approach**

The power adjustment approach allows nodes to vary their transmission power in order to reduce energy incurred in transmission. Rather than transmitting at maximum transmission power, nodes collaborate to adjust and find the appropriate transmission power, yielding to a connected network. For example, in Figure 2, the links N1 - N4 and N2 - N4 are unused by reducing the radio range of these nodes.

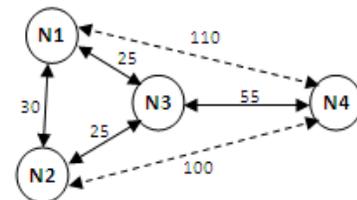


Figure 2. Topology control by adjusting the radio range.

Protocols representing this technique are Minimum Energy Communication Network (MECN) [34] where each node uses the minimum power level to communicate, and COMPOW [20] which uses a common minimum power level for all nodes in order guarantee the connectivity of the network.

**E.2. Power Mode Approach**

In addition to the techniques used by the MAC layer protocols and when the number of deployed nodes is sufficient, redundancy of nodes can be exploited to get a better topology by changing the state of a node between active and sleep. GAF (Geographical Adaptive Fidelity)

[21] is an example of a protocol that uses a principle of equivalent nodes, it allows a node to switch between sleep, discovery and active states as shown in Figure 3.

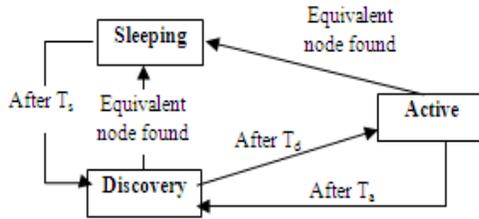


Figure 3. State Transitions in GAF.

Also, in ASCENT (Adaptive Self-Configuring Sensor Network Topologies) [22], a self-reconfigurable algorithm that allows nodes to locally measure the operating conditions is presented. Based on these conditions, nodes then decide whether they need to participate in routing or not.

### E.3. Hierarchical structures

This technique is to find a structure with hierarchical clusters for the network. Choosing cluster heads presents a problem for this technique. Several contributions have tried to propose approaches to obtain more efficient clustering. Most algorithms construct a virtual backbone based on the connected dominating set concept (CDS). From these algorithms, we can talk about PACDS (Power Aware Connected Dominating Set) [23], ECDS (Energy Efficient Distributed Connecting Dominating Sets) [35], and TMPO (Topology Management by Priority Ordering) [24].

### E.4. Hybrid approaches

The hybrid approach for topology control uses in general a combination between a clustering method and other techniques, like CLUSTERPOW algorithm [25] (Figure 4).

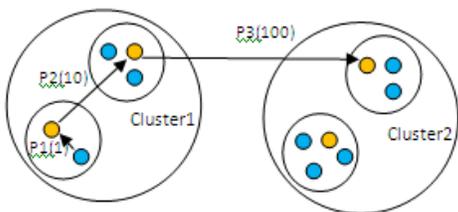


Figure 4. Principle of the CLUSTERPOW algorithm.

This algorithm defines several levels of clusters with different radio powers for communication within and between clusters.

## III. PROPOSED METHOD

### A. The problem

Our contribution consists of proposing a hybrid and distributed method using MAS for wireless sensor network topology control. This method bears on local decisions taken by the node itself using a function of several parameters: residual energy, number of neighbors, links

quality, etc. The main objective is to have at any time a connected, homogeneous and fault-tolerant network which should be capable to predict and avoid as much failures as possible. We have been inspired by influence systems [36] which require strong cooperation between nodes.

When transmitting data, the principle consists of selecting the most fault tolerant nodes that ensure safe transfer. The use of MAS seems to be a suitable approach according to the distributed, cooperative and emergent principles that characterizes this operation.

The role of MAS here is to calculate for each node a parameter determining its state and its capacity to go further without energy depletion or congestion failure before the end of data transfer. The calculated parameter is called the degree of tolerance of a node.

Some works in this context have a similar principle which is based on computing one or more metrics to control the topology such as Rong-rong et al. [26] which calculates the probability of node's failure, and Bo-Chao et al. [27] based on the evaluation of the link quality between two nodes to predict the lifetime of each node. So the main difference between these works is the choice of network parameters and how to calculate these parameters. In our approach, we propose another method, where we will use the link quality evaluation of [28] then we add the battery status and the number of hops to the. This principle allows us to express the lifetime and the fault tolerance ability. A thresholding mechanism is implemented to avoid frequent changes in the topology due to minor differences.

### B. Related work

We are interested here in the works which are based on local settings of network to predict or estimate other values or states in order to optimize the process of topology control.

We start with Yin et al. [26] which proposes an adaptive method for fault tolerance topology control by calculating the node failure probability FP based on the ratio of the consumed energy  $E_c$ , the initial energy  $E_{init}$  and another fault probability P associated to hardware and software components. We have:

$$PD = P \cdot \frac{E_c}{E_{init}} \quad (1)$$

In [29], Dario Bruneo et al. show that the introduction of Markov techniques allows estimating the lifetime of a node by taking the active-sleep cycle as a model of transitions with probabilities for each transition. But in reality, lifetime also depends on the node activities when it is at active state.

Failures in a sensor network can be detected by application of "fuzzy inference" according to Safdar Abbas Khan et al. [30] where the sensor measures are compared with expected values by a neural network; the differences in behavior allow detecting anomalies.

The lifetime of a hierarchical network is studied by Bo-Chao et al. in [27]. It proposes an algorithm for lifetime prediction in better and worst cases in a WSN with one hop

clusters where the cluster heads should communicate directly with the sink. The objective of this work is to find the best deployment for this type of topology. A cooperative approach for topology control is proposed by Paolo Costa et al. in [31]. The construction of the topology consists of choosing nodes that guarantee a degree of k-connectivity by using a minimum radio range. The stability of topology is obtained by cooperation between nodes until obtaining the optimal radio level under k-connectivity constraint.

### C. Proposal details

#### C.1. The network model

Related to an application domain, the nodes of network are deployed randomly in a known field to capture specific types of information; the captured data are then transmitted to a control station or sink. When transferring data, the choice of path is based on the choice of the next hop or the parent from the current node obeying to some routing protocol. Here, the next hop is selected from the neighbors using the degree of tolerance at the time of transfer, so the node that has the highest degree of tolerance will be qualified for this transfer. This task is performed by agents implanted on nodes.

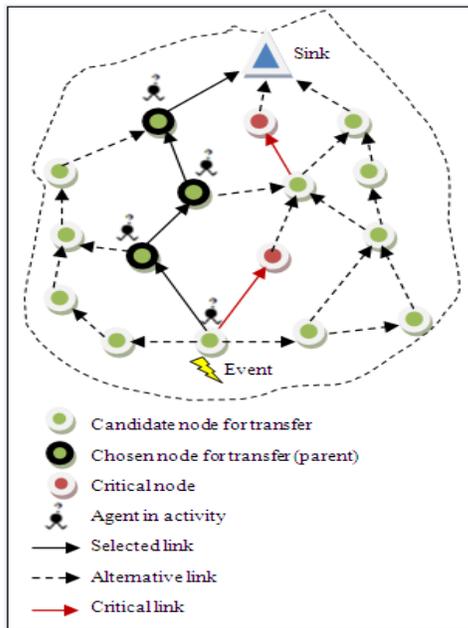


Figure 5. Network model.

Figure 5 shows this principle where the agent is responsible for selecting a parent among candidate neighbors for the node that wants to transmit packets. Some nodes may become critical, so they cannot be parents to other nodes.

#### C.2. The topology construction

After deployment of nodes, the sink diffuses an initialization message *Init* which is based on the HC (Hop

Count) value. For the sink itself, the value is null. The neighborhood discovery is included in this method.

Each node  $n$  which receives the *Init* message considers the sender of the message as the next hop for the next transmissions if the HC value of the latter, is less than the HC value of the receiver node  $n$ . So, it does:

$$\text{if } HC(n) > HC(\text{init}) \text{ then } HC(n) \leftarrow HC(\text{init})$$

Then, it rebroadcasts the *Init* message. At the beginning, the HC values are set to infinity for all nodes except the sink which is initialized to zero.

#### C.3. Topology control method

After the stability of the topology obtained by the initialization process (there is no node which rebroadcasts the *Init* message), the nodes calculate their degree of tolerance as follows. Let us consider:

- $E_{\text{init}}$ : initial energy of a node.
- $E_v$ : residual energy of the node  $v$ .
- $N$ : set of neighbors of  $v$ ,  $N_i \in N$  is a neighbor of  $v$ .
- $\text{Pin}(N_i)$ : number of received packets by  $v$  from  $N_i$  during a period  $t$ .
- $\text{Pout}(N_i)$ : number of broadcast packets by  $N_i$  during a period  $t$ .
- $\text{Poutc}(N_i)$ : number of correctly received packets by the neighbors of  $N_i$  (with acknowledgment).
- $\text{HC}(v)$ : hop count from  $v$  to the sink;
- $\text{NH}(v)$ : the next hop from  $v$  to the sink also said parent of  $v$ .

For the calculation of tolerance degree TD, we propose to use the link quality and battery status of each node. The calculation of the first parameter is inspired from the link quality estimation proposed by Omprakash et al. [28], where CTP (Collection Tree Protocol) is defined. It is a routing protocol that computes unicast routes to a single route or a small number of designated sinks in a wireless sensor network basing only on the link quality estimation network parameter. It uses periodic messages called beacons to maintain topology. A beacon is a packet that contains the link quality estimation between two nodes.

In our case, for battery status, we consider the relationship between the residual energy and the initial energy.

We define the quality of outgoing links QS between node  $v$  and its neighbor  $N_i$  as follows:

$$QS_v(N_i) = \text{Poutc}(v)/\text{Pout}(v) \quad (2)$$

Similarly, the quality of incoming links QE between a node  $v$  and its neighbor  $N_i$ :

$$QE_v(N_i) = \text{Pin}(v)/\text{Pout}(N_i) \quad (3)$$

The node  $v$  calculates its TD (Tolerance Degree) in function of its battery status and quality  $Q$  of an outgoing or incoming link like it is shown in equation (4), where  $Q$  is QS for outgoing links and  $Q$  is QE for incoming links as follows:

$$TD(v) = \left( \frac{E_v}{E_{\text{init}}} \right) \cdot (\alpha_{TD} \cdot Q + (1 - \alpha_{TD}) \cdot Q) \cdot TD_{\text{old}} \quad (4)$$

where  $\alpha_{TD}$  is a weighting constant that can take values between 0 and 1, it is 0.9 for our case.

The parent change procedure is based on the parameter values of the previous parent and candidate neighbors.

We use mainly the energy  $E$ , the degree of tolerance  $TD$  (old value) and the hops count  $HC$ . By cooperation, nodes use a control message to inform neighbors when there is change in values of energy,  $HC$  and  $TD$ . To avoid parent change when small variations in these parameters happen, a threshold principle is used for each one. So the node  $v$  decides to choose a neighbor  $N_i$  as its new parent  $NH$  if the following conditions are satisfied:

$$\begin{aligned} HC(N_i) &\leq HC(v) + \text{Threshold\_HC} \\ E(N_i) &> E(NH(v)) + \text{Threshold\_E} \\ TD(N_i) &> TD(NH(v)) + \text{Threshold\_TD} \end{aligned}$$

It is necessary to know that the verification order of these conditions is very important; it also depends on the nature of application using our approach. Here, we chose the hops count in first with a threshold that depends on the network size (number of nodes in the network) in order to avoid long traffic paths. This may be not useful if the application has no real time constraint. The algorithm below shows the parent selection process according to the previous conditions.  $T_E$ ,  $T_{TD}$  and  $T_{HC}$  represent the threshold values for the energy, degree of tolerance and the hops count respectively. The function  $\text{Change\_Parent}$  represents a switching tool between the active parent and the candidate one. The  $TD$  function implements the estimation of the tolerance degree.

#### Algorithm : Parent Selection

```

1: NH = Current Parent (Next Hop)
2: Ni = Candidate neighbor
3: If HC(Ni) < HC(NH) + THC then
4: if Energy(NH) = Energy(Ni) then
5: if TD(NH) > TD(Ni) + TTD then
6: Change_Parent(Ni)
7: end if
8: else if Energy(NH) < Energy(Ni) then
9: if TD(NH) ≥ TD(Ni) + TTD then
10: Change_Parent(Ni)
11: end if
12: else if Energy(NH) > Energy(Ni) and
    Energy(NH) ≤ Energy(Ni) + TE then
13: if TD(NH) ≥ TD(Ni) then
14: Change_Parent(Ni)
15: end if
16: end if

```

This algorithm allows the node to change its parent as soon as finding a better one. Also, it is used for selecting a new parent among candidate neighbors when detecting a fault (e.g., the current parent dies or moves). A fault of parent can be declared after failing in packets transfer.

## IV. IMPLEMENTATION AND RESULTS

### A. Simulation

Actually, the MAS platforms are not used in practice to implement WSN simulation because there are no network properties integrated in these software tools.

In order to validate the proposed solution and study its performances, as well as the adaptation of a multi-agent model for this type of distributed algorithm, we have implemented our approach using the Castalia simulator which is based on the simulator Omnet++ [32]. CASTALIA is a commonly used tool in recent years for WSNs simulation due to its gratuity and easy integration of new protocols in its software layer structure presenting a configurable environment as needed. For our study, we used version 3.2 of Castalia with OMNET++ 4.2 turning on an UBUNTU machine. Our distributed algorithm consists of implementing agent on each sensor node. It is the simple manner to view a WSN as MAS where cooperation is provided by exchanging messages at the moment of data transfer to select at each hop, the most fault tolerant node. It is clear that the principle of our algorithm implies that agents must communicate the necessary information like the latest values of  $TD$ , energy level and the number of hops. The diagram in Figure 6 shows that the agent on node 3, for example, has the choice to transfer its packets through nodes 2, 6 or 7 depending on the status of each one of these nodes.

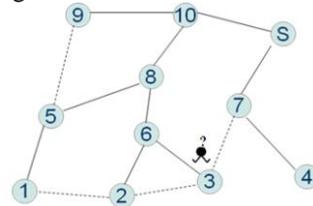


Figure 6. Principle of our algorithm.

### B. Hypothesis

To perform simulations, we consider the following hypothesis: initially, each node has an initial energy. All sensor nodes are battery powered with limited energy, except the Sink. Network size is specified at the beginning of each simulation, node 0 is chosen as Sink. Other simulation parameters are shown in the following table:

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Number of nodes	Up to 300
Field Deployment	250 x 250 meters
Deployment type	Random
Radio model	CC2420
Radio power	0dBm
Initial energy	18720 joules
Simulation time	100, 200, 300, 1000 sec ..
Thresholds: $T_{HC}$ , $T_E$ and	20% of difference for each

C. Results and analysis

To demonstrate the performance of our approach, we make a comparison with CTP protocol. On figures, our approach is noted ATC for agent topology control. Figure 7 illustrates the execution of our protocol and shows that traffic is distributed across all nodes. This is assured by changing the parents according to the parameters of each node. Applying the algorithm gives the following:

- (A) Construction of the initial tree
- (B) Node 7 becomes a parent of 3
- (C) Node 3 becomes a parent of 2
- (D) Node 2 becomes a parent of 3 and 1 becomes a parent of node 2
- (E) Node 9 becomes a parent of node 5

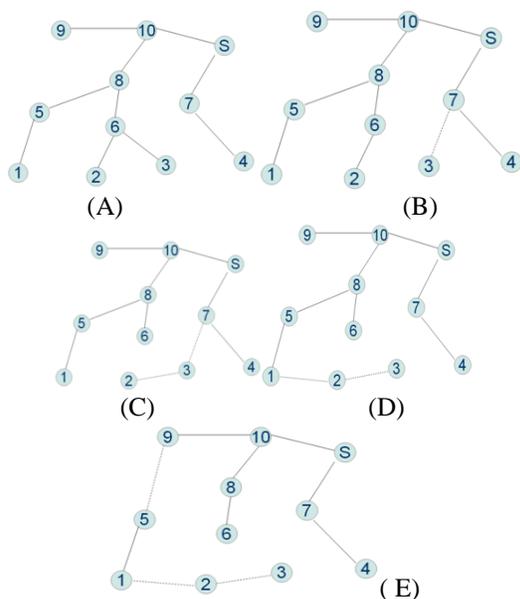


Figure 7. Illustration.

Before giving some results on the performances of our approach compared with CTP method, for the lifetime definition, we consider the duration between the network initialization time until the moment it becomes non convex or disconnected.

The initialization begins by broadcasting the 'init' message, and then nodes start changing their parents basing on hop count metric until the stabilization of the topology. Time of this operation depends on density of network which is defined by its size and the transmission ray of nodes. We can express this density by using the average number of node's neighbors in the network like it is shown in Figure 8.

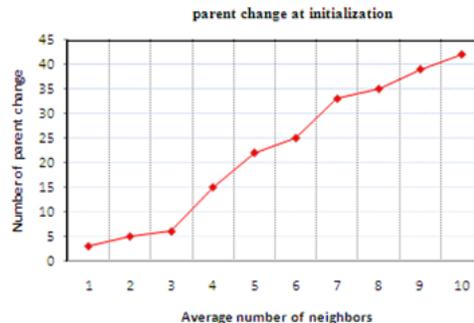


Figure 8. Relationship between WSN density and initialization process.

A proportional relationship is remarked between the average number of neighbors in the entire network and the average number of parent change at initialization. Having a lot of neighbors implies frequent recursive parent changing.

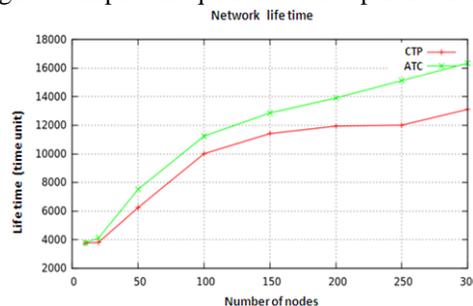


Figure 9. Impact on lifetime.

The curves in Figure 9 show a divergence in case of high density networks because our approach tries to find other paths to conserve energy of those used. However, in CTP, congestion or over-use of a path leads to the premature death of nodes. In case of low density, there is not a big difference because the topology is almost fixed with a small number of nodes.

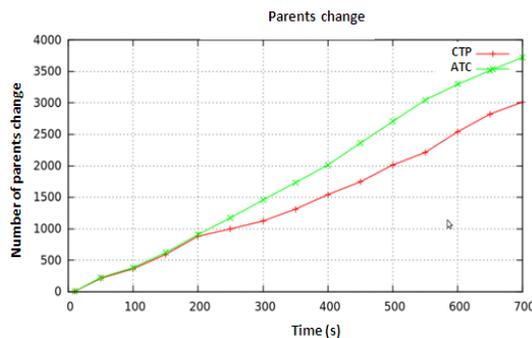


Figure 10. Parent change in the time.

We note from Figure 10 that the difference in the number of parents change in both cases CTP and ATC are not important at first, but, over time, a divergence becomes more significant. This is explained by the diminution of the degree of tolerance calculated by our protocol, which requires change of parents in order to ensure load balancing.

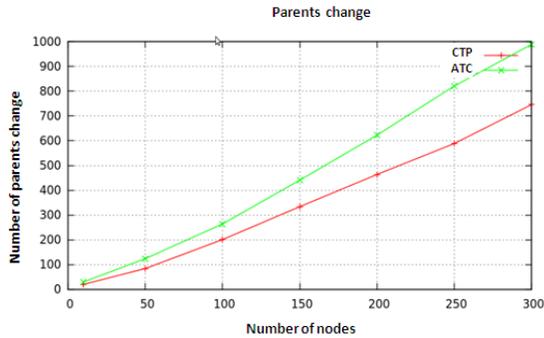


Figure 11. Change of parents with network size.

Figure 11 also shows the curves of parent change, but this time based on the network size. By increasing the size, we observe a difference in the number of parent changes for both protocols CTP and ATC. It is higher for large-scale networks that ensure the existence of other paths where the best one will be selected. The use of several parameters by our protocol gives more opportunities than CTP which uses only the quality of links.

## V. CONCLUSION AND FUTURE WORK

In this paper, we have presented a state of the art concerning the use of MAS in wireless sensor networks. In this context, we have proposed an agent-based topology control method for WSNs. According to the state of node and its neighbors, the main objective is to have a fault-tolerant network with an extended lifetime by optimizing the choice of paths from the nodes to the sink. This choice is based on the changing parent method which uses the concept of tolerance degree. Also, to respect the multi-agent principles, the choice of parent node is achieved in a cooperative and distributed manner. The design and implementation using Omnet++/Castalia of our proposal shows the relevance of multi-agent systems approach compared to CTP method.

Indeed, the simulation results show that our solution allows assessing at any time the fault tolerance level of each node leading to a better path selection process, and therefore, a longer lifetime of nodes. Limitations of the proposed approach are related to the reliability of link quality estimation mechanism which is best effort delivery. So, it is well adapted for relatively low traffic rate applications.

The performance of our approach can be enhanced by providing a formula that uses other network parameters to calculate the tolerance degree or even use a probabilistic approach to predict the activities of a sensor node such as packet traffic. In terms of implementation, it is interesting to use a real agent based platform for WSN in order to study performances of MAS and WSN when coupled. Other perspective works could concern to embed security parameters and develop more complex formulas.

## REFERENCES

- [1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey" *Computer networks*, vol. 38, no. 4, 2002, pp. 393–422.
- [2] J.P. Jamont and O. Michel, "Design of embedded multi agent systems: discussion about some specificities", VII Agent-Oriented Software Engineering Technical Forum, Paris, French, December 15, 2010.
- [3] J.P. Jamont, "MWAC: Multi-Wireless-Agent Communication", *Revue Technique et science informatiques RSTI série TSI*, vol. 25, no. 5, 2006, pp. 661–688.
- [4] K. Smarsly, K.H. Law, and M. König, "Resource-Efficient Wireless Monitoring based on Mobile Agent Migration", In proceedings of the SPIE: Health Monitoring of Structural and Biological Systems 2011. San Diego, CA, USA, 06 Mar. 2011.
- [5] W. Xue, B.Dao D. Liang, and W. Sheng, "Collaborative Target Classification and Localization in Wireless Sensor Networks Agent". *Journal of Sensors*, vol. 7, no. 8, 2007, pp. 1359–1386.
- [6] Y. Lingyun, W. Xingchao, G. Jianhou, and Z. Yanfang, "Data Gathering Algorithm Based on Mobile Agent and Emergent Event Driven in Cluster-Based WSN", *Journal of Networks*, vol. 5, no. 10, 2010, pp. 1160–1168.
- [7] N. Jabeur and P. A.Graniero, "Agent-Based Clusters to Virtually Manage Spatially Distributed Sensors", Projects TDMDS08 and SII-73 University of Windsor, Department of Earth and Environmental Sciences, Windsor, Ontario, Canada, 2008.
- [8] L. Min, X. Shijun, and S.Siyi, "An agent-based routing algorithm assistedQoS-forwireless sensor networks", *Journal of Network and Computer Applications*, vol. 35, no. 1, 2012, pp. 29–36.
- [9] D. Wajgi and N.V. Thakur, "Load Balancing Algorithms in Wireless Sensor Network : A Survey" *International Journal of Computer Networks and Wireless Communications (IJCNWC)*, ISSN: 2250-3501, vol. 2, no. 4, Aug. 2012, pp. 456–460.
- [10] Z. Bidai, M. Maimour, and H. Haffaf, "Multipath Extension of the ZigBee Tree Routing in Cluster-Tree Wireless Sensor Networks", *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)*, vol. 4, no. 2, Apr. 2012, pp. 30–48
- [11] B. Rijubrata, "Mobile Agent based Architecture for Wireless Sensor Networks", In proceedings of TKK T-110.5190 Seminar on Internetworking, Helsinki University of Technology, Finland, Apr. 27, 2009.
- [12] K.B. Pratik, Q.Hairong, and Y. Xu, "Collaborative Mobile-agent-based sensor fusion", *Information Fusion*, vol. 9, 2008, pp. 399–411.
- [13] R.H.B. Rahal and B.Bounabat, "Wireless Sensor Network Simulation Of The Energy Consumption By A Multi Agent System", *Journal of Theoretical and Applied Information Technology*, vol. 25, no. 1, 2011, pp. 50–57.
- [14] B. Dario, S. Marco, B. Andrea, C. Davide, and G. Marco, "Markovian agent modeling swarm intelligence algorithms in wireless sensor networks". *Journal of Performance Evaluation*, vol. 69, no. 3–4, 2012, pp. 135–149.
- [15] D.C. Matthew, A. Djamel, P.H. Barry, and H.Alan, "A Bayesian network approach to a biologically inspired motion strategy for mobile wireless sensor networks", *Journal of Ad Hoc Networks*, vol. 7, no. 6, 2009, pp. 1217–1228.
- [16] V. Saamaja, K.R. Kiran, and R.M.C. Siva, "Using mobile data collectors to Improve network lifetime of wireless sensor networks with reliability constraints", *Journal of Parallel and Distributed Computing*, vol. 70, no. 7, 2010, pp. 767–778.
- [17] G. Wenzhong, X. Naixue, C. Han-Chieh, H. Sajid, and C.Guolong, "Design and Analysis of Self-Adapted Task Scheduling Strategies in Wireless Sensor Networks". *Journal of Sensors*, vol. 11, 2011, pp. 6533–6554.
- [18] H. Hai-ping, W. Ru-chuan, S. Li-juan, W. Hai-yuan, and X. Fu, "Research on tasks schedule and data transmission of video sensor

- networks based on intelligent agents and intelligent algorithms", *The Journal of China Universities of Posts and Telecommunications*, vol. 16, no. 6, 2009, pp. 84-91.
- [19] A.A. Azrina, Y.A. Sekercioglu, P. Fitzpatrick, and M.Ivanovich, "A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks", *Communications Surveys and Tutorials, IEEE*, vol. 15, no. 1, 2013, pp. 121-144.
- [20] S. Narayanaswamy, V. Kawadia, R.S. Sreenivas, and P.R. Kumar, "Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the protocol compow", In proceedings of European Wireless Conference (2002), pp. 156-162.
- [21] X. Ya, J. Heidemann and D.Estrin, "Geography-informed energy conservation for ad hoc routing", In proceedings of 7th Annual international conference on mobile computing and networking-MOBICOM, New York, USA, 2001, pp. 70-84.
- [22] A. Cerpa and D. Estrin, "Ascent: Adaptive Self-Configuring sensor networks topologies", *IEEE Transactions on Mobile Computing*, vol. 3, no. 3, 2004, pp. 272-285.
- [23] D.Z. Du and P. Pardalos, "Handbook of Combinatorial Optimization", Kluwer Academic Publishers, 2004.
- [24] L. Bao and J.J. Garcia-Luna-Aceves, "Topology management in ad hoc networks", In proceedings of 4th ACM International Symposium on Mobile AdHoc Networking and Computing, ACM New York, NY, USA, 2003, pp. 129-140.
- [25] V. Kawadia and P.R. Kumar, "Power control and clustering in ad hoc networks", In INFOCOM 2003. Twenty-second annual joint conference of the IEEE Computer and Communications. IEEE Societies, 30 March-3 Apr. 2003, pp. 459-469.
- [26] Y. Rong-rong, L. Bin, L. Ya-qian, and H. Xiao-chen, "Adaptively fault-tolerant topology control algorithm for wireless sensor networks", *The Journal of China Universities of Posts and Telecommunications*, vol. 19, no. Suppl. 2, 2012, pp. 13-18.
- [27] C. Bo-Chao, L. Guo-Tan, T. Ryh-Yuh, and H. Ping Hsu, "Network lifetime bounds for hierarchical wireless sensor networks in the presence of energy constraints", *Journal of Computer Networks*, vol. 56, no. 2, 2012, pp. 820-831.
- [28] G. Omprakash, F. Rodrigo, J. Kyle, M. David, and L. Philip, "Collection Tree Protocol", In proceedings of the 7th ACM Conference on embedded networked sensor systems, Berkeley, CA, USA, Nov. 2009, pp. 1-14.
- [29] B. Dario, D. Salvatore, L. Francesco, P. Antonio, and S. Marco, "Evaluating wireless sensor node longevity through Markovian techniques", *Journal of Computer Networks*, vol. 56, no. 2, 2012, pp. 521-532.
- [30] A.K. Safdar, D. Boubaker, and D. Karim, "Application of fuzzy inference systems to detection of faults in wireless sensor networks", *Journal of Neurocomputing*, vol. 94, no. 1, 2012, pp. 111-120.
- [31] C. Paolo, C. Matteo, B. Stefano, and C. Luca, "A Cooperative Approach for Topology Control in Wireless Sensor Networks", *Pervasive and Mobile Computing*, vol. 5, no. 5, 2009, pp. 526-541.
- [32] OMNeT++ Homepage <http://www.omnetpp.org>.
- [33] F. Celik, A. Zengin, and S.Tuncel, "A survey on swarm intelligence based routing protocols in wireless sensor networks", *International Journal of the Physical Sciences*, vol. 5, no. 14, 2010, pp. 2118-2126.
- [34] V. Rodoplu and T. H. Meng, "Minimum energy mobile wireless networks", *IEEE J. Salt. Areas Communications*, vol. 17, no. 8, 1999, pp. 1333-1344.
- [35] Z. Yuanyuan, X. Jia, and H. Yanxiang, "Energy efficient distributed connected dominating sets in wireless sensor networks building", In proceedings of the 2006 International conference on wireless communications and mobile computing, New York, USA, ACM 2006, pp. 797-802.
- [36] B. Chazelle, "Natural algorithms and influence systems", *Communications of the ACM*, vol. 55 no. 12, Dec. 2012, pp. 101-110.

# Outage Performance and Derivation due to Adjacent Channel Interference

YoungKeun Yoon, YoungJun Chong  
 Radio Technology Research Department  
 Electronics and Telecommunications Research Institute  
 Daejeon, Republic of Korea  
 ykyoon@etri.re.kr, yjchong@etri.re.kr

**Abstract**—This paper describes the mathematical formula derivation method of the outage probability and compares the mathematical formula with statistical Monte-Carlo (MC) simulation results. The outage probability is the factor considered in the adjacent channel interference (ACI) impact between a victim terminal station and multiple interfering terminal stations. For protection of a victim terminal station from harmful interference, we would calculate the out of band emission limit of an interfering terminal station, too. The distribution of multiple interfering terminal stations within a cell follows the Poisson point process. The propagation model is a composite median pathloss and shadow fading with Log-normal distribution. Outage probability is obtained and evaluated with various parameters based on Long Term Evolution (LTE) Time Division Duplex (TDD) terminal stations.

**Keywords**—mask; interference; block; channel; adjacent

## I. INTRODUCTION

Due to the increase of the rapid data transmission with high capacity and the scarcity of available frequencies, the capacity performance of a mobile communication system may be reduced. When multiple terminal stations are located within the close proximity and the frequency bands are assigned in the adjacent bands like the reverse FDD (Frequency Division Duplex) assignment, the harmful interference among terminal stations may happen. Due to the aggregate interference of interfering terminal stations, out of band emission of an interfering terminal station should be suppressed as lower level than the reference limit level of the block edge mask of a terminal station for the protection of a victim station. The calculation of the block edge mask is still an open research issue. Both a deterministic minimum coupling loss (MCL) and a stochastic approach based on the Monte-Carlo (MC) simulation have been suggested in [1]. MC simulation is a computerized mathematical scheme and provides the decision-maker with a range of possible outcomes and the probabilities it will occur for any choice of action. The result generated is a probability of the interference or the outage. MC approach is a statistical scheme, which is to distribute a victim terminal station amongst a population of interferers. MC method is capable of modeling highly complex systems including a cellular system like LTE. MCL approach is relatively straight forward. MCL method is capable of modeling only a single interferer to a single victim station.

Statistical distribution of the aggregate ACI from multiple

interferers has been studied in relation to dynamic spectrum sharing on the legacy radio systems and the interference protection [2]-[3]. Log-normal distribution is used to approximate the probability distribution function (PDF) of aggregate interfering signals received at the center from multiple terminal stations distributed uniformly in an annual region with inner radius and outer radius [2]. Log-normal approximation does not match well due to a large difference of the interference received both from near and far away multiple terminal stations, when outer radius is several ten times larger than inner radius. In the other hands, the Log-normal approximation does work well for a system with an exclusive region such as the cognitive radio [3].

In this paper, we derive the mathematical formula of the outage probability optimized in the approximation for fitting with the coexistence of terminal stations of a small radius and in the LTE TDD system. This formula is capable of calculating the block edge mask and out of band limit for the adjacent channel sharing between operators or terminal stations. The outage probability means the total outages counted as the calculated signal reception level is lower than the reference threshold of the signal reception level. To offer practical protection limit from the aggregate interference, the derived equation's analytic results are in good agreement with the Monte-Carlo simulation ones.

## II. SCENARIO

Let us consider the scenario shown in Figure 1.

Two LTE systems use adjacent channel frequency bands

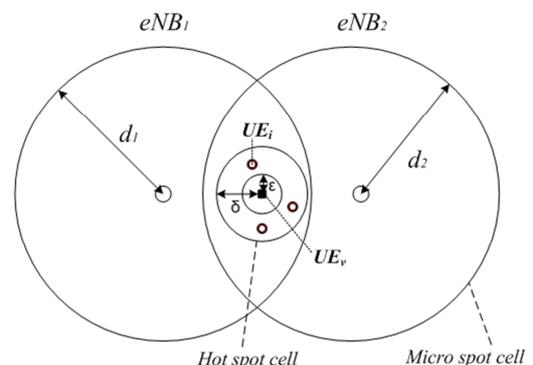


Figure 1. Geometry for the scenario

and their cell areas are overlapped. LTE terminal stations (UEs) are assumed to be spread homogeneously in the coverage of each LTE base station (eNB). One of LTE base

stations is a wanted transmitter ( $eNB_i$ ) and the other is an interfering transmitter ( $eNB_j$ ). A terminal station of a wanted transmitter is the victim station ( $UE_v$ ) and terminal stations of an interfering transmitter are interfering stations ( $UE_{is}$ ). The assumption is given that interfering stations ( $UE_{is}$ ) are located around any victim station ( $UE_v$ ) and  $UE_{is}$  cause the potential ACI to a  $UE_v$ . The distribution of the received aggregate ACI to a  $UE_v$  depends on random variable factors. Random variable factors are the transmission power of  $UE_{is}$ , the median pathloss, shadow fading, the adjacent channel interference ratio (ACIR), and a number of  $UE_{is}$ . It is assumed that ACIR is a certain value fixed identically for all interfering stations. The distribution and the concurrent transmission number of terminal stations in the cell are random variables. These random variables are used to calculate the distribution of the aggregate ACI.

The assumption of the geometry as shown in Figure 1 is as follows.  $UE_v$  is randomly distributed in the circle of the radius  $d_1$  of  $eNB_i$ .  $UE_{is}$  are distributed uniformly within the radius  $d_2$  and an angular ring with both inner radius of  $\epsilon$  and outer radius of  $\delta$ . The inner radius of  $\epsilon$  defines a minimum distance among terminal stations of  $UE_v$  and  $UE_{is}$ . The outer radius of  $\delta$  is determined to become the maximum radius where the distribution area of multiple interfering stations becomes an effective interference region because the interfering reception power to a victim station is within the range of valid values. The propagation model is a composite median pathloss and shadow fading with Log-normal distribution. Shadow fading model is assumed to have the path correlation among terminal stations.

### III. ANALYTIC DERIVATION OF OUTAGE PROBABILITY

Considering ACI dominant environment in the interference, the received signal to interfering power ratio (SIR) to  $UE_v$  is represented as follows.

$$\begin{aligned} SIR &= \frac{P_{eNB} P_{L_{eNB-UE_v}}}{I_{UE_i}} = \frac{P_{eNB} P_{L_{eNB-UE_v}}}{\sum P_{UE_i} \partial_{UE_i} P_{L_{UE_i-UE_v}}} \\ &= \frac{P_{eNB} C_1 d_1^{-\gamma_1} \vartheta_1}{\sum P_{UE_i} \partial_{UE_i} C_2 d_2^{-\gamma_2} \vartheta_2^i} = K \cdot \frac{d_1^{-\gamma_1} \vartheta_1}{\sum d_2^{-\gamma_2} \vartheta_2^i} \end{aligned} \quad (1)$$

where:

$P_{eNB}$  is the transmit power of a wanted transmitter

$\sum P_{UE_i}$  is the transmit power of interfering stations

$P_{L_{eNB-UE_v}}$  is the composite median pathloss and shadow fading of a wanted transmitter to a victim station

$P_{L_{UE_i-UE_v}}$  is the composite median pathloss and shadow fading of an interfering station to a victim station

$\partial_{UE_i}$  is ACIR, which means the adjacent channel interference ratio

Adjacent channel interference gives rise to extraneous

power of the received signal to a victim station. The adjacent channel interference is the sum of the power both that interfering stations emit into a victim station's channel known as the unwanted emission and that interfering stations pick up from a victim station's channel known as the adjacent channel selectivity (ACS). ACS occurs and do not completely eliminate an interfering signals because radio frequency (RF) filters required a roll-off. Therefore, a victim station emits some power in the adjacent channel picked up by an interfering station. An interfering station receives some emissions from a victim station's channel due to the roll off of the selectivity filters.

$C_1$ ,  $d_1$ , and  $\gamma_1$  are the parameters of the median pathloss model between a victim station and a wanted transmitter,  $C_2$ ,  $d_2$ , and  $\gamma_2$  are the parameters of the median pathloss model between a victim station and an interfering station.  $\vartheta_1$  and  $\vartheta_2^i$  are the Log normal random variables having zero-means, variances of  $\sigma_1^2$ ,  $\sigma_2^2$  between a victim station and a wanted transmitter link and between a victim station and an interfering station, respectively.  $K$  is a constant to substitute all the constant values.

With the log-normal approximation in the denominator of (1) PDF and cumulative distribution function (CDF) of SIR can be derived. Firstly, we transform the SIR of (1) into a SIR in the decibel scale for ease of the derivation.

$$\begin{aligned} SIR \text{ (dB)} &= 10 \log_{10} \left( K \cdot \frac{d_1^{-\gamma_1} \vartheta_1}{\sum d_2^{-\gamma_2} \vartheta_2^i} \right) = 10 \log_{10} \left( K \cdot \frac{\tau \cdot \mathfrak{N}_1}{\mathfrak{N}_I} \right) \\ &= 10 \log_{10}(K) + 10\beta(\hat{\tau} + \widehat{\mathfrak{N}}_1 - \widehat{\mathfrak{N}}_I) \end{aligned} \quad (2)$$

where,  $\hat{x}$  is defined as  $\ln(x)$  and  $\beta$  is  $1/\ln(10)$ .

With a few mathematical manipulations the PDF of  $\hat{\tau}$  is obtained.

$$\text{PDF}(\hat{\tau}) = \begin{cases} \frac{2}{\gamma_1 D^2} e^{-\frac{2\hat{\tau}}{\gamma_1}}, & -\gamma_1 \ln(D) \leq \hat{\tau} < \infty \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

Because, both  $\widehat{\mathfrak{N}}_1$  and  $\widehat{\mathfrak{N}}_I$  have a Gaussian distribution and the submission of the Gaussian random variables is a Gaussian random variable [4]. And,  $\widehat{\mathfrak{N}}_1$  or  $\widehat{\mathfrak{N}}_I$  has a Gaussian distribution:

$$f(\widehat{\mathfrak{N}}) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\widehat{\mathfrak{N}} - \mu)^2}{2\sigma^2}\right) \quad (4)$$

where,  $\mu = -\mu_I$  and  $\sigma^2 = \sigma_1^2 + \sigma_I^2$ .

The PDF of SIR in dB is obtained through the convolution of PDF ( $\hat{\tau}$ ) and  $f(\widehat{\mathfrak{N}})$  as follows.

$$f_X(x) = \frac{1}{5\gamma_1\beta} e^{-20x_t} Q(-x_t + \theta) \quad (5)$$

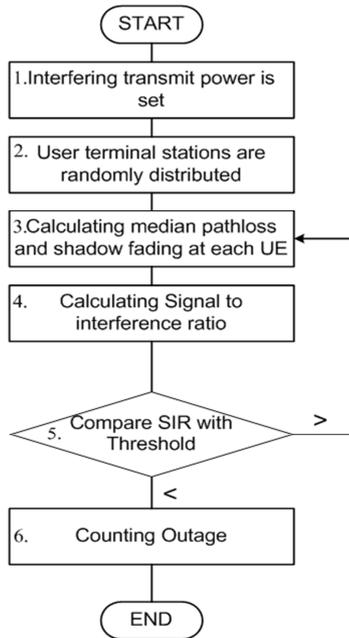


Figure 2. Monte-Carlo Simulation

where,  $x_t = \frac{1}{\sigma} \left( \frac{x}{10\beta} - \ln \left( \frac{K}{D^{\gamma_1}} \right) - \mu - \frac{\sigma^2}{\gamma_1} \right)$ ,  $\theta = \frac{\sigma}{\gamma_1}$ , and  $Q(y)$  is defined as  $\int_y^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

CDF of SIR in dB is derived using integration by parts and results are as follows.

$$F_X(x) = \int_{-\infty}^x f_x(t) dt = Q(-x_t - \theta) - e^{-2\theta x_t} Q(-x_t + \theta) \quad (6)$$

where,  $x_t$  and  $\theta$  are the same as those in (5)

Finally, we can have a formulation for the outage probability defining  $\Delta$ , which is the threshold of the outage as follows.

$$\begin{aligned} Prob(SR < \Delta) &= F_x(\Delta) \\ &= Q(-\Delta_t - \theta) - e^{-2\theta \Delta_t} Q(-\Delta_t + \theta) \quad (7) \end{aligned}$$

where,  $\Delta_t = \frac{1}{\sigma} \left( \frac{\Delta}{10\beta} - \ln \left( \frac{K}{D^{\gamma_1}} \right) - \mu - \frac{\sigma^2}{\gamma_1} \right)$ ,  $\theta = \frac{\sigma}{\gamma_1}$

#### IV. SIMULATION

Figure 2 shows the flow diagram of the Monte-Carlo simulation.

Step1: The emission power of  $UE_i$  is set to 23dBm.

Step2: All LTE terminal stations are randomly distributed in the cell.

Step3: The median pathloss and shadow fading between a wanted transmitter and a victim station is calculated. The median pathloss and shadow fading between an interfering station and a victim station is calculated.

Step4: The received signal to interfering power ratio (SIR) to a victim station is calculated.

Step5: SIR value is compared with reference threshold value of the predefined SIR. Reference threshold value of the received signal to interfering power ratio is defined from the quality of service of LTE terminal station. If a calculation value in Step4 is smaller than a reference threshold, the outage is happened and the outage means the blocking event. If a calculation value in Step4 is larger than a reference threshold, a call of a terminal station is not dropped, goes to Step 3.

Step6: Total outage is counted.

#### V. RESULTS

Let us consider the validation of the derived equation using the Monte-Carlo simulation.

We assumed the coexistence scenario of two E-UTRA (LTE) systems, which have the bandwidth of 10MHz and assign at adjacent channel bands without the guard band.

The base station of LTE system known as Evolved Node B (eNB) has the transmit power ( $P_{eNB}$ ) of 46dBm and the antenna gain with 12dBi including the feeder loss of -3dB. The terminal station of LTE system (UE) has the transmit power of 23dBm and the antenna gain of 0dBi. Cell radius ( $d$ ) of eNB is 500m. Inner radius ( $\epsilon$ ) is 1m when MCL is about 30dB (including 2dB body loss). Outer radius ( $\delta$ ) is 19m. UEs are uniformly distributed within the cell. ACS of a UE is 33dB. For predicting the radio propagation characteristics, both extended Hata model and Motley-Keenan formula are used. The extended Hata model [1] applies for calculating to the pathloss of the desired link between a wanted transmitter (eNB) and a victim station (UE). The Motley-Keenan formula [4] applies for the interfered link between interfering stations and a victim station in the small cell environment. Long term fading known as shadow fading is Log normal random variable having zero-mean and variances. The variance of  $\sigma_1^2$  is 12dB for the desired link and the variance of  $\sigma_2^2$  is 4dB for the interfered link. Total pathloss value of the communication link is the sum of the median pathloss and long term fading value.

Attenuation factors and the constant of the applied median pathloss model are used as follows: attenuation exponent ( $\alpha_1$ ) 3.52 and constant ( $C_1$ )  $10^{-2}$  in the extended Hata model and ( $\alpha_2$ ) 2.0 or 3.5 and constant ( $C_2$ )  $10^{-3.15}$  in the Motley-Keenan. The used pathloss equations are shown in (8) and (9).

$$PL_{eNB-UE} = C_1 \cdot d_1^{-\gamma_1} = 10^{-2} \cdot d_1^{-3.52} \quad (8)$$

$$PL_{UE-UE} = C_2 \cdot d_2^{-\gamma_2} = 10^{-3.15} \cdot d_2^{-2} \quad (9)$$

Figure 3 shows CDF of the outage probabilities on both pathloss attenuation ( $\alpha_2$ ) of 2.0 or 3.5 in the indoor interfered links and interfering station densities  $\lambda$  of 2 or 4. BEM OOB limit is assumed as  $-10\text{dBm}/10\text{MHz}$ . ACIR is calculated as  $1/((1/\text{ACS})+(1/\text{ACLR}))$ , where, ACLR means the adjacent channel leakage ratio. Normally, ACLR of LTE UE defines as  $30\text{dB}+X$  at the first adjacent channel. In Figure 3, the analytic derivation results are in good agreement with the Monte-Carlo simulation results. Also, we can find that both larger number of interfering stations and lower pathloss exponents enhance the accuracy of derivation. This enhancement is the reason that large numbers of samples make the approximation more precise by central limit theorem, and large pathloss exponent does the deviation of the interference levels to increase according to the distance between interfering users and a victim station.

Figure 4 shows CDF of outage probabilities on both BEM levels of  $0\text{dBm}$  or  $-20\text{dBm}$  in the indoor interfered links. Interfering station density ( $\lambda$ ) is 2 and the pathloss attenuation ( $\alpha_2$ ) is 3.5. Outage probabilities are almost consistent on analytical as well as simulation results. In Figure 4, we can identify that as we allow more interference into adjacent band by increasing BEM OOB limit, the outage probability increases due to the increased interference. If we set a minimum SIR of  $0\text{dB}$  for 0.1 error rate (outage probability) in the application, then, BEM OOB limit of  $-20\text{dBm}$  should be selected. Finally, the approximation equation of the derived outage probability can be used instead of the simulation results and it can also be applied for calculating BEM OOB limits that require the outage probability of a victim system.

## VI. CONCLUSIONS

In this paper, we derived an optimized probability formula for outages due to the unwanted emission of the interfering LTE stations in the adjacent channel bands. Analytic results of the mathematical formula were compared with statistical Monte-Carlo (MC) simulation results. For Monte-Carlo simulation, it is assumed that terminal stations are uniformly distributed around the hot spot cell area. For optimized formulation derivation, the composite median pathloss and long-term fading have log-normal approximation.

As a result, the Log-normal approximation performs well in spite of a large deviation of interference received from both near and far away terminals although there are some mismatches in absolute values, when outer radius is  $19\text{m}$  and inner radius of  $1\text{m}$ . Also, we can apply for calculating the out of band emission limit of block edge mask. The analytic and Monte-Carlo simulation results are useful for current and future network system performance analysis.

## ACKNOWLEDGMENT

This work was supported by the IT R&D program of

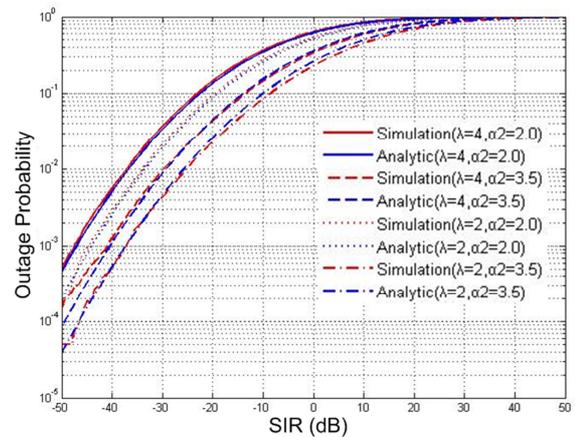


Figure 3. Outage Probabilities for pathloss exponents & an interfering user density

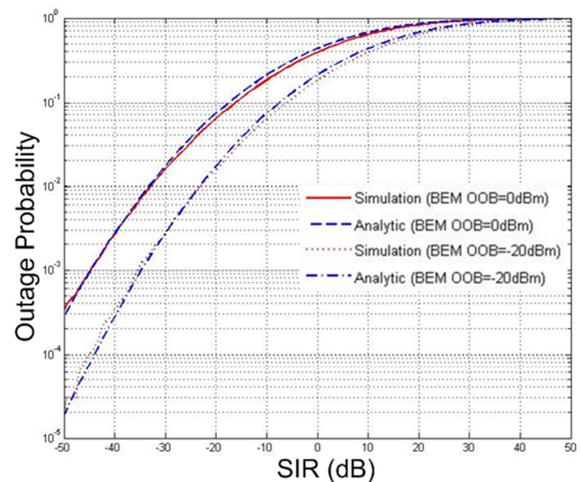


Figure 4. Outage Probabilities for BEM OOB limits

## MSIP/KCA

## REFERENCES

- [1] ECC Report 131, "Derivation of a Block Edge Mask (BEM) for Terminal Stations in the 2.6GHz Frequency Bands (2055-2690MHz)," Dublin, Jan. 2009.
- [2] R. Menon, R. M. Buehrer, and J. H. Reed, "On the impact of dynamic spectrum sharing techniques on legacy radio systems," *IEEE Transaction on Wireless Communications*, vol. 7, no. 11, Nov. 2008, pp. 4198 – 4207.
- [3] K. W. Sung, M. Tercero, and J. Zander, "Aggregate interference in secondary access with interference protection," *IEEE Communication Letters*, vol. 15, no. 6, June 2011, pp. 629 – 631.
- [4] M. J. Nawrochi, M. Dohler, and A. H. Aghvami, *Understanding UMTS Radio Network Modelling, Planning and Automated Optimisation*. Chichester: John Wiley and Sons, 2006.
- [5] A. Leon-Garcian, *Probability and Random Processes for Electrical Engineering*, 2<sup>nd</sup> Edition, Addison-Wesley Publishing Company, 1993.
- [6] Y. Shim, I. Lee, and S. Park, "The impact of LTE UE on Audio Devices," *ETRI Journal*, vol. 35, no.2, Apr. 2013, pp. 332-335.

# Uplink Performance Evaluation of Broadband Systems which Adopt a Massive MU-MIMO Approach

Paulo Torres

EST/PCB - Instituto Politecnico de Castelo Branco  
Castelo Branco, Portugal  
Email: paulo.torres@ipcbr.pt

Luis Charrua<sup>†</sup>, Antonio Gusmao<sup>‡</sup>

CAPS/IST - Instituto Superior Tecnico, Universidade de Lisboa  
Lisboa, Portugal  
Email: <sup>†</sup>luischarrua@enautica.pt, <sup>‡</sup>gus@ist.utl.pt

**Abstract**—This paper deals with an Orthogonal Frequency Division Multiplexing (OFDM)-based uplink within a Multi User (MU)-Multi-Input Multi-Output (MIMO) system where a "massive MIMO" approach" is adopted. In this context, either an optimum Minimum Mean-Squared Error (MMSE) linear detection or a reduced-complexity Matched Filter (MF) linear detection are considered. Regarding performance evaluation by simulation, several semi-analytical methods are proposed: one performance evaluation method in the optimum (MMSE) case; two performance evaluation methods in the MF case. This paper includes performance results for uncoded 4-Quadrature Amplitude Modulation (QAM)/OFDM transmission and a MU-MIMO channel with uncorrelated Rayleigh fading, under the assumptions of perfect power control and perfect channel estimation. The accuracy of performance results obtained through the semi-analytical simulation methods is assessed by means of parallel conventional Monte Carlo simulations [10]. The performance results are discussed in detail and we also emphasize the achievable "massive MIMO" effects, even for the reduced-complexity detection techniques, provided that the number of BS antennas is much higher than the number of antennas which are jointly employed in the terminals of the multiple autonomous users.

**Keywords**- *Broadband Wireless Communications; MU-MIMO Systems; Massive MIMO; Performance Evaluation; OFDM.*

## I. INTRODUCTION

The development of MIMO technologies has been crucial for the "success story" of broadband wireless communications in the last two decades [1]. Through spatial multiplexing schemes, following and extending ideas early presented in [2], MIMO systems are currently able to provide very high bandwidth efficiencies and a reliable radiotransmission at data rates beyond 1 Gigabit/s. Appropriate MIMO detection schemes, offering a range of performance/complexity tradeoffs, have been essential for the technological improvements in this area [1][3][4]. In the last decade, MU-MIMO systems have been successfully implemented and introduced in several broadband communication standards [5]; in such "space division multiple access" systems, the more antennas the Base Station (BS) is equipped with, the more users can simultaneously communicate in the same time-frequency resource.

In recent years, the adoption of MU-MIMO systems with a very large number of antennas in the BS, much larger than the number of Mobile Terminal (MT) antennas in its cell, was

proposed by Marzetta [6]. This "massive MIMO" approach has been shown to be recommendable for several reasons [6][7][8][9]: simple linear processing for MIMO detection becomes nearly optimal; both MultiUser Interference (MUI) effects and fast fading effects of multipath propagation tend to disappear; both power efficiency and bandwidth efficiency become substantially increased.

This paper deals with an OFDM-based uplink within a MU-MIMO system where the BS is constrained to adopt simple, linear detection techniques but can be equipped with a large number of receiver antennas. In this context, either an optimum (MMSE) linear detection or a reduced-complexity MF linear detection are considered in Section II. Regarding performance evaluation by simulation, several semi-analytical methods are proposed in Section III, all of them combining simulated channel realizations and analytical computations of BER performance which are conditional on those channel realizations: one performance evaluation method in the optimum (MMSE) case; two performance evaluation methods in the MF case.

In Section IV, this paper includes performance results for uncoded 4-QAM/OFDM transmission and a MU-MIMO channel with uncorrelated Rayleigh fading effects regarding the several transmitter/receiver (TX/RX) antenna pairs, under the assumptions of perfect power control and perfect channel estimation. The accuracy of performance results obtained through the semi-analytical simulation methods is assessed by means of parallel conventional Monte Carlo simulations (involving an error counting procedure). The performance results are discussed in detail and we also emphasize the achievable "massive MIMO" effects, even for the reduced-complexity detection techniques, provided that the number of BS antennas is much higher than the number of antennas which are jointly employed in the terminals of the multiple autonomous users. Section V includes the main conclusions of the paper.

## II. SYSTEM MODEL

### A. OFDM-based Radiotransmission

We consider here a Cyclic Prefix (CP)-assisted, OFDM-based, block transmission, within a MU-MIMO system with  $N_T$  TX antennas and  $N_R$  RX antennas - for example (but not necessarily) one antenna per MT, as depicted in Figure 1(a).

We assume, in the  $j$ th TX antenna ( $j = 1, 2, \dots, N_T$ ), a length- $N$  block  $\mathbf{S}^{(j)} = [S_0^{(j)}, S_1^{(j)}, \dots, S_{N-1}^{(j)}]^T$  of frequency-domain data symbols in accordance with the corresponding binary data block. The insertion of a length- $L_s$  CP, long enough to cope with the time-dispersive effects of multipath propagation, is also assumed after the IDFT that is required to bring the data block information to the time domain.

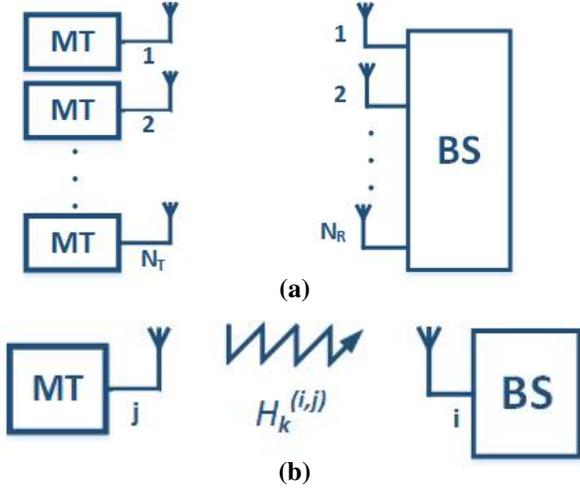


Figure 1. Uplink of an  $N_T \times N_R$  MU-MIMO system (a) and channel characterization, at subcarrier  $k$ , concerning the antenna pair  $(i, j)$  (b).

The frequency-domain data symbols  $S_k^{(j)}$  ( $k = 0, 1, \dots, N-1; j = 1, 2, \dots, N_T$ ) are randomly and independently selected from a QAM alphabet ( $E[S_k^{(j)}] = 0$  and  $E[|S_k^{(j)}|^2] = \sigma_S^2$  for any  $(j, k)$ ).

For any subcarrier, the frequency domain transmission rule can be described as follows:

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{S}_k + \mathbf{N}_k, \quad (1)$$

where  $\mathbf{S}_k = [S_k^{(1)}, S_k^{(2)}, \dots, S_k^{(N_T)}]^T$  is the "input" vector,  $\mathbf{N}_k = [N_k^{(1)}, N_k^{(2)}, \dots, N_k^{(N_R)}]^T$  is the Gaussian noise vector ( $E[N_k^{(i)}] = 0$  and  $E[|N_k^{(i)}|^2] = \sigma_N^2 = N_0 N$ ),  $\mathbf{H}_k$  denotes the  $N_R \times N_T$  channel matrix with entries  $H_k^{(i,j)}$ , concerning a given channel realization (RX antenna  $i$  and TX antenna  $j$ , for each subcarrier, as shown in Figure 1(b)) and  $\mathbf{Y}_k = [Y_k^{(1)}, Y_k^{(2)}, \dots, Y_k^{(N_R)}]^T$  is the resulting "output" vector.

By assuming  $E[H_k^{(i,j)}] = 0$  and a constant

$$E[|H_k^{(i,j)}|^2] = P_\Sigma, \quad (2)$$

for any  $(i, j, k)$ , and a 4-QAM/OFDM block transmission, the average bit energy at each BS antenna is given by

$$E_b = \frac{\sigma_S^2}{2\eta N} P_\Sigma, \quad (3)$$

where  $\eta = \frac{N}{N+L_s}$ , with  $L_s$  denoting the CP length.

## B. Optimum (MMSE) and MF Linear Detection Techniques

With regard to subcarrier  $k$ , two linear detection techniques are considered, as shown in Figure 2, both directly providing frequency-domain decisions  $\hat{\mathbf{S}}_k$  based on the frequency-domain detector output  $\tilde{\mathbf{Y}}_k$  ( $k = 0, 1, \dots, N-1$ ).

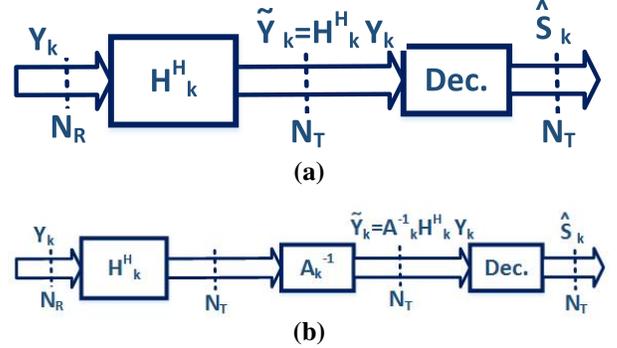


Figure 2. Reduced-complexity (MF) (a) and optimum (MMSE) (b) linear detection techniques, regarding subcarrier  $k$ .

As Figure 2(a) indicates,

$$\tilde{\mathbf{Y}}_k = \mathbf{H}_k^H \mathbf{Y}_k \quad (4)$$

for the reduced-complexity linear detection technique. The  $\tilde{Y}_k^{(j)}$  components of  $\tilde{\mathbf{Y}}_k$  ( $k = 0, 1, 2, \dots, N-1; j = 1, 2, \dots, N_T$ ) are given by

$$\tilde{Y}_k^{(j)} = \sum_{i=1}^{N_R} H_k^{(i,j)*} Y_k^{(i)}, \quad (5)$$

which clearly means, for any  $(j, k)$  pair, a Maximal Ratio Combining (MRC) procedure, involving MF for each component of the length- $N_R$  received vector  $\mathbf{Y}_k$ .

For the optimum (MMSE) linear detection technique, it can be shown that [4]

$$\tilde{\mathbf{Y}}_k = \mathbf{A}_k^{-1} \mathbf{H}_k^H \mathbf{Y}_k, \quad (6)$$

where

$$\mathbf{A}_k = \mathbf{H}_k^H \mathbf{H}_k + \alpha \mathbf{I}_{N_T}, \quad (7)$$

with  $\alpha = \frac{\sigma_N^2}{\sigma_S^2} = N_0 \frac{N}{\sigma_S^2}$ .

## III. SEMY-ANALYTICAL METHODS FOR PERFORMANCE EVALUATION

### A. Performance Evaluation Method in the Optimum (MMSE) Case

The frequency-domain output  $\tilde{\mathbf{Y}}_k$  of the MMSE detector in Figure 2 (b) includes Gaussian noise and residual MUI terms in its  $N_T$  components. Regarding the  $j$ th component of  $\tilde{\mathbf{Y}}_k$ , the resulting Signal-to-Interference-plus-Noise-Ratio (SINR) can be derived by resorting to well-known "MMSE estimation" principles. It can be written as

$$\text{SINR}_{j,k} = \frac{\Gamma_k^{(j,j)}}{1 - \Gamma_k^{(j,j)}}, \quad (8)$$

where  $\Gamma_k^{(j,j)}$  is the  $(j, j)$  entry of

$$\mathbf{\Gamma}_k = [\mathbf{H}_k^H \mathbf{H}_k + \alpha \mathbf{I}_{N_T}]^{-1} \mathbf{H}_k^H \mathbf{H}_k, \quad (9)$$

$$\left( \Gamma_k^{(j,j)} = \frac{E[S_k^{(j)*} \tilde{Y}_k^{(j)}]}{\sigma_S^2}, \right. \quad \text{since}$$

$$\tilde{Y}_k^{(j)} = \Gamma_k^{(j,j)} S_k^{(j)} + \text{'uncorrelated noise - like term'} \left. \right).$$

For 4-QAM/OFDM, the resulting  $BER_{j,k}$  - conditional on the channel realization  $\mathbf{H}_k$  ( $k = 0, 1, \dots, N-1$ ) - is given by

$$BER_{j,k} \approx Q\left(\sqrt{SINR_{j,k}}\right), \quad (10)$$

with  $SINR_{j,k}$  according to (8). The average BER for the overall channel realization  $\mathbf{H}_k$  ( $k = 0, 1, \dots, N-1$ ) can be computed as follows:

$$BER = \frac{1}{N_T} \sum_{j=1}^{N_T} BER_j, \quad (11)$$

where

$$BER_j = \frac{1}{N} \sum_{k=0}^{N-1} BER_{j,k}, \quad (12)$$

#### B. Performance Evaluation Methods in the MF Case

The  $N_T$  components of the frequency domain output  $\tilde{\mathbf{Y}}_k$ , in the MF detector of Figure 2 (a), can be decomposed - into "useful signal", MUI and "Gaussian noise" - as follows:

$$\tilde{Y}_k^{(j)} = \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2 S_k^{(j)} + \sum_{l \neq j} \sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*} S_k^{(l)} + \sum_{i=1}^{N_R} H_k^{(i,j)*} N_k^{(i)}, \quad (13)$$

( $j = 1, 2, \dots, N_T$ ). Regarding the  $j$ th component of  $\tilde{\mathbf{Y}}_k$ , the resulting SINR can be given by

$$SINR_{j,k} = \frac{\sigma_S^2 \left( \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2 \right)^2}{\sigma_S^2 \sum_{l \neq j} \left| \sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*} \right|^2 + NN_0 \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}$$

$$= \frac{1}{\alpha_k^{(j)}} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2, \quad (14)$$

where

$$\alpha_k^{(j)} = \alpha + \sum_{l \neq j} \alpha_k^{(l,j)}, \quad (15)$$

with  $\alpha = N_0 \frac{N}{\sigma_S^2}$  and

$$\alpha_k^{(l,j)} = \frac{\left| \sum_{i=1}^{N_R} H_k^{(i,j)*} H_k^{(i,l)} \right|^2}{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}, \quad (16)$$

For 4-QAM/OFDM, the BER computation for the overall channel realization  $\mathbf{H}_k$  ( $k = 0, 1, \dots, N-1$ ) can be performed in accordance with eqns. (10), (11) and (12), through the use of  $SINR_{j,k}$  given by eqns. (14), (15) and (16).

The second semi-analytical method for performance evaluation considers the MUI as it is, and does not necessarily regard it as a "quasi-Gaussian interference". By assuming  $\sigma_S = \sqrt{2}$  (i.e.,  $S_k^{(j)} = \pm 1 \pm j$ ), it is easy to derive

$$BER_{j,k} = \frac{1}{2^{2N_T-1}} \sum_{\{\tilde{S}_k^{(l)}; l \neq j\}} \left[ Q \left( \frac{H_\Sigma + \Re e \left\{ \sum_{l \neq j} \left( \sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*} \right) \tilde{S}_k^{(l)} \right\}}{\sqrt{N \frac{N_0}{2} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}} \right) + \right.$$

$$\left. + Q \left( \frac{H_\Sigma + \Im m \left\{ \sum_{l \neq j} \left( \sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*} \right) \tilde{S}_k^{(l)} \right\}}{\sqrt{N \frac{N_0}{2} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}} \right) \right], \quad (17)$$

with  $H_\Sigma = \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2$ ; then, by resorting to eqns. (11) and (12), we can get the average BER for the "overall channel" realization  $\mathbf{H}_k$  ( $k = 0, 1, \dots, N-1$ ).

#### IV. NUMERICAL PERFORMANCE RESULTS AND DISCUSSION OF MASSIVE MIMO EFFECTS

In the following, we present a set of performance results for uncoded 4-QAM/OFDM uplink block transmission, with  $N = 256$  and  $L_s = 64$ , within a MU-MIMO  $N_T \times N_R$  Rayleigh fading channel. The fading effects regarding the several TX/RX antenna pairs are assumed to be uncorrelated, with all zero-mean complex Gaussian  $H_k^{(i,j)}$  channel coefficients having the same variance  $P_\Sigma$  (see Section II-A).

Figure 5 and Figure 6 involve subsets of BER performance curves taken from Figure 3 and Figure 4, respectively: in both cases, we selected the  $N_T \times N_R$  MU-MIMO MF linear detection and the SIMO  $1 \times N_R$  (single-path) performance, for several values of  $N_R$ . Figure 7 shows BER performance results, under MF linear detection, for  $N_R = 100$  and several values of  $N_T$  [SIMO  $1 \times N_R$  (single-path) case also included.].

With regard to both linear detection techniques of Section II-B, the several performance results concerning the MU-MIMO system have been obtained by random generation of a large number of channel realizations, analytical BER computation - according to the methods of Section III - for each channel realization, and an averaging operation over the set of channel realizations. The accuracy of performance results obtained through these semi-analytical simulation methods was assessed by means of parallel conventional Monte Carlo simulations (involving an error counting procedure). As expected, having in mind the subcarrier-by-subcarrier detection procedure in the uncoded QAM/OFDM block transmission context, the achieved performances turned out to be the same for frequency-flat and frequency-selective fading conditions (under the assumption of a CP long enough to cope with the channel time dispersion).

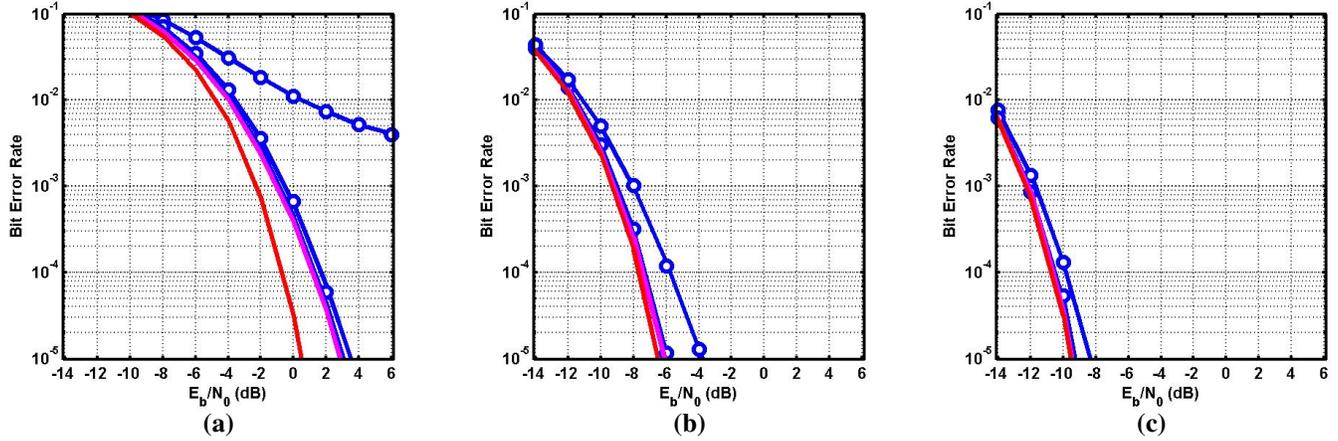


Figure 3. BER performances for OFDM-based MU-MIMO with  $N_T = 2$ , and  $N_R = 10$  (a), 50 (b) or 100 (c), under linear (MF, MMSE) and ML detection [SIMO  $1 \times N_R$  (single-path, multipath) reference BER performances are also included, and the five BER performances are ordered, from the worst to the best, as explained in section IV].

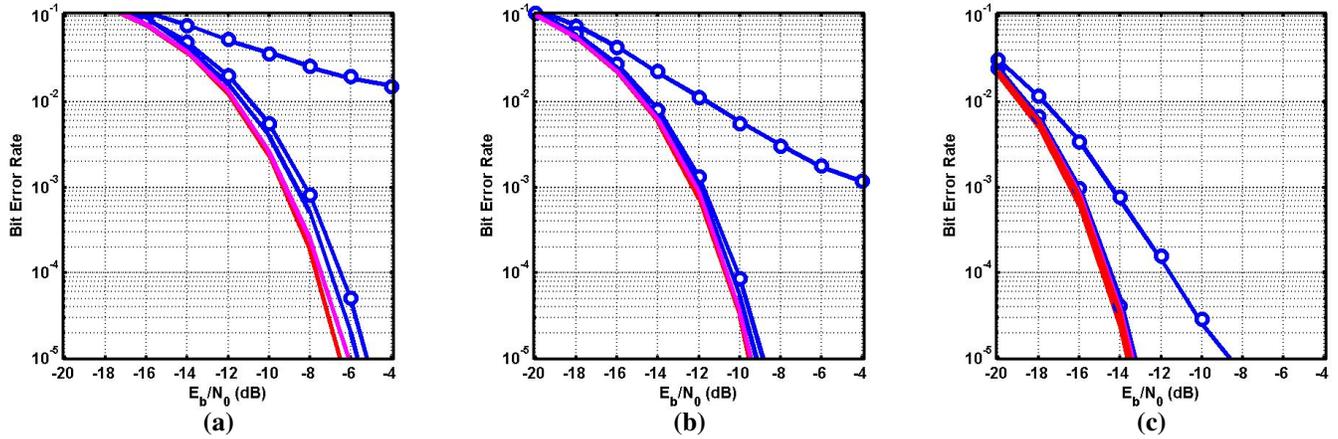


Figure 4. BER performances for OFDM-based MU-MIMO with  $N_T = 10$ , and  $N_R = 50$  (a), 100 (b) or 250 (c), under linear (MF, MMSE) and ML detection [SIMO  $1 \times N_R$  (single-path, multipath) reference BER performances are also included, and the five BER performances are ordered, from the worst to the best, as explained in section IV].

When  $N_R \gg N_T$ , both the MUI effects and the fading effects of multipath propagation tend to disappear: consequently, the BER performances for the MU-MIMO  $N_T \times N_R$  Rayleigh fading channel become very close to those concerning a SIMO  $1 \times N_R$  channel with single-path propagation for all  $N_R$  TX/RX antenna pairs. The achievable performances under a "truly massive" MU-MIMO implementation can be analytically derived as explained in the following.

Entries of  $\mathbf{H}_k$  are i.i.d. Gaussian-distributed random variables with zero mean and variance  $P_\Sigma$ . According to the law of large numbers [10],

$$\lim_{N_R \rightarrow \infty} \left[ \frac{1}{N_R} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2 \right] = E \left[ |H_k^{(i,j)}|^2 \right] = P_\Sigma \quad (18)$$

and

$$\begin{aligned} \lim_{N_R \rightarrow \infty} \left[ \frac{1}{N_R} \sum_{\substack{i=1 \\ (l \neq j)}}^{N_R} H_k^{(i,j)*} H_k^{(i,l)} \right] &= \quad (19) \\ &= \frac{E}{(l \neq j)} \left[ H_k^{(i,j)*} H_k^{(i,l)} \right] = 0. \end{aligned}$$

Therefore, having in mind (14), (15), (16) and (3),

$$\begin{aligned} \lim_{N_R \rightarrow \infty} \left( \frac{SINR_{j,k}}{N_R} \right) &= \frac{\sigma_S^2}{N_0 N} \lim_{N_R \rightarrow \infty} \left( \frac{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}{N_R} \right) \\ &= \frac{\sigma_S^2}{N_0 N} P_\Sigma = 2\eta \frac{E_b}{N_0} \quad (20) \end{aligned}$$

When  $N_R \gg N_T$  (which also implies  $N_R \gg 1$ )

$$SINR_{j,k} \approx N_R \lim_{N_R \rightarrow \infty} \left[ \frac{SINR_{j,k}}{N_R} \right] = 2\eta N_R \frac{E_b}{N_0}, \quad (21)$$

leading to

$$BER \approx Q \left( \sqrt{2\eta N_R \frac{E_b}{N_0}} \right). \quad (22)$$

Figures 3 and 4 show the simulated BER performances for an OFDM-based MU-MIMO uplink and several possibilities regarding  $N_T$  and  $N_R$ , when using the linear (MF and

MMSE) detection techniques of Section II. In both figures, for the sake of comparisons, we include Maximum Likelihood (ML) detection results concerning the  $N_T \times N_R$  system; we also include SIMO  $1 \times N_R$  reference performances, for both the multipath propagation channel - which implies a Rayleigh fading concerning each TX/RX antenna pair - and an ideal single-path propagation channel. For the linear detection techniques, the semi-analytical methods of Section III have been adopted; the complementary conventional Monte Carlo simulation (involving error counting) results correspond to the superposed circles in the solid lines.

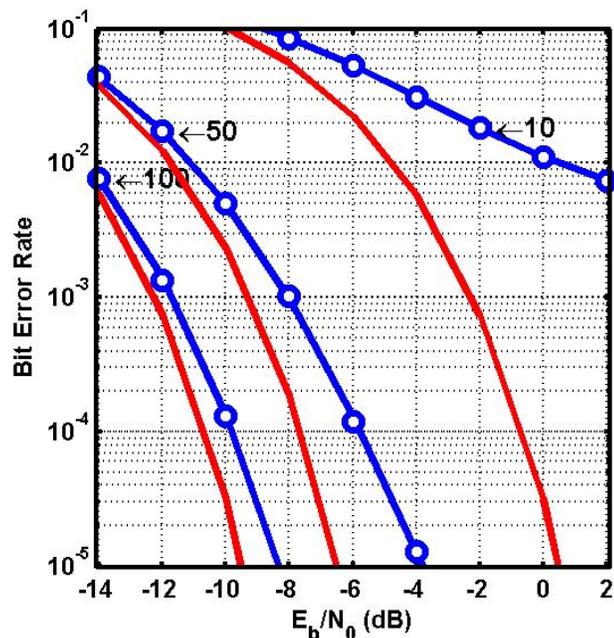


Figure 5. BER performances for OFDM-based MU-MIMO with  $N_T = 2$ , and  $N_R = 10, 50$  or  $100$ , under MF linear detection [SIMO  $1 \times N_R$  (single-path) reference performances are also included].

In all figures, where the  $1 \times N_R$  (single-path case) SIMO detection performance was analytically computed according to (22) - an excellent agreement of the semi-analytical simulation results with conventional Monte Carlo simulation results can be observed.

In the simulation results concerning all subfigures of both Figure 3 and Figure 4, the five BER performance curves have been shown to be ordered, from the worst to the best, as follows:  $N_T \times N_R$  MU-MIMO with MF linear detection;  $N_T \times N_R$  MU-MIMO with MMSE linear detection;  $N_T \times N_R$  MU-MIMO with ML detection;  $1 \times N_R$  (multipath case) SIMO detection;  $1 \times N_R$  (single-path case) SIMO detection. These figures clearly show that the performance penalty which is inherent to the reduced-complexity (MF) linear detection - as compared with the optimum (MMSE) linear detection and even the optimum (ML) detection - can be made quite small, by increasing  $N_R$  significantly; they also show that, under highly increased  $N_R$  values, the "MUI-free" SIMO (multipath) performance and the ultimate bound - the "MUI-free and fading-free" SIMO (single-path) performance - can be closely approximated, even when adopting the reduced-complexity (MF) linear detection. Figure 5 and Figure 6 emphasize the performance benefits of an increased  $N_R$ , for a given  $N_T$ , and

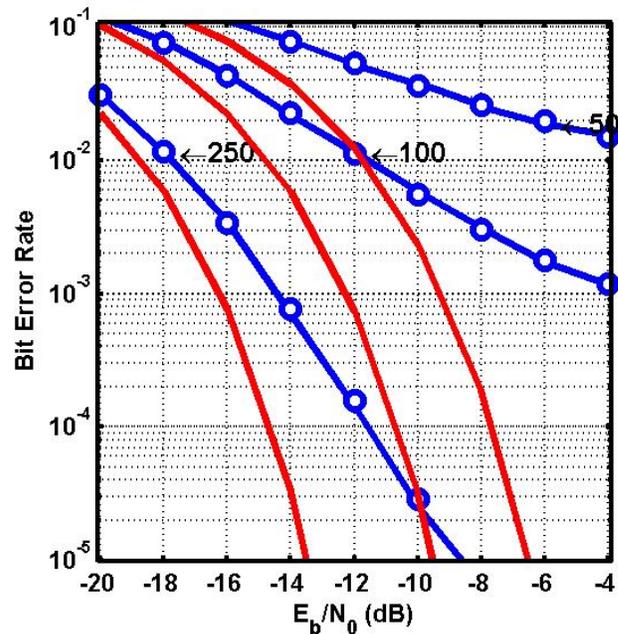


Figure 6. BER performances for OFDM-based MU-MIMO with  $N_T = 10$ , and  $N_R = 50, 100$  or  $250$ , under MF linear detection [SIMO  $1 \times N_R$  (single-path) reference performances are also included].

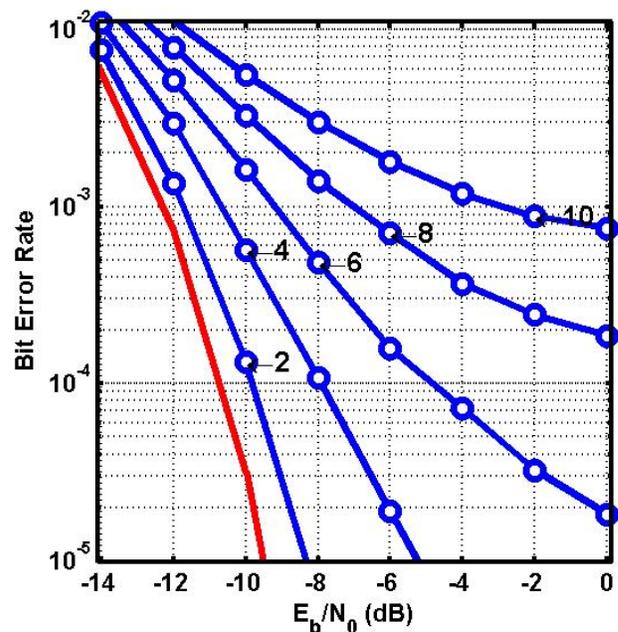


Figure 7. BER performances for OFDM-based MU-MIMO with  $N_R = 100$ , and  $N_T = 2, 4, 6, 8$  or  $10$ , under MF linear detection [SIMO  $1 \times N_R$  (single-path) reference performance is also included].

Figure 7 emphasizes the more or less acceptable performance degradation levels which are unavoidable when  $N_R$  is kept fixed and  $N_T$  increases (under the reduced-complexity detection in all cases). This set of figures emphasizes a "massive MIMO" effect when  $N_R \gg 1$ , especially when  $N_R \gg N_T$  too, which leads to BER performances very close to the ultimate "MUI-free and fading-free" SIMO (single-path) performance

bound.

## V. CONCLUSIONS

This paper was dedicated to the uplink performance evaluation of a MU-MIMO system with OFDM transmission, when adopting a large number of antennas and linear detection techniques at the BS. The accuracy of performance results obtained by semi-analytical means, as proposed in Section III, was demonstrated.

The numerical performance results, discussed in detail in Section IV, show the "massive MIMO" effects provided by a number of BS antennas much higher than the number of antennas which are jointly employed in the terminals of the multiple autonomous users, even when a reduced-complexity (MF) linear detection technique is adopted.

## REFERENCES

- [1] A.J. Paulraj, D.A. Gore, R.U. Nabar, and H. Bolcskei. An overview of mimo communications - a key to gigabit wireless. *Proceedings of the IEEE*, 92(2):198–218, Feb 2004.
- [2] G. Foschini. "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when Using Multi-element Antennas". *Bell Labs Technology Journal*, (2):1, 1996.
- [3] J. Mietzner, R. Schober, L. Lampe, W.H. Gerstacker, and P.A. Hoeher. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *Communications Surveys Tutorials, IEEE*, 11(2):87–105, Second 2009.
- [4] E.G. Larsson. Mimo detection methods: How they work [lecture notes]. *Signal Processing Magazine, IEEE*, 26(3):91–95, May 2009.
- [5] D. Gesbert, M. Kountouris, R.W. Heath, Chan-Byoung Chae, and T. Salzer. Shifting the mimo paradigm. *Signal Processing Magazine, IEEE*, 24(5):36–46, Sept 2007.
- [6] T.L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *Wireless Communications, IEEE Transactions on*, 9(11):3590–3600, November 2010.
- [7] F. Rusek, D. Persson, Buon Kiong Lau, E.G. Larsson, T.L. Marzetta, O. Edfors, and F. Tufvesson. Scaling up mimo: Opportunities and challenges with very large arrays. *Signal Processing Magazine, IEEE*, 30(1):40–60, Jan 2013.
- [8] Hien Quoc Ngo, E.G. Larsson, and T.L. Marzetta. Energy and spectral efficiency of very large multiuser mimo systems. *Communications, IEEE Transactions on*, 61(4):1436–1449, April 2013.
- [9] J. Hoydis, S. ten Brink, and M. Debbah. Massive mimo in the ul/dl of cellular networks: How many antennas do we need? *Selected Areas in Communications, IEEE Journal on*, 31(2):160–171, February 2013.
- [10] H. Cramer. "Random Variables and Probability Distributions". *Cambridge University Press*, 1970.

# Evaluative Study of Detect-Split-Forward Scheme over MIMO Relays

Suhaib M. Al-Basit, Samir Al-Ghadhban and Salam A.Zummo

Department of Electrical Engineering  
King Fahd University of Petroleum & Minerals  
Dhahran, Saudi Arabia  
Email: {suhaib, samir, zummo}@kfupm.edu.sa

**Abstract**—Virtual multi-input multi-output (vMIMO) schemes in wireless communication systems improve coverage, throughput, capacity, and quality of service. In this paper, we propose two uplink vMIMO relaying schemes based on detect-split-and-forward (DSF). In addition, we investigate the effect of several physical parameters on the performance of the relaying systems, such as distance, modulation type and number of relays. In addition, we provide analytical tools to evaluate the performance of the proposed vMIMO relaying schemes.

**Keywords**—Virtual MIMO, distributed antenna, relaying, detect-split-and-forward, STBC, V-BLAST.

## I. INTRODUCTION

Since the introduction of the multiple-input-multiple-output (MIMO) technology, there have been great advancements in data rate speeds and wireless network efficiency. The main purpose of MIMO implementation is to boost the transmission rate by exploiting the randomness of parallel channels. Using MIMO technology, the capacity of a propagation environment decreases with increasing the correlation of the channel coefficients. Practically, for none-line-of-sight (NLOS) and omni-directional wireless mobile communications, there are restrictions on handset manufacturing caused by wavelenghts. Hence, the designers should select applicable wavelengths to realize the full potential of MIMO receivers. Obviously, large antenna arrays of different sizes are not always practical for handsets mobile. Furthermore, the developers of the next generation wireless systems are investing in virtual MIMO (vMIMO) [1]. Virtual MIMO (vMIMO) is a recent model adapted from the broadcasting model of wireless channels where all communication nodes (relays) support each other. The goal of vMIMO is to provide better quality-of-service (QoS) at higher data rates, especially for users who are at the cell edge. This technique acts in a way similar to multi-user (MU-MIMO) technique in the uplink side, also called network MIMO [1]. vMIMO is based on the concept of relaying over virtual antenna arrays, with results in enhancing the end-to-end link performance, offering good QoS and extending coverage range in NLOS environment. vMIMO systems execute the communication process in a distributed manner to take advantages of the MIMO system while reducing consumption of battery, improving capacity and expanding network lifetime [2].

Relay structures have evolved by the introduction of virtual antenna arrays and MIMO relays [3] and [4]. The relaying

technique, as introduced by Van der Meulen [5], has transpired through the years as the most well-known approach to improve the reliability and performance of wireless networks. It makes use of node cooperation and it allows a network to extend its coverage without exhausting its power resources. The two well-known relaying protocols, are: amplify-and forward (AF) and decode-and forward (DF) [6]. AF is the simplest as it only amplifies the received signal then forwards it. However, its drawback is that it amplifies the noise in addition to the signal. This technique, as described in the IEEE802.16j [7] standard, does not require the mobile station (MS) to be aware of intermediate relays. On the other hand, DF is a protocol that uses error detection and correction as it decodes data once received and confirms its correctness then forwards the data. This technique is generally used with hybrid automatic repeat request (HARQ) to ensure that correct data was decoded and intact [6].

The authors of [8] and [9] extended relaying concepts to MIMO by considering several relay transmission and topology schemes (e.g., parallel, serial and hybrid) and taking into consideration practical MIMO systems. The authors of [10] and [11] investigated the performance of space-time block coding (STBC) with MIMO relaying using AF as an effective way to introduce spatial diversity. Virtual spatial multiplexing with AF relaying was investigated in [12] and closed-form expressions were derived at high-SNRs. In [13], IEEE802.16e described the uplink virtual MIMO (UL-vMIMO) as follows: each user is equipped with a single antenna and shares the same channel resources with other users. By utilizing simultaneous transmissions over a common burst, vMIMO increases the peak transmission rate and improves the system performance. In addition, DF virtual relaying scheme for MIMO systems was analyzed in [14].

In [15], an AF virtual spatial multiplexing scheme is proposed in which each transmitter is equipped with a single antenna. The transmitters form a virtual antenna array and send identical signals to relays that amplify-and-forward different portions of the signal at a reduced data rate to the destination. The receiver is equipped with multiple antennas in order to null and cancel the interference from the different relays and detect the original signal transmitted from the source. Another approach proposed by [16] is to let the relays detect a sub-stream from the original stream. Then, all relays forward their low rate sub-streams simultaneously over the same physical channel. This scheme has the advantage of controlling noise, as

in digital systems, so that it is not amplified. Another advantage is that the vMIMO relay can send data with lower modulation rates which improves the bit error rate (BER). Because of the practical difficulty associated with antenna coupling, another alternative technique is proposed in [16]. In this paper we refer to the scheme of [16] as detect-split-forward (DSF) using vertical-bell laboratories layered space-time (VBLAST).

Up to the authors knowledge, there is no analysis for the error probability of DSF schemes over vMIMO channels. Thus, in this paper we provide analytical tools to evaluate the performance of DSF-vMIMO schemes, for both VBLAST and STBC. Both simulations and analysis are conducted to evaluate the system performance in terms of several physical parameters such as distance, modulation type, and number of relays.

The remainder of this paper is organized as follows. Section II introduces the system and channel models. The analysis of system performance and the average capacity are conducted in Section III and Section IV. In Section V, some simulation and numerical results are presented and discussed. Finally, conclusions are given in Section VI.

## II. SYSTEM MODEL

We consider a  $1 \times N_R \times N_D$  uplink system, where 1 indicates a source with a single antenna,  $N_R$  is number of relays, each equipped with a single antenna.  $N_D$  is number of receiving antennas at the destination. The source modulates a block of  $k$  information bits and transmits an  $2^k$ -ary modulated symbol  $x$ , which is received by all relays. Then, each relay detects the information bits and splits them into  $N_R$  blocks of length  $m$  bits, where  $mN_R = k$ . At each relay,  $m$ -bit block is modulated using lower level modulation schemes ( $2^m$ -ary symbol) and will be transmitted through  $N_R$  relays, which creates a vMIMO system. At the relays, two vMIMO schemes are considered for relaying the  $N_R$  symbols. The first scheme uses spatial multiplexing based on VBLAST and the second scheme uses STBC.

### A. DSF-vMIMO Schemes

To illustrate DSF vMIMO schemes, consider a  $1 \times 2 \times 2$  system with 2 bps/Hz efficiency. The relays receive the following signals from the source:

$$Y_{R_1} = h_{SR_1}x + n_{R_1} \quad (1)$$

$$Y_{R_2} = h_{SR_2}x + n_{R_2}, \quad (2)$$

where  $x$  is the transmitted symbol,  $Y_{R_i}$  is the received signal at relay  $i$ , and  $h_{SR_i}$  is the complex Gaussian fading coefficient, with zero mean and variance of one, from the source to relay  $i$ . In addition,  $n_{R_i}$  is the additive white Gaussian noise at relay  $i$  with zero-mean and variance  $N_o$ , where  $N_o$  is the noise power spectral density. Each relay detects the transmitted symbol  $x$ . The detection, splitting and forwarding of each vMIMO scheme is explained next.

1) *DSF-VBLAST Scheme*: The DSF-VBLAST scheme detects the  $2^k$ -ary symbol, splits it into  $N_R$  parallel symbols, where each symbol carries  $m$  bits and then forwards the signals simultaneously to the destination as shown in Figure 1. As an example, for a 2 bps/Hz efficiency, the source modulates

4 bits using 16-QAM and sends it to the relays. Each relay detects the 16-QAM symbol and demodulates the 4 bits. Then, these 4 bits will be splitted into two blocks, each consisting of two bits. These two bits will be modulated using QPSK and then spatially multiplexed and forwarded to the destination. Since 4 bits have been transmitted using two hops, the system efficiency is 2 bps/Hz.

The destination receives the following vMIMO signals:

$$\begin{pmatrix} Y_{D_1} \\ Y_{D_2} \end{pmatrix} = \begin{pmatrix} h_{R_1D_1} & h_{R_2D_1} \\ h_{R_2D_2} & h_{R_1D_2} \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} + \begin{pmatrix} n_{D_1} \\ n_{D_2} \end{pmatrix}, \quad (3)$$

where  $Y_{D_j}$  is the received signal at antenna  $j$  at the destination,  $\hat{x}_i$  is the transmitted signal from relay  $i$ , and  $h_{R_iD_j}$  is the complex Gaussian fading coefficient, with zero mean and unit variance, between relay  $i$  and antenna  $j$  at the destination. In addition,  $n_{D_j}$  is the additive white Gaussian noise at antenna  $j$  at the destination.

The destination applies the VBLAST detection algorithm with successive interference cancellation (SIC) and it performs nulling, based on zero-forcing (ZF), cancelation and ordering.

2) *DSF-STBC Scheme*: To improve the diversity order and provide more link reliability, STBC is used in the second hop. At the relays, the  $2^k$ -ary symbol is detected and split into  $N_R$  symbols, each carries  $k/N_R$  bits. After that, the splitted symbols are mapped to an STBC as shown in Figure 2. As an example, for a 2 bps/Hz, the source sends 6 bits using a 64-QAM symbol to the two relays. Then, each relay detects the 64-QAM symbol and demodulates the 6 bits and split them into two blocks, each consists of 3 bits. Each block of 3 bits is modulated using 8PSK and then mapped to Alamouti STBC code. The STBC codes will be forwarded to the destination over two time slots. Since 6 bits have been transmitted during three time slots, the system efficiency is 2 bps/Hz. The signals received by the destination could be expressed as:

$$\begin{pmatrix} Y_{D_1}^2 \\ Y_{D_2}^2 \\ Y_{D_1}^{3*} \\ Y_{D_2}^{3*} \end{pmatrix} = \begin{pmatrix} h_{R_1D_1}^2 & h_{R_1D_2}^2 \\ h_{R_2D_1}^2 & h_{R_2D_2}^2 \\ h_{R_1D_1}^{3*} & -h_{R_1D_2}^{3*} \\ h_{R_2D_1}^{3*} & -h_{R_2D_2}^{3*} \end{pmatrix} \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} + \begin{pmatrix} n_{D_1}^2 \\ n_{D_2}^2 \\ n_{D_1}^{3*} \\ n_{D_2}^{3*} \end{pmatrix}, \quad (4)$$

where  $Y_{D_j}^t$  is the received signal at time slot  $t$  and antenna  $j$ ,  $h_{R_iD_j}^t$  is the channel from relay  $j$  to antenna  $j$  at the destination at time slot  $t$ . The MIMO channel from the relay to the destination is assumed to be quasi-static complex Gaussian channel, where each coefficient has a zero mean and unit variance.

## III. PERFORMANCE ANALYSIS

### A. DSF-VBLAST Scheme

To calculate the block error rate (BLER) for the DSF VBLAST scheme, we analyze the detection process at each hop. Assume that  $P_{B,H_1}$  is the BLER of the first hop and  $P_{B,H_2}$  is the BLER of the second hop. Then the total BLER at the destination will be:

$$P_B = 1 - (1 - P_{B,H_1})(1 - P_{B,H_2}), \quad (5)$$

Since in the first hop, the source transmit the modulated symbol to  $N_R$  single antenna relays, the BLER at the first hop is

$$P_{B,H_1} = 1 - (1 - P_{e,R})^{N_R}. \quad (6)$$

Where  $P_{e,R}$  is the symbol error rate (SER) at each relay for M-ary modulation over fading channels [17], taking into account the appropriate signal set levels and energy after splitting. Thus SER at each relay is:

$$P_{e,R} = \left( \frac{M-1}{M} \right) \left( 1 - \sqrt{\frac{1.5\gamma_R}{M^2 - 1 - 3\gamma_R}} \right), \quad (7)$$

where  $M = 2^m$  is the new cardinality of the signal set after splitting and  $\gamma_R$  denotes the average received SNR at each relay.

Since VBLAST detection is used in the second hop, each layer of the V-BLAST scheme has a different error probability depending on its diversity order [18].

Assume that  $P_{e,i}$  is the SER for layer  $i$  over Rayleigh fading channels, then the BLER of the second hop is:

$$P_{B,H_2} = 1 - \prod_{i=1}^{N_R} (1 - P_{e,i}), \quad (8)$$

For M-QAM signals,  $P_{e,i}$  is [19]:

$$\begin{aligned} P_{e,i} = & 4\left(1 - \frac{1}{\sqrt{M}}\right)\left(\frac{1-\zeta_i}{2}\right)^{D_i} \sum_{j=0}^{D_i-1} \binom{D_i-1+j}{j} \left(\frac{1+\zeta_i}{2}\right)^j \\ & - 4\left(1 - \frac{1}{\sqrt{M}}\right)^2 \left\{ \frac{1}{4} - \frac{\zeta_i}{\Pi} \left\{ \left(\frac{\Pi}{2} - \tan^{-1}\zeta_i\right) \sum_{j=0}^{D_i-1} \binom{2j}{j} \left(\frac{1-\zeta_i}{4}\right)^j \right. \right. \\ & \left. \left. + \sin(\tan^{-1}\zeta_i) \sum_{j=1}^{D_i-1} \sum_{r=1}^j \frac{J_{rj}}{(1+\beta_i)^j} [\cos(\tan^{-1}\zeta_i)]^{2(j-r)+1} \right\} \right\}, \end{aligned} \quad (9)$$

where the diversity order of layer  $i$  is:

$$D_i = N_D - N_R + i, \quad (10)$$

and  $N_D$  is the total number of receiver antenna at the destination. Let  $d$  be the distance between the relays and the source, and  $v$  be the path loss, then the parameters in (9) are defined as

$$\zeta_i = \frac{\beta_i}{1 + \beta_i} \quad (11)$$

$$\beta_i = \frac{3d^{-v}\gamma_D}{2N_R(M-1)} \quad (12)$$

$$J_{rj} = \frac{\binom{2j}{j}}{\binom{2(j-r)}{j-r} 4^i (2(j-r)+1)}. \quad (13)$$

where  $\gamma_D$  denotes the average received SNR at each receive antenna at the destination.

For M-PSK case,  $P_{e,i}$  is [19]:

$$P_{e,i} = \frac{M-1}{M} - \frac{\mu_t}{\sqrt{\mu_t^2 + 1}} \left( \frac{1}{2} + \frac{\omega_t}{\Pi} \right) \sum_{\tau=0}^{D_t-1} \binom{2\tau}{\tau} [4(\mu_t^2 + 1)]^{-\tau}$$

$$- \frac{\mu_t}{\sqrt{\mu_t^2 + 1}} \frac{1}{\Pi} \sin(\omega_t) \sum_{\tau=1}^{D_t-1} \sum_{i=1}^{\tau} \frac{J_{i\tau}}{(\mu_t^2 + 1)^\tau} [\cos(\omega_t)]^{2(\tau-i)+1}, \quad (14)$$

where

$$\mu_t = \sqrt{\rho_t} \sin\left(\frac{\Pi}{M}\right) \quad (15)$$

$$\rho_t = d^{-v} \cdot \gamma_D \quad (16)$$

$$\omega_t = \tan^{-1}\left(\frac{\sqrt{\rho_t} \cos\left(\frac{\Pi}{M}\right)}{\sqrt{\mu_t^2 + 1}}\right). \quad (17)$$

## B. DSF-STBC Scheme

For M-QAM STBC, the SEP at the second hop could be calculated using (9) with a diversity order  $N_D \cdot N_R$  and with one layer ( $L = 1$ ). Therefore, the second hop BLER is  $P_{B,H_2} = P_{e,1}$  with a diversity  $D = N_D \cdot N_R$ . Then the total BLER  $P_B$  is calculated using (5).

## IV. CAPACITY ANALYSIS

The DSF-vMIMO capacity is presented in this section. The analysis of the capacity of DSF-vMIMO for V-BLAST and STBC is based on the fact that the instantaneous capacity of the two hop relay system is determined by the weakest link. The first hop consists of two SISO channels and the second hop consists of a MIMO channel. The second MIMO hop can be either V-BLAST or STBC. Therefore, the instantaneous capacity of DSF-VBLAST can be calculated as:

$$C_{DSF-VBLAST} = \frac{\min\{C_{SISO1}, C_{SISO2}, C_{VBLAST}^{ZF}\}}{N_H}, \quad (18)$$

and the capacity of DSF STBC is:

$$C_{DSF-STBC} = \frac{\min\{C_{SISO1}, C_{SISO2}, C_{STBC}\}}{N_H}, \quad (19)$$

where  $N_H$  is the total number of hops.

In the above equations,  $C_{SISO}$  is the instantaneous capacity of single-input single-output flat Rayleigh fading channels [18] at the first hop, which is expressed as:

$$C_{SISO} = \log_2(1 + \gamma_R |h|^2) \text{ bps/Hz}, \quad (20)$$

where the channel coefficient  $h$  is a complex Gaussian random variable with zero mean and unit variance, and  $\gamma_R$  is the average SNR at each relay. For a given  $h$ , there is only one way to increase the capacity of the SISO channel and that is by increasing SNR. Also, the capacity increases logarithmically with increasing SNR.

The second hop is either a VBLAST or an STBC. For VBLAST, the instantaneous capacity with  $N_R$  relays and with ZF interference nulling and serial cancelation is given by [20]:

$$C_{VBLAST}^{ZF} = N_R \cdot \min_{j=1,2,\dots,N_R} \left\{ \log_2 \left( 1 + \frac{\gamma_D}{N_R \|W_{ZF,j}\|_F^2} \right) \right\}, \quad (21)$$

where  $W_{ZF,j}$  is the ZF projection vector of the  $j^{\text{th}}$  relay,  $\gamma_D$  is the SNR per receive antenna at the destination, and  $\|\cdot\|_F^2$  is the squared Frobenius norm.

For STBC, the instantaneous capacity rate  $r_c$  code and  $N_R$  relays is [21]

$$C_{STBC} = r_c \left\{ \log_2 \left( 1 + \frac{\gamma_D}{N_R} \|\mathbf{H}\|_F^2 \right) \right\}. \quad (22)$$

## V. NUMERICAL RESULTS

In this section, the performance of DSF V-BLAST and DSF-STBC schemes are evaluated using the analytical results and verified by Monte-Carlo simulations. The channel models used in the simulation are Rayleigh flat-fading channels characterized by complex Gaussian random variables with zero-mean and 0.5 variance per dimension. In addition additive white Gaussian noise is added at each receive antenna. Both BLER performance and average capacity are evaluated. The spectral efficiency is computed as the total number of bits received at the destination divided by the total number of transmission time slots.

The BLER performance of  $1 \times 2 \times 2$  DSF-VBLAST and DSF-STBC are shown in Fig. 3 and 4, respectively. The performance is examined at several spectral efficiencies as shown in Tables I and II. The relays are placed at a normalized distance of 0.3 from the source. The results in this figure illustrate the accuracy of our analysis.

BLER comparison between DSF-VBLAST and DSF-STBC schemes is shown in Fig. 5. The result shows the effect of relay location on the performance of both systems and the inherent tradeoffs. The performance is examined at low, medium, and high SNRs and the normalized distance is changed from zero to one. The result shows that when the relays are closer to the source, the DSF-STBC performs better than DSF-V-BLAST. However, when the relays are placed further than 0.4, the performance of DSF-VBLAST becomes better. From this result, we propose to design a hybrid system where the relays use adaptive techniques to determine the best scheme to be used based on the distance from the source.

The average capacities of DSF-VBLAST and DSF-STBC schemes are shown in Fig. 6. At a certain distance and since both systems consist of SISO and MIMO hops, the SISO channel will dominate the overall channel capacity. The result in Fig. 6 shows that at distances greater than  $d = 0.5$ , the DSF-VBLAST channel capacity will perform the same as the DSF-STBC since both systems are dominated by the weakest channel, which is the SISO channel at the first hop.

## VI. CONCLUSION

In this paper, we analyzed detect-split-forward uplink virtual MIMO relaying schemes based on VBLAST and STBC. The relays split the incoming source signal into lower modulation levels. This will enhance the energy efficiency of the relaying system. Block error rate expressions and outage capacity rates were presented. The analysis presented in this paper matched the simulation results and it showed the effect of several physical parameters such as distance, modulation type and number of relays. For future work, there exists a tradeoff between VBLAST and STBC which leads to designing an adaptive relaying system that can adapt the MIMO scheme based on the location of the relay to satisfy a certain quality of service.

## ACKNOWLEDGMENT

The authors would like to acknowledge the support provided by King Fahd University of Petroleum and Minerals (KFUPM) and King Abdulaziz City for Science and Technology (KACST) through the Science and Technology Unit at KFUPM for funding this work through project number 09-ELE781-4 as part of the National Science, Technology and Innovation Plan.

TABLE I.  $1 \times 2 \times 2$  DSF-VBLAST

Time Slots	1 <sup>st</sup> Hop	2 <sup>nd</sup> Hop	bps/Hz
2	QPSK	BPSK	1
2	16-QAM	QPSK	2
2	64-QAM	8PSK	3
2	256-QAM	16-QAM	4

TABLE II.  $1 \times 2 \times 2$  DSF-STBC

Time Slots	1 <sup>st</sup> Hop	2 <sup>nd</sup> Hop	bps/Hz
3	QPSK	BPSK	0.66
3	16-QAM	QPSK	1.33
3	64-QAM	8PSK	2
3	256-QAM	16-QAM	2.66

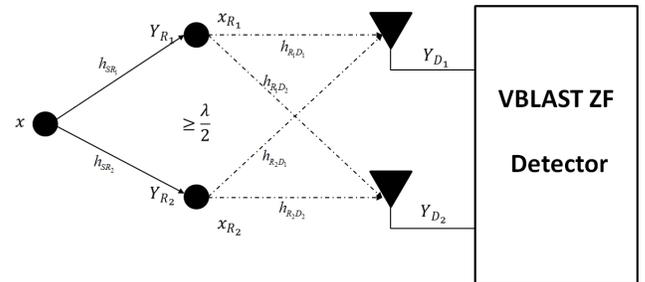


Figure 1. System model of  $1 \times 2 \times 2$  DSF-VBLAST relaying scheme.

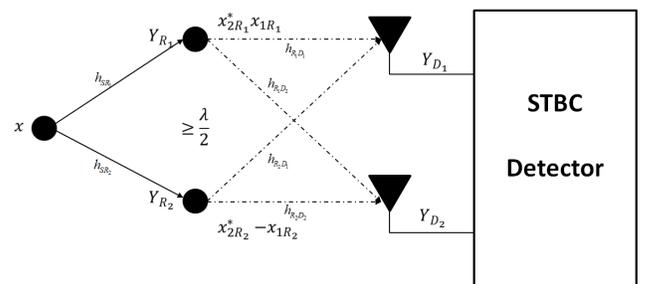


Figure 2. System model of  $1 \times 2 \times 2$  DSF-STBC relaying scheme.

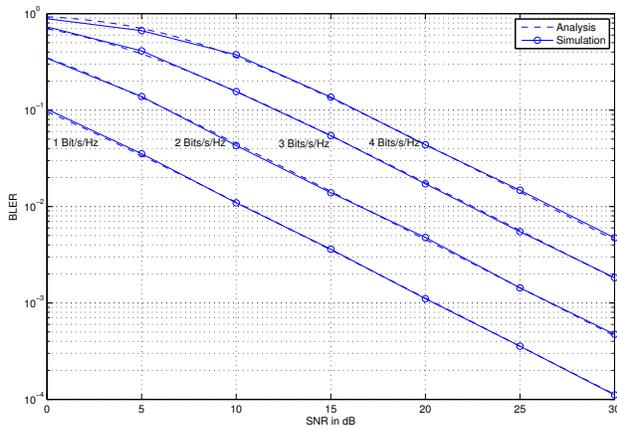


Figure 3. BLER performance comparison of DSF-VBLAST relaying schemes for  $1 \times 2 \times 2$  vMIMO system.

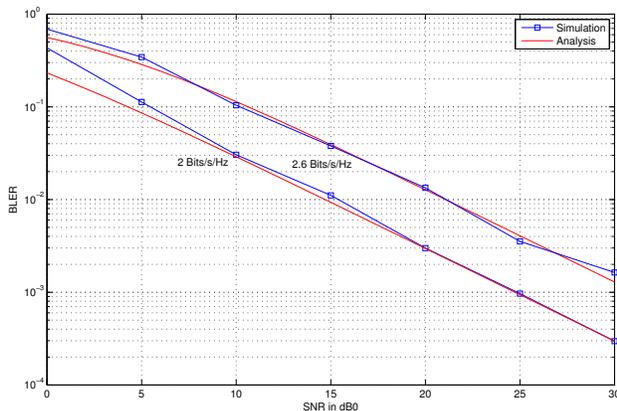


Figure 4. BLER performance of DSF-STBC relaying schemes at 0.3 normalized distance from the source.

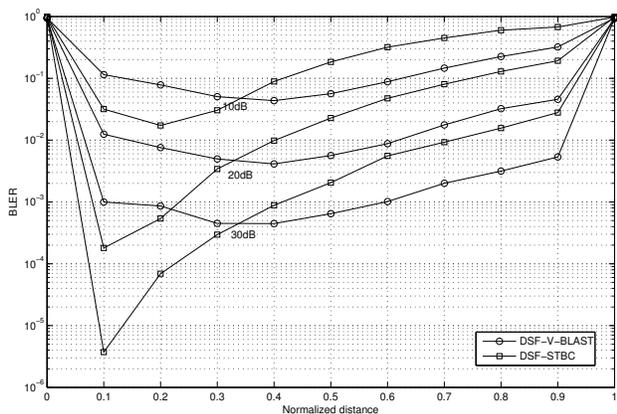


Figure 5. Effect of the source-relay distance on the BLER performance of DSF-vMIMO relaying schemes.

REFERENCES

[1] M. Rumney, LTE and the Evolution to 4G Wireless: Design and Measurement Challenges. Wiley, 2009.

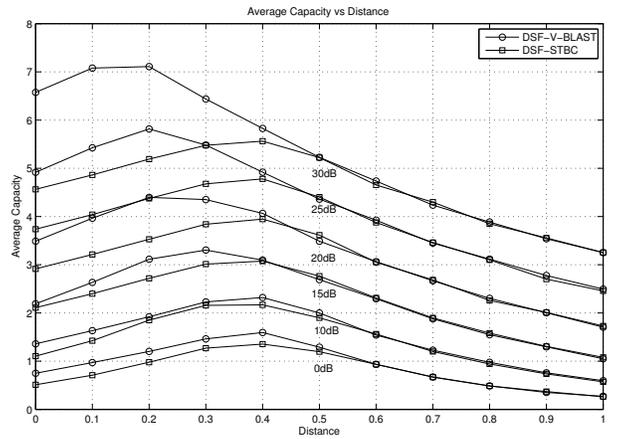


Figure 6. Average capacity of DSF-vMIMO relaying schemes at different source-relay distances.

[2] K. J. Rayliu, A. K. Sadek, and A. Kwasinski, Cooperative Communications and Networking. Cambridge University Press, 2009.

[3] M. Dohler, A. Gkelias, and H. Aghvami, "A resource allocation strategy for distributed MIMO multi-hop communication systems," IEEE Communications Letters, vol. 8, Feb. 2004, pp. 99–101.

[4] B. Wang, J. Zhang, and A. Host-Madsen, "On the capacity of MIMO relay channels," IEEE Transactions on Information Theory, vol. 51, Jan 2005, pp. 29–43.

[5] E. C. van der Meulen, "Three-terminal Communication Channels," Journal of Advances in Applied Probability, vol. 3, June 1971, pp. 120–154.

[6] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," IEEE Transactions on Information Theory, vol. 50, Dec. 2004, pp. 3062–3080.

[7] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems Amendment 1: Multiple Relay Specification. IEEE Std 802.16j-2009 (Amendment to IEEE Std 802.16-2009), 2009.

[8] Y. Kim and H. Liu, "Infrastructure Relay Transmission with Cooperative MIMO," IEEE Transactions on Vehicular Technology, vol. 57, Jul. 2008, pp. 2180–2188.

[9] A. Nessa, Q. Yang, and K. S. Kwak, "Performance analysis of dual-hop cooperative MIMO transmission with relay selection in Rayleigh fading channel," Int. Arab J. Inf. Technol., vol. 8, 2011, pp. 2180–2188.

[10] A. Abdaoui, S. S. Ikki, M. H. Ahmed, E., and Chatelet, "On the Performance Analysis of a MIMO-Relaying Scheme With Space Time Block Codes," IEEE Transactions on Vehicular Technology, vol. 59, Sep. 2010, pp. 3604–3609.

[11] P. Huo and L. Cao, "Distributed STBC With Soft Information Relay Based on Gaussian Approximation," IEEE Signal Processing Letters, vol. 19, Oct. 2012, pp. 599–602.

[12] C. Song, K. Lee, and I. Lee, "Performance Analysis of Amplify-And-Forward Spatial Multiplexing MIMO Relaying Systems," The IEEE International Conference on Communications, May 2010, pp. 1–5.

[13] X. Zhu, Y. Song, H. Yang, and L. Cai, "2-D Switching Diversity Aided Collaborative Spatial Multiplexing for Uplink Wireless Access," IEEE Wireless Communications and Networking Conference, Apr. 2009, pp. 1–4.

[14] H. V. Khuong and T. Le-Ngoc, "Performance analysis of a decode-and-forward cooperative relaying scheme for MIMO systems," The 25th Biennial Symposium on Communications, May 2010, pp. 400–403.

[15] A. Darmawan, S. Kim, and H. Morikawa, "Amplify-and-Forward Scheme in Cooperative Spatial Multiplexing," Mobile and Wireless Communications Summit, Jul. 2007, pp. 1–7.

[16] S. Kim and R. Cherukuri, "Cooperative spatial multiplexing for high-

- rate wireless communications,” IEEE Workshop Signal Processing Advances in Wireless Communications, Jun. 2005, pp. 181–185.
- [17] M. K. Simon and M. S. Alouini, Digital Communication over Fading Channels. Wiley-IEEE Press, 2004.
- [18] S. Al-Ghadhban, Multi-layered Space Frequency Time Codes. Phd Dissertation at Virginia Polytechnic Institute and State University, Dec. 2005.
- [19] F. Al-Shalan, Performance of Quadrature Amplitude Modulation in Nakagami Fading Channels with Diversity. Thesis at King Fahd University of Petroleum and Minerals, Mar. 2000.
- [20] C. Papadias and G. J. Foschini, “On the Capacity of Certain Space-Time Coding Schemes,” EURASIP Journal on Applied Signal Processing, vol. 5, May 2002, pp. 447–458.
- [21] S. Sandhu and A. Paulraj, “Space-Time Block Codes: A Capacity Perspective,” IEEE Comm. Letters, vol. 4, Dec. 2000, pp. 384–386.

# On the SC/FDE Uplink Alternative to OFDM in a Massive MU-MIMO Context

Paulo Torres

EST - Instituto Politecnico de Castelo Branco  
Castelo Branco, Portugal  
Email: paulo.torres@ipcb.pt

Luis Charrua<sup>†</sup>, Antonio Gusmao<sup>‡</sup>

CAPS - Instituto Superior Tecnico, Universidade de Lisboa  
Lisboa, Portugal  
Email: <sup>†</sup>luischarrua@enautica.pt, <sup>‡</sup>gus@ist.utl.pt

**Abstract**—This paper deals with Single Carrier (SC)/Frequency Domain Equalization (FDE) as an uplink alternative to Orthogonal Frequency Division Multiplexing (OFDM) for a Multi User (MU)-Multi-Input Multi-Output (MIMO) system where a "massive MIMO" approach is adopted. In this context, either an optimum Minimum Mean-Squared Error (MMSE) linear detector or appropriate reduced-complexity linear detection techniques are considered. Regarding performance evaluation by simulation, two semi-analytical methods are proposed - one method in the optimum (MMSE) case and the other one in the reduced-complexity cases. This paper includes performance results for uncoded 4-Quadrature Amplitude Modulation (QAM) SC/FDE transmission and a MU-MIMO channel with uncorrelated Rayleigh fading, under the assumptions of perfect power control and perfect channel estimation. The accuracy of performance results obtained through the semi-analytical simulation methods is assessed by means of parallel conventional Monte Carlo simulations. The performance results are discussed in detail and we also emphasize the achievable "massive MIMO" effects, even for the reduced-complexity detection techniques, provided that the number of BS antennas is much higher than the number of antennas which are jointly employed in the terminals of the multiple autonomous users. Appropriate "SC/FDE vs OFDM" comparisons are also included in this discussion of performance results.

**Keywords**-Broadband Wireless Communications; MU-MIMO Systems; Massive MIMO; Performance Evaluation; SC/FDE.

## I. INTRODUCTION

Cyclic Prefix (CP)-assisted block transmission schemes were proposed and developed, in the last two decades, for broadband wireless systems, which have to deal with strongly frequency-selective fading channel conditions. These schemes take advantage of current low-cost, flexible, Fast Fourier Transform (FFT)-based signal processing technology, with both OFDM and SC/FDE alternative choices [1][2][3]. Mixed air interface solutions, with OFDM for the downlink and SC/FDE for the uplink, as proposed in [2], are now widely accepted; the main reason for replacing OFDM by SC/FDE, with regard to uplink transmission, is the lower envelope fluctuation of the transmitted signals when data symbols are directly defined in the time domain, leading to reduced power amplification problems at the mobile terminals.

Also in the last two decades, the development of MIMO technologies has been crucial for the "success story" of broadband wireless communications. Through spatial multiplexing schemes, early introduced by Foschini [4], and appropriate

MIMO detection schemes [5][6] - offering a range of performance/complexity tradeoffs -, MIMO systems are currently able to provide very high bandwidth efficiencies and a reliable radiotransmission at very high data rates. In the last decade, MU-MIMO systems [7] - able to serve multiple autonomous users in the same time-frequency resource, thereby providing a true "space division multiple access" - also have been successfully implemented and introduced in several broadband communication standards.

In recent years, the adoption of a very large number of antennas in the BS, much larger than the number of Mobile Terminal (MT) antennas in its cell, was proposed in [8][9]. This "massive MIMO" approach has been shown to be recommendable for several reasons [8][9]: simple linear processing for MIMO detection becomes nearly optimal; both MultiUser Interference (MUI) effects and fast fading effects of multipath propagation tend to disappear; both power efficiency and bandwidth efficiency become substantially increased.

This paper deals with SC/FDE as an uplink alternative to OFDM [10] for a MU-MIMO system where the BS is constrained to adopt simple, linear detection techniques, but can be equipped with a large number of receiver antennas. In this context, either an optimum (MMSE) linear detector or appropriate reduced-complexity, Matched Filter (MF)-based, linear detection techniques are considered in Section II. Regarding performance evaluation by simulation, two semi-analytical methods are proposed in Section III - one method in the optimum (MMSE) case and the other one in the reduced-complexity cases -, both combining simulated channel realizations and analytical computations of BER performance which are conditional on those channel realizations.

In Section IV, this paper includes performance results for uncoded 4-QAM SC/FDE transmission and a MU-MIMO channel with uncorrelated Rayleigh fading effects regarding the several transmitter/receiver (TX/RX) antenna pairs, under the assumptions of perfect power control and perfect channel estimation. The accuracy of performance results obtained through the semi-analytical simulation methods is assessed by means of parallel conventional Monte Carlo simulations (involving an error counting procedure). The performance results are discussed in detail and we also emphasize the achievable "massive MIMO" effects, even for the reduced-complexity detection techniques, provided that the number of BS antennas is much higher than the number of antennas which are jointly employed in the terminals of the multiple

autonomous users. Appropriate "SC/FDE vs OFDM" comparisons are also included in this discussion of performance results. Section V includes the main conclusions of the paper.

## II. SYSTEM MODEL

### A. SC/FDE-based Radiotransmission

We consider here a CP-assisted, SC/FDE-based, block transmission, within a MU-MIMO system with  $N_T$  TX antennas and  $N_R$  RX antennas - for example (but not necessarily) one antenna per MT. We assume, in the  $j$ th TX antenna ( $j = 1, 2, \dots, N_T$ ), a length- $N$  block  $\mathbf{s}^{(j)} = [s_0^{(j)}, s_1^{(j)}, \dots, s_{N-1}^{(j)}]^T$  of time-domain data symbols in accordance with the corresponding binary data block. The insertion of a length- $L_s$  CP, long enough to cope with the time-dispersive effects of multipath propagation, is also assumed.

The time-domain data symbols  $s_n^{(j)}$  ( $n = 0, 1, \dots, N - 1$ ;  $j = 1, 2, \dots, N_T$ ) are randomly and independently selected from a QAM alphabet ( $E[s_n^{(j)}] = 0$  and  $E[|s_n^{(j)}|^2] = \sigma_s^2$ ) for any  $(j, n)$ .

By using the frequency-domain version of the time-domain data blocks  $\mathbf{s}^{(j)} = [s_0^{(j)}, s_1^{(j)}, \dots, s_{N-1}^{(j)}]^T$ , given by  $\mathbf{S}^{(j)} = [S_0^{(j)}, S_1^{(j)}, \dots, S_{N-1}^{(j)}]^T = DFT(\mathbf{s}^{(j)})$  ( $j = 1, 2, \dots, N_T$ ), we can describe the frequency-domain transmission rule as follows, for any subchannel  $k$ :

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{S}_k + \mathbf{N}_k, \quad (1)$$

where  $\mathbf{S}_k = [S_k^{(1)}, S_k^{(2)}, \dots, S_k^{(N_T)}]^T$  is the "input vector",  $\mathbf{N}_k = [N_k^{(1)}, N_k^{(2)}, \dots, N_k^{(N_R)}]^T$  is the Gaussian noise vector ( $E[N_k^{(i)}] = 0$  and  $E[|N_k^{(i)}|^2] = \sigma_N^2 = N_0 N$ ),  $\mathbf{H}_k$  denotes the  $N_R \times N_T$  channel matrix with entries  $H_k^{(i,j)}$ , concerning a given channel realization, and  $\mathbf{Y}_k = [Y_k^{(1)}, Y_k^{(2)}, \dots, Y_k^{(N_R)}]^T$  is the resulting, frequency-domain, "output" vector.

As to a given MIMO channel realization, it should be noted that the Channel Frequency Response (CFR)  $\mathbf{H}^{(i,j)} = [H_0^{(i,j)}, H_1^{(i,j)}, \dots, H_{N-1}^{(i,j)}]^T$ , concerning the antenna pair  $(i, j)$ , is the DFT of the Channel Impulse Response (CIR)  $\mathbf{h}^{(i,j)} = [h_0^{(i,j)}, h_1^{(i,j)}, \dots, h_{N-1}^{(i,j)}]^T$ , where  $h_n^{(i,j)} = 0$  for  $n > L_s$  ( $n = 0, 1, \dots, N - 1$ ). Regarding a statistical channel model - which encompasses all possible channel realizations -, let us assume that  $E[h_n^{(i,j)*} h_{n'}^{(i,j)}] = 0$  for  $n' \neq n$ .

By assuming, for any  $(i, j, k)$ , a constant

$$E[|H_k^{(i,j)}|^2] = \sum_{n=0}^{N-1} E[|h_n^{(i,j)}|^2] = P_\Sigma, \quad (2)$$

(of course, with  $h_n^{(i,j)} = 0$  for  $L_s < n \leq N - 1$ ) and a 4-QAM SC/FDE block transmission, the average bit energy at

each Base Station (BS) antenna is given by

$$E_b = \frac{\sigma_s^2}{2\eta} P_\Sigma, \quad (3)$$

where  $\eta = \frac{N}{N+L_s}$ .

### B. Optimum (MMSE) Linear Detection Techniques

Linear detection techniques are considered in this paper for dealing with both MUI and Inter Symbol Interference (ISI). An appropriate linear detector can be implemented by resorting to frequency-domain processing, so as to jointly perform frequency-domain MultiUser Detection (MUD) and FDE procedures. After CP removal, a DFT operation leads to the required set  $\{\mathbf{Y}_k; k = 0, 1, \dots, N - 1\}$  of length- $N_R$  inputs to the frequency-domain detector ( $\mathbf{Y}_k$  given by (1)); it works, for each  $k$ , as shown in Fig. 1(a), leading to a set  $\{\tilde{\mathbf{Y}}_k; k = 0, 1, \dots, N - 1\}$  of length- $N_T$  outputs  $\tilde{\mathbf{Y}}_k = [\tilde{Y}_k^{(1)}, \tilde{Y}_k^{(2)}, \dots, \tilde{Y}_k^{(N_T)}]^T$  ( $k = 0, 1, \dots, N - 1$ ).

For the optimum (MMSE) linear detection technique, in the frequency-domain, it can be shown that

$$\tilde{\mathbf{Y}}_k = \mathbf{A}_k^{-1} \mathbf{H}_k^H \mathbf{Y}_k, \quad (4)$$

(see Fig. 1(a)), where

$$\mathbf{A}_k = \mathbf{H}_k^H \mathbf{H}_k + \alpha \mathbf{I}_{N_T}, \quad (5)$$

with  $\alpha = \frac{\sigma_N^2}{\sigma_s^2} = \frac{N_0}{\sigma_s^2} \left( \sigma_S^2 = E[|S_k^{(j)}|^2] = N \sigma_s^2 \right)$ .

For each user  $j$ , the required time-domain decisions are then based on the IDFT of the length- $N$  block  $\tilde{\mathbf{Y}}^{(j)} = [\tilde{Y}_0^{(j)}, \tilde{Y}_1^{(j)}, \dots, \tilde{Y}_{N-1}^{(j)}]^T$  ( $j = 1, 2, \dots, N_T$ ), as shown in Fig. 1(c).

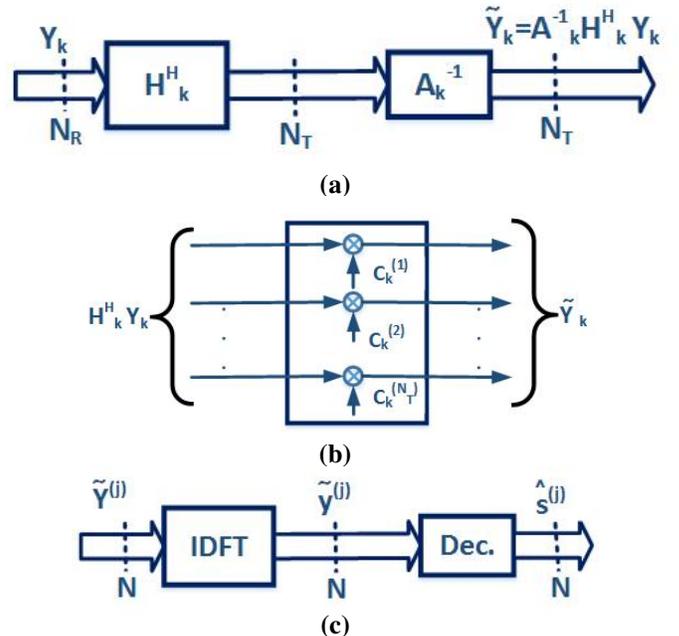


Figure 1. Linear, frequency-domain detection procedure ( $k = 0, 1, \dots, N - 1$ ) (a), reduced-complexity implementation regarding  $\mathbf{A}_k^{-1}$  (b) and time-domain decision procedure for user  $j$  ( $j = 1, 2, \dots, N_T$ ) (c).

### C. Reduced-complexity Linear Detection Techniques

Instead of the optimum (MMSE) linear detector, a reduced-complexity linear detection technique can be implemented by replacing the  $\mathbf{A}_k$  matrix in (4) by a diagonal  $\mathbf{A}'_k$  matrix sharing the same entries

$$A'_k{}^{(j,j)} = A_k{}^{(j,j)} = \alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2, \quad (6)$$

in the main diagonal. Therefore, the required matrix inversion in (4) becomes a very easy task, and the corresponding reduced-complexity implementation of the second block in Fig. 1 (a) can be done according to Fig. 1 (b), with

$$C_k^{(j)} = \frac{1}{\alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}, \quad (7)$$

Consequently, the  $N_T$  components of  $\tilde{\mathbf{Y}}_k$  can be decomposed - into "useful signal", MUI and "Gaussian noise" - as follows:

$$\begin{aligned} \tilde{Y}_k^{(j)} &= C_k^{(j)} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2 S_k^{(j)} + \\ &+ C_k^{(j)} \sum_{l \neq j} \sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*} S_k^{(l)} + \\ &+ C_k^{(j)} \sum_{i=1}^{N_R} H_k^{(i,j)*} N_k^{(i)}, \end{aligned} \quad (8)$$

( $j = 1, 2, \dots, N_T$ ).

As an alternative to the  $C_k^{(j)}$  coefficients given by (7) - in the context of a simplified signal processing structure as jointly depicted in Fig. 1(a) and Fig. 1(b) - one can adopt  $\alpha'_k{}^{(j)}$  coefficients so as to meet the MMSE criterion.

It can be shown that, under the reduced-complexity constraint, the resulting MMSE coefficients can be written as

$$C_k^{(j)} = \frac{1}{\alpha'_k{}^{(j)} + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}, \quad (9)$$

where

$$\alpha'_k{}^{(j)} = \alpha + \sum_{l \neq j} \alpha_k^{(l,j)}, \quad (10)$$

with  $\alpha = \frac{N_0}{\sigma_s^2}$  and

$$\alpha_k^{(l,j)} = \frac{\left| \sum_{i=1}^{N_R} H_k^{(i,j)*} H_k^{(i,l)} \right|^2}{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2} \quad (11)$$

### III. SEMI-ANALYTICAL METHODS FOR PERFORMANCE EVALUATION

#### A. Performance Evaluation Method in the Optimum (MMSE) Case

The frequency-domain output  $\tilde{\mathbf{Y}}_k$  of the MMSE detector in Fig. 1(a) can be written as

$$\tilde{\mathbf{Y}}_k = \mathbf{\Gamma}_k \mathbf{S}_k + \text{'Noise-like term'} \quad (12)$$

where

$$\mathbf{\Gamma}_k = [\mathbf{H}_k^H \mathbf{H}_k + \alpha \mathbf{I}_{N_T}]^{-1} \mathbf{H}_k^H \mathbf{H}_k \quad (13)$$

For 4-QAM SC/FDE transmission and optimum (MMSE) detection, the resulting  $BER_j$  ( $j = 1, 2, \dots, N_T$ ) - conditional on the channel realization  $\{\mathbf{H}_k; k = 0, 1, \dots, N-1\}$  - is given by

$$BER_j \approx Q\left(\sqrt{SINR_j}\right) \quad (14)$$

where  $SINR_j$  denotes the signal-to-"Interference plus Noise" Ratio regarding the components of the  $j$ th, time-domain, output block  $\tilde{\mathbf{y}}^{(j)} = [\tilde{y}_0^{(j)}, \tilde{y}_1^{(j)}, \dots, \tilde{y}_{N-1}^{(j)}]^T = IDFT\left([\tilde{Y}_0^{(j)}, \tilde{Y}_1^{(j)}, \dots, \tilde{Y}_{N-1}^{(j)}]^T\right)$ . It can be shown that

$$SINR_j = \frac{\gamma^{(j)}}{1 - \gamma^{(j)}} \quad (15)$$

in eqn. (14), with

$$\gamma^{(j)} = \frac{1}{N} \sum_{k=0}^{N-1} \Gamma_k^{(j,j)}, \quad (16)$$

where  $\Gamma_k^{(j,j)}$  denotes the  $(j, j)$  entry of  $\mathbf{\Gamma}_k$  defined in (13) ( $\gamma^{(j)} = \frac{E[s_n^{(j)*} \tilde{y}_n^{(j)}]}{\sigma_s^2}$ , since  $\tilde{y}_n^{(j)} = \gamma^{(j)} s_n^{(j)} + \text{'uncorrelated noise-like term'}$ ).

Of course, the average BER for the channel realization  $\{\mathbf{H}_k; k = 0, 1, \dots, N-1\}$  can be easily derived from (14):

$$BER = \frac{1}{N_T} \sum_{j=1}^{N_T} BER_j, \quad (17)$$

#### B. Performance Evaluation Method in the Reduced-complexity Cases

When using the MMSE criterion under the constraint of a simplified detection structure, based on Fig. 1(a) and Fig. 1(b), the  $C_k^{(j)}$  coefficients are given by (9). In this case, we can still write (12), but now with

$$\mathbf{\Gamma}_k = \mathbf{A}'_k{}^{-1} \mathbf{H}_k^H \mathbf{H}_k, \quad (18)$$

where  $\mathbf{A}'_k$  is a diagonal matrix with entries

$$A'_k{}^{(j,j)} = \frac{1}{C_k^{(j)}} = \alpha'_k{}^{(j)} + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2, \quad (19)$$

which replaces  $\mathbf{A}_k$  ( $\alpha_k^{(j)}$  given by (10) and (11)).

Therefore, eqns. (15) and (16), for the  $SINR_j$  are still valid; however, due to the different  $\mathbf{\Gamma}_k$  matrix,

$$\gamma^{(j)} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}{\alpha_k^{(j)} + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2} \quad (20)$$

in (15), in this case. Of course, the resulting BER (conditional on the channel realization  $\{\mathbf{H}_k; k = 0, 1, \dots, N-1\}$ ) can then be computed according to eqns. (14) and (17).

With regard to the reduced-complexity technique which uses  $C_k^{(j)}$  coefficients according to (7) (not according to (9)), it should be noted that it cannot be regarded, in general, as a true "MMSE technique". The only exception is the case where  $N_T = 1$ , leading to  $\alpha_k^{(j)} = \alpha$ : in this special, Single User (SU) case - with ISI but not MUI- we still could adopt (15), with

$$\gamma^{(j)} = \frac{1}{N} \sum_{k=0}^{N-1} \frac{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}{\alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}; \quad (21)$$

consequently, for  $N_T = 1$ , we should get

$$SINR_{j,SU} = \frac{1}{\alpha} \frac{\sum_{k=0}^{N-1} \frac{\sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}{\alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}}{\sum_{k=0}^{N-1} \frac{1}{\alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2}} \quad (22)$$

For  $N_T > 1$ , an appropriate MUI term should be added to the "ISI + noise term". The resulting  $SINR_j$  is given by

$$SINR_j = \frac{SINR_{j,SU}}{1 + SINR_{j,SU} \frac{\sigma_{MUI}^2(j)}{\sigma_s^2}}, \quad (23)$$

where

$$\sigma_{MUI}^2 = \frac{\sigma_s^2}{N (\gamma^{(j)})^2} \sum_{l \neq j} \sum_{k=0}^{N-1} \left| \frac{\sum_{i=1}^{N_R} H_k^{(i,l)} H_k^{(i,j)*}}{\alpha + \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2} \right|^2, \quad (24)$$

with  $\gamma^{(j)}$  obtained from (21), concerns the MUI when  $N_T > 1$ .

#### IV. NUMERICAL PERFORMANCE RESULTS, DISCUSSION OF MASSIVE MIMO EFFECTS AND SC/FDE VS OFDM COMPARISONS

The set of performance results which are presented here are concerned to 4-QAM SC/FDE uplink block transmission, with  $N = 256$  and  $L_s = 64$ , in a MU-MIMO  $N_T \times N_R$  Rayleigh fading channel. The fading effects regarding the several TX/RX antenna pairs are assumed to be uncorrelated, and two possibilities are considered for the CIRs of the channel realizations (the first possibility only for some performances of Fig. 4):

- A zero-mean, complex Gaussian  $h_0^{(i,j)}$  with variance  $P_\Sigma$ , and  $h_n^{(i,j)} = 0$  for  $n = 1, \dots, 255$  (i.e. a frequency-flat Rayleigh fading);
- Independent zero-mean complex Gaussian  $h_n^{(i,j)}$  coefficients, all of them with variance  $\frac{P_\Sigma}{64}$ , for  $n = 0, 1, \dots, 63$ , and  $h_n^{(i,j)} = 0$  for  $n = 64, 65, \dots, 255$  (i.e. a strongly frequency-selective Rayleigh fading).

With regard to the linear detection techniques of Sections II-B and II-C, the several performance results concerning the MU-MIMO system have been obtained by random generation of a large number of channel realizations, analytical BER computation - according to the methods of Section III - for each channel realization, and an averaging operation over the set of channel realizations. The accuracy of performance results obtained through these semi-analytical simulation methods was assessed by means of parallel conventional Monte Carlo simulations (involving an error counting procedure).

When  $N_R \gg N_T$ , both the MUI effects and the effects of multipath propagation (fading, ISI) tend to disappear: consequently, the BER performances for the MU-MIMO  $N_T \times N_R$  Rayleigh fading channel become very close to those concerning a Single-Input Multi-Output (SIMO)  $1 \times N_R$  channel with single-path propagation for all  $N_R$  TX/RX antenna pairs. The achievable performances under a "truly massive" MU-MIMO implementation can be analytically derived as shown below.

Entries of  $\mathbf{H}_k$  are i.i.d. Gaussian-distributed random variables with zero mean and variance  $P_\Sigma$ . According to the law of large numbers,

$$\lim_{N_R \rightarrow \infty} \left[ \frac{1}{N_R} \sum_{i=1}^{N_R} |H_k^{(i,j)}|^2 \right] = E \left[ |H_k^{(i,j)}|^2 \right] = P_\Sigma, \quad (25)$$

and

$$\lim_{N_R \rightarrow \infty} \left[ \frac{1}{N_R} \sum_{\substack{i=1 \\ (i \neq j)}}^{N_R} H_k^{(i,j)*} H_k^{(i,l)} \right] = \underset{l \neq j}{E} \left[ H_k^{(i,j)*} H_k^{(i,l)} \right] = 0.$$

Consequently, when  $N_R \gg N_T$ ,  $\tilde{Y}_k^{(j)} \approx \left[ S_k^{(j)} N_R P_\Sigma + \text{'Gaussian noise with variance } N N_0 N_R P_\Sigma \text{' } \right] \times C_k^{(j)}$ , with

$$C_k^{(j)} \approx \frac{1}{\alpha + N_R P_\Sigma} = C \text{ (for any } (j, k)). \quad (26)$$

Therefore, practically there is neither MUI nor ISI and fading at the time-domain outputs:  $\tilde{y}_n^{(j)} \approx \left[ s_n^{(j)} N_R P_\Sigma + \text{'Gaussian noise with variance } N_0 N_R P_\Sigma \text{' } \right] \times C$ .

The resulting BER performance becomes as follows:

$$BER \approx Q \left( \sqrt{\frac{N_R P_\Sigma \sigma_s^2}{N_0}} \right) = Q \left( \sqrt{2\eta N_R \frac{E_b}{N_0}} \right), \quad (27)$$

Figures 2 and 3 show the simulated BER performances for an SC/FDE-based MU-MIMO uplink and several possibilities regarding  $N_T$  and  $N_R$ , when using the linear detection techniques of Section II: optimum (MMSE) detection; reduced-complexity detection, under  $C_k^{(j)}$  coefficients given by eq. (7)

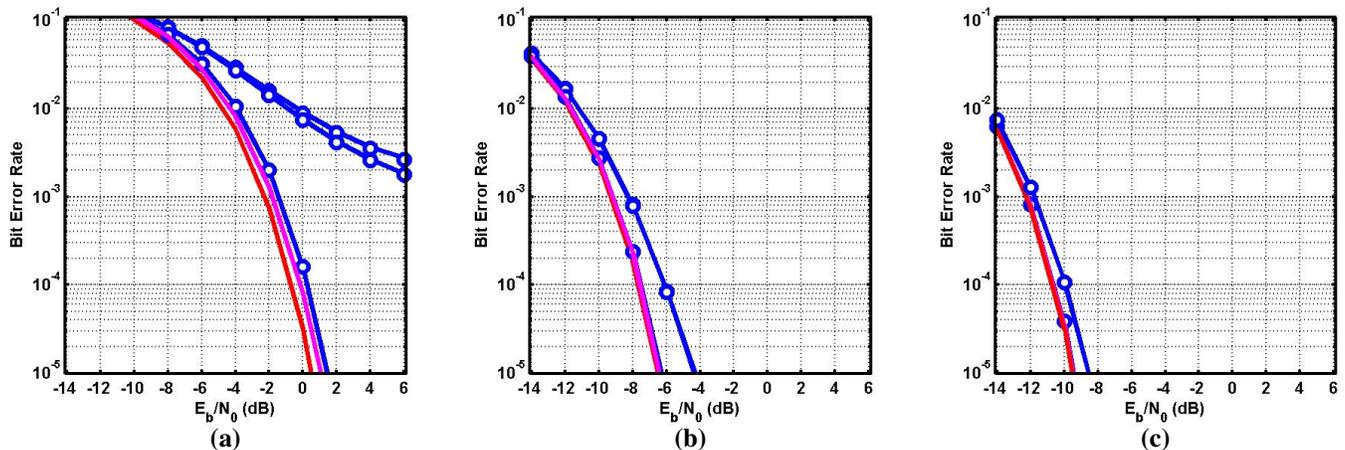


Figure 2. BER performances for SC/FDE-based MU-MIMO, with  $N_T = 2$ , and  $N_R = 10$  (a), 50 (b) or 100 (c), under reduced-complexity (I, II) and MMSE linear detection [ $1 \times N_R$  (single-path, multipath) reference BER performances are also included, and the five BER performances are ordered, from the worst to the best, as explained in section IV].

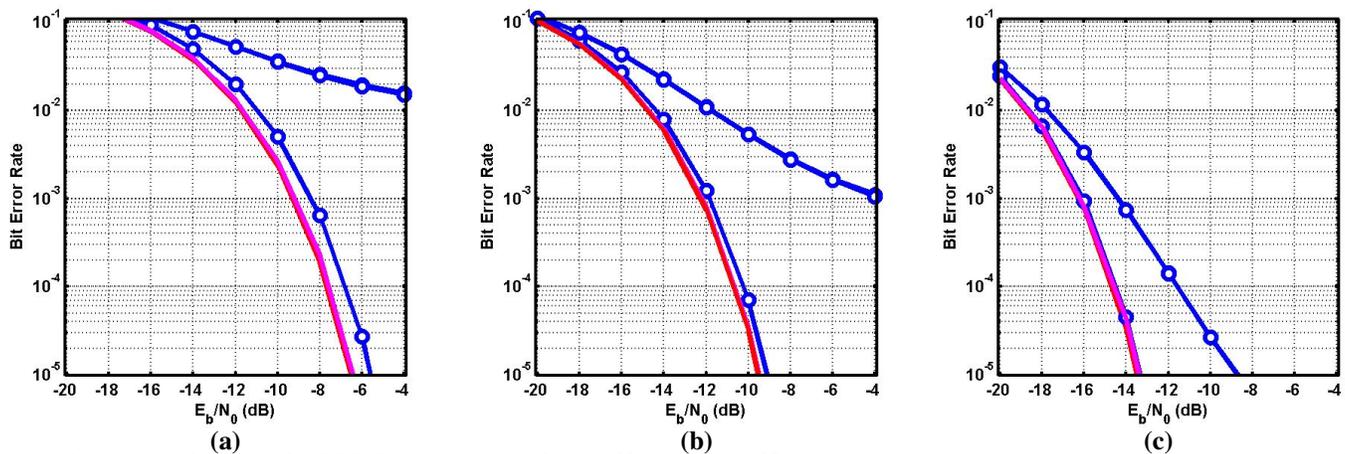


Figure 3. BER performances for SC/FDE-based MU-MIMO, with  $N_T = 10$ , and  $N_R = 10$  (a), 50 (b) or 100 (c), under reduced-complexity (I, II) and MMSE linear detection [ $1 \times N_R$  (single-path, multipath) reference BER performances are also included, and the five BER performances are ordered, from the worst to the best, as explained in section IV].

(I) or eq. (9) (II). In both figures, for the sake of comparisons, we also include  $1 \times N_R$  reference performances, for both the multipath propagation channel - which implies a Rayleigh fading concerning each TX/RX antenna pair - and an ideal single-path propagation channel. For the linear detection techniques, the semi-analytical methods of Section III have been adopted; the complementary conventional Monte Carlo simulation (involving error counting) results correspond to the superposed circles in the solid lines.

Fig. 4 is dedicated to an 'SC/FDE vs OFDM' comparison of BER performances, for the strongly frequency-selective fading channel case and the MMSE linear detection. In fact, the "OFDM results" shown here could have been obtained by resorting to the SC/FDE simulation software, by replacing the strongly frequency-selective fading by a frequency-flat fading. This is due to the following reasons: under frequency-flat fading, uncoded SC/FDE and OFDM provide identical performances; uncoded OFDM performance does not depend on the frequency-selectivity of the fading effects.

In all figures, where the SIMO detection performance was analytically computed according to (22), an excellent

agreement of the semi-analytical simulation results with conventional Monte Carlo simulation results can be observed.

In the simulation results concerning all subfigures of both Fig. 2 and Fig. 3, the five BER performance curves have been shown to be ordered, from the worst to the best, as follows:  $N_T \times N_R$  MU-MIMO with reduced-complexity (I) linear detection;  $N_T \times N_R$  MU-MIMO with reduced-complexity (II) linear detection;  $N_T \times N_R$  MU-MIMO with MMSE detection;  $1 \times N_R$  (multipath case) SIMO detection;  $1 \times N_R$  (single-path case) SIMO detection. These figures clearly show that the performance degradation which is inherent to the reduced-complexity linear detection techniques (I and II) - as compared with the optimum (MMSE) linear detection - can be made quite small, by increasing  $N_R$  significantly; they also show that, under highly increased  $N_R$  values, the "MUI-free" SIMO (multipath) performance and the ultimate bound - the "MUI-free and ISI & fading-free" SIMO (single-path) performance - can be closely approximated, even when adopting the reduced-complexity (I) linear detection. These two figures emphasize a "massive MIMO" effect when  $N_R \gg 1$ , especially when  $N_R \gg N_T$  too, which leads to BER performances very

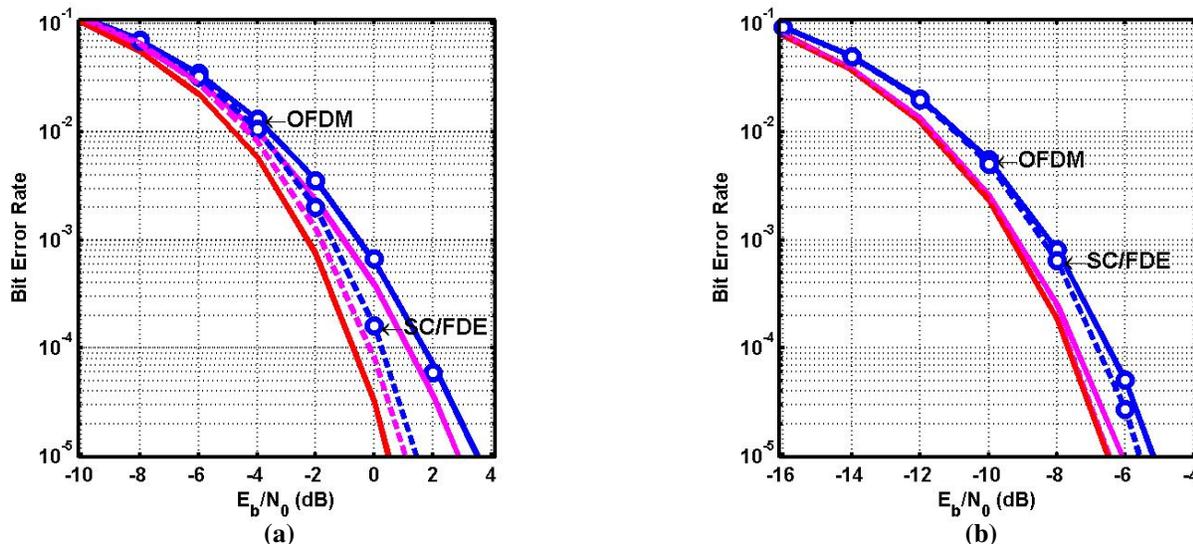


Figure 4. SC/FDE (dashed lines) vs OFDM (solid lines) BER performances, for  $2 \times 10$  (a) and  $10 \times 50$  (b), under strongly frequency-selective Rayleigh fading [SIMO  $1 \times N_R$  (multipath, single-path) reference performances are also included].

close to the ultimate "MUI-free and ISI & fading-free" SIMO (single-path) performance bound.

With regard to 'SC/FDE vs OFDM' (uncoded) BER performance comparisons, we can remember the significant performance advantage of SC/FDE when the Rayleigh fading is frequency-selective and  $N_R = 1$  or 2 [2][3]. However, Fig. 4 clearly shows that - in spite of the strongly frequency-selective fading considered here - the performance advantage of SC/FDE practically vanishes when  $N_R \gg 1$ , even for a moderate  $\frac{N_R}{N_T}$  (e.g., equal to 5).

Not surprisingly - having in mind the comparison depicted in Fig. 4 - Fig. 2 and Fig. 3 of this paper are very similar, respectively, to Fig. 3 and Fig. 4 of [10] (where the OFDM-based MU-MIMO alternative is considered).

## V. CONCLUSIONS

This paper was dedicated to the uplink performance evaluation of a MU-MIMO system with SC/FDE transmission, when adopting a large number of antennas and linear detection techniques at the BS. The numerical performance results, discussed in detail in Section IV, show the "massive MIMO" effects provided by a number of BS antennas much higher than the number of antennas which are jointly employed in the terminals of the multiple autonomous users, even when reduced-complexity linear detection techniques are adopted.

The accuracy of performance results obtained by semi-analytical means, much less time-consuming than conventional, 'error counting'-based, Monte Carlo simulations - was also demonstrated. The proposed performance evaluation method can be very useful for rapidly knowing "how many antennas do we need in the BS?", so that a "massive MIMO" effect can be achievable, for a given number of antennas jointly employed in the user terminals.

The performance results of this paper also clearly show that the SC/FDE detection performance, in a MU-MIMO context with a large number of BS antennas, cannot be significantly

better than that of OFDM: in fact, the SC/FDE performance advantage practically vanishes when  $N_R \gg N_T$ , even for a strongly frequency-selective fading channel. Nevertheless, we can say that SC/FDE is a better choice than OFDM for up-link transmission, due to its well-known "power amplification advantage" [2] and to the fact that it does not suffer from a detection performance disadvantage.

## REFERENCES

- [1] H. Sari, G. Karam, and I. Jeanclaude. An analysis of orthogonal frequency-division multiplexing for mobile radio applications. In *Vehicular Technology Conference, 1994 IEEE 44th*, pages 1635–1639 vol.3, Jun 1994.
- [2] A. Gusmao, R. Dinis, J. Conceicao, and N. Esteves. Comparison of two modulation choices for broadband wireless communications. In *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, volume 2, pages 1300–1305 vol.2, may 2000.
- [3] A. Gusmao, R. Dinis, and N. Esteves. On frequency-domain equalization and diversity combining for broadband wireless communications. *Communications, IEEE Transactions on*, 51(7):1029–1033, July 2003.
- [4] G. Foschini. "Layered Space-Time Architecture for Wireless Communication in a Fading Environment when Using Multi-element Antennas". *Bell Labs Technology Journal*, (2):1, 1996.
- [5] J. Mietzner, R. Schober, L. Lampe, W.H. Gerstacker, and P.A. Hoeher. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *Communications Surveys Tutorials, IEEE*, 11(2):87–105, Second 2009.
- [6] E.G. Larsson. Mimo detection methods: How they work [lecture notes]. *Signal Processing Magazine, IEEE*, 26(3):91–95, May 2009.
- [7] D. Gesbert, M. Kountouris, R.W. Heath, Chan-Byoung Chae, and T. Salzer. Shifting the mimo paradigm. *Signal Processing Magazine, IEEE*, 24(5):36–46, Sept 2007.
- [8] T.L. Marzetta. Noncooperative cellular wireless with unlimited numbers of base station antennas. *Wireless Communications, IEEE Transactions on*, 9(11):3590–3600, November 2010.
- [9] J. Hoydis, S. ten Brink, and M. Debbah. Massive mimo in the ul/dl of cellular networks: How many antennas do we need? *Selected Areas in Communications, IEEE Journal on*, 31(2):160–171, February 2013.
- [10] P. Torres, L. Charrua, and A. Gusmao. "Uplink Performance Evaluation of Broadband Systems which Adopt a Massive MU-MIMO Approach". *International Conference on Wireless and Mobile Communications, ICWMC*, June 2014.

# Capacity Estimation of IEEE 802.15.4 Chains of Sensor Nodes

Evaggelos Chatzistavros, George Stamatelos

Department of Electrical and Computer Engineering,  
Democritus University of Thrace  
Xanthi, 67100, Greece  
Email: ec3526@ee.duth.gr

**Abstract**—In the context of an integrated water management project, we derive an analysis based on numerical modelling for throughput estimation in an IEEE 802.15.4 wireless sensor network (WSN). It takes into account the number of nodes in a cluster network, as well as transmitted packet size. We focus on the IEEE 802.15.4 compliant slotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) algorithm. Theoretical estimations are verified using the ns-2 network simulator. The configuration examined is a chain network formed by wireless sensor nodes where the first node is the source of data packets and the last is the traffic sink. Simulation results show that there is a satisfactory approximation between the theoretically estimated and the simulated values.

**Keywords**—Performance Evaluation; Wireless Sensor Networks; Analytical Model; Personal Area Network.

## I. INTRODUCTION

Wireless sensor networks are widely deployed in a variety of applications, including home monitoring and automation [1]–[3], environmental monitoring [4] [5] and health monitoring [6]–[8]. Depending on system requirements and the type of application, sensor nodes are usually deployed in three different topologies, i.e., peer-to-peer (also called point to point), star and mesh. In the peer-to-peer network, nodes can directly communicate with each other, without the mediation of a Personal Area Network (PAN) Coordinator. All sensor nodes in a star topology are connected to a PAN coordinator. Node communications are routed through the coordinator. A tree topology is considered a hybrid of both the peer to peer and star configurations. It consists of different hierarchical levels, the lowest of which forms one or more star networks.

An environmental monitoring system is proposed in the context of the CYBERSENSORS project [9]. The system architecture under consideration employs two distinct subsystems: a physical /chemical parameters monitoring subsystem and a visual monitoring subsystem, as shown in Fig. 1. Regarding the first, it consists of wireless sensor probes communicating with a PAN coordinator, forming an IEEE 802.15.4-compliant WSN [10]. In the second subsystem, Visual Sensor Nodes communicate with an IEEE 802.11n router (forming a Wireless Local Area Network (WLAN)) [11]. Data collected from both networks are forwarded to a remote server through a 3G/HSPA broadband link [12]. Using a peer-to-peer setup can extend WSN range, especially when efficient self-configuration and large area coverage are important [13].

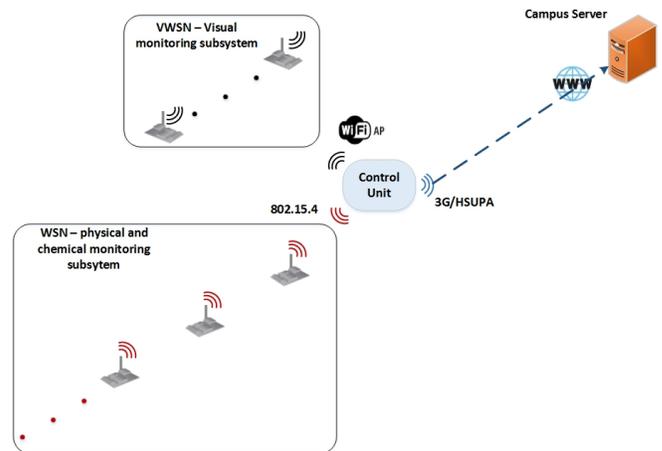


Figure 1. System architecture for integrated water management.

In this paper, we look into performance issues associated with the sensors subsystem responsible for sampling and storing of physical (i.e., suspended particles, temperature) and chemical (i.e., conductivity, dissolved oxygen, nitrate, pH, heavy metals) quality metrics [4]. Sensor nodes operate in an IEEE 802.15.4/Zigbee framework for the wireless data transmission of the collected measurements. Within this high frequency monitoring platform, sensor nodes will be deployed along a river bank.

In [14], authors proposed an analytical performance model for a sensor network with specific delay and packet delivery ratio requirements. Their model is validated using the ns-2 simulator [15] for a star-topology sensor network. A Markov renewal process [16] is embedded for the calculation of saturation throughput. Authors in [17] developed an analytical model using Markov chains, for both a one-hop star topology, as well as a multi-hop sensor network. They examine the unslotted IEEE 802.15.4 CSMA/CA algorithm, validated by Monte Carlo simulations. A worst-case modeling methodology based on Network Calculus [18] [19] is described in [20]. Sensor nodes forming a cluster-tree network compete for channel access following the slotted CSMA/CA algorithm.

In the following sections, we propose a method for calculating network capacity when a number of remote sensor nodes are employed for data acquisition and the rest for packet forwarding to the PAN coordinator. Throughput is derived as a function of the number of relay nodes in the network and

packet size. These results will be used for the deployment of an operational sensor network.

The rest of the paper is organized as follows: the analytical model is described in Section II. Model evaluation is presented in Section III, while conclusions are drawn in the last section.

## II. NUMERICAL ANALYSIS

Two distinct data transfer modes are supported by an IEEE 802.15.4 network, i.e., beacon enabled and non - beacon enabled. In the beacon enabled mode, the PAN coordinator transmits periodic beacons, allowing node synchronization. This scheme employs the superframe structure. The *macBeaconOrder* (BO) attribute describes the interval at which the coordinator shall transmit beacon frames. The Beacon Interval (BI) is related to the *macBeaconOrder* as follows:

$$BI = aBaseSuperframeDuration * 2^{BO}$$

$$\text{for } 0 \leq BO \leq 14,$$

where,

$$aBaseSuperframeDuration = aNumSuperframeSlots * aBaseSlotDuration$$

The length of the active portion of the superframe is described by the *macSuperframeOrder* (SO) attribute. The Superframe Duration (SD) is related to the SO attribute as follows:

$$SD = aBaseSuperframeDuration * 2^{SO}$$

$$\text{for } 0 \leq SO \leq BO$$

The active portion of each superframe is divided into *aNumSuperframeSlots* = 16 equally spaced slots and is composed of three parts: a beacon, a Contention Access Period (CAP) and a Contention Free Period (CFP). If  $SO < BO$ , the active superframe portion is shorter compared to the BI, allowing nodes to enter sleep mode.

A numerical analysis model for throughput estimation is developed in this paper. As stated in the previous paragraph, we assume there is no CFP period.

### A. Two nodes transmission

Our analysis begins with the simplest scenario, where a node transmits its data directly to the PAN coordinator. We also assume equal BO and SO values greater than 4. In this case, a node transmits its packets during a single time slot of a superframe. If BO (and consequently SO)  $< 4$ , a node will not complete its transmission during the current time slot and should wait for the next superframe to complete transmission, as indicated by the Medium Access Control (MAC) mechanism. Also in this analysis, there are no Guaranteed Time Slots (GTS) during a superframe.

We assume that there are no other contending nodes to gain channel access within the range of the sensor nodes. The Backoff Exponent (BE) parameter is related to how many backoff periods a device shall wait before attempting to assess a channel. In the case of two transmitting nodes, it is limited to its lowest value  $BE=3$ , due to low contention levels. Based on

the slotted CSMA/CA algorithm, the total duration of a frame transmission is calculated as:

$$T_{tot} = (d_{BE} + d_{fr} + d_{LIFS} + d_{trn} + d_{CW} + d_{ack}) \times T_s \quad (1)$$

where

- $d_{BE}$  : backoff period duration, in symbols
- $d_{fr}$  : the duration of a frame transmission, in symbols
- $d_{LIFS}$  : Interframe Spacing Time, the number of symbols separating two successive frames,
- $d_{trn}$  : turnaround time, the number of symbols required to switch the RF transceiver from receive to transmit mode in order to transmit an acknowledgement,
- $d_{CW}$  : time required for the necessary clear back-off periods prior to the transmission. There are two Contention Window (CW) slots, each of 20 symbols length,
- $d_{ack}$  : transmission time required for an acknowledgement frame, in symbols.

The 250 kbps data rate supported by the IEEE 802.15.4 standard are equivalent to  $62.5 \frac{ksymbols}{sec}$ . The duration of a symbol,  $T_s$  equals to  $\frac{1}{62500 \frac{symbols}{sec}} = 0.016 \frac{ms}{symbol}$

The maximum frame size is 133 bytes (102 bytes application data + 25 bytes PHY overhead + 6 bytes MAC overhead).  $d_{fr}$  expressed in symbols is equal to:

$$d_{fr} = \frac{133 \frac{bytes}{frame} \times 8 \frac{bits}{byte}}{4 \frac{bits}{symbol}} = 266 \frac{symbols}{frame} \quad (2)$$

As the number of symbols forming the basic time period used by the CSMA/CA algorithm is 20 symbols, each time period described below must be expressed as an integer multiple of 20 symbols.  $d_{LIFS}$ ,  $d_{trn}$ ,  $d_{CW}$  and  $d_{ack}$  values are defined by the IEEE 802.15.4 standard.

Therefore,

- $d_{BE} = (2^{BE}-1) * aUnitBackoffPeriod$ , where *aUnitBackoffPeriod* is equal to 20 symbols
- $d_{ack} = 10$  symbols. A total of 20 symbols are reserved
- $d_{LIFS} = 40$  symbols
- $d_{trn} = 12$  symbols. As in  $t_{ack}$  20 symbols occupied
- $d_{CW} = 2*20=40$  symbols

Based on the above estimations and (1),  $T_{tot}$  is equal to:

$$\begin{aligned} T_{tot} &= (d_{BE} + d_{fr} + d_{ack} + d_{LIFS} + d_{trn} + d_{CW}) \times T_s \\ &= 526 \frac{symbols}{frame} \times 0.016 \frac{ms}{symbol} = 0.008416 \frac{sec}{frame} \quad (3) \end{aligned}$$

Subsequently, the chain throughput approximation is equal to:

$$S = \frac{1}{0.008416 \frac{sec}{frame}} = 118 \frac{frame}{sec} \times 133 \frac{bytes}{frame} \times 8 \frac{bits}{byte} \quad (4)$$

Taking into account only the application payload, application throughput is equal to:

$$S_{ap} = 118 \times 102 \times 8 \simeq 96.29kbps \quad (5)$$

### B. N-nodes chain analysis

When increasing chain size to three nodes, the second node forwards packets received from the last (third), to the coordinator. The absence of collisions maintains node contention low, so the BE parameter is still limited to its lowest value  $BE = 3$  for the first two nodes of the chain. The PAN coordinator is not involved in contention, as it only receives packets. The number of symbols now required for a packet transmission is twice the number of those involved in the two-node analysis, i.e., 1052 symbols.

The successive packets of a single connection interfere with each other as they move down the chain, forcing contention in the MAC protocol. As node contention increases with network size, the BE index increases as well. Contention reaches maximum level when the chain consists of more than 3 sensor nodes. According to the slotted CSMA/CA algorithm (shown in Fig. 2 [10]), the BE index reaches its maximum value  $macMaxBE = 5$ .

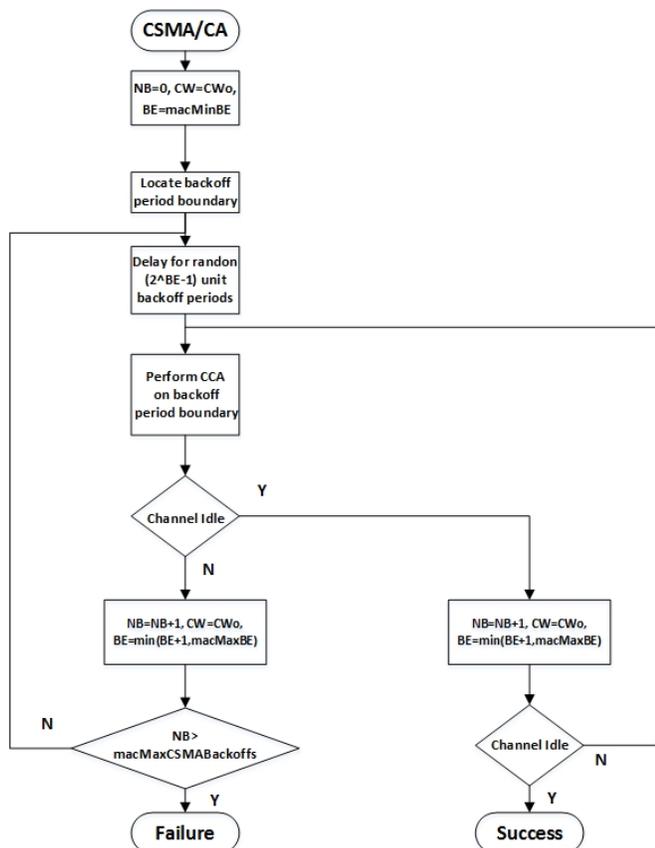


Figure 2. Slotted IEEE 802.15.4 CSMA/CA algorithm (from [10]).

An 802.15.4 node's ability to send is affected by the amount of competition it experiences. For example, node 3 in a 7-node chain experiences interference from 4 other nodes, while node 1 is interfered with by two other nodes. This means that node 1 could actually inject more packets into the chain than the subsequent nodes can forward. In the n-nodes chain illustrated in Fig. 3, the last is the source and the first is the data sink. Packets travel along a chain of n-1 intermediate hops until they reach their destination. For the first transmission between nodes n-1 and n-2 contention is low, thereby the BE value is kept at its lowest value, i.e., 3. For the subsequent n-2 hops contention levels are higher since three or more nodes potentially attempt to gain access to the channel at the same time. Due to the increased contention levels, the BE index maintains the  $macMaxBE$  value.

Subsequently, we proceed to the calculation of the throughput equation. We assume a Constant Bit Rate (CBR) application (to model the periodical sensor sampling) which transmits packets of size k bytes (data payload). The frame size (send by the PHY layer) is k + 33 (PHY + MAC overhead) bytes, which is equal to:

$$(k + 33) \frac{bytes}{frame} \times \frac{8 \frac{bits}{byte}}{4 \frac{bits}{symbol}} = 2 \times (k + 33) \frac{symbols}{frame} \quad (6)$$

The maximum backoff duration,  $d_{BE}$ , depends on the number of nodes forming the network. The remaining terms in (1), i.e.,  $d_{ack}$ ,  $d_{LIFS}$ ,  $d_{trn}$  are constant. Based on the above estimation and (4), data throughput can be approximated by the following equation:

$$S = \frac{1}{\frac{T_{tot} \frac{symbols}{frame}}{62500 \frac{symbols}{sec}}} \times k \frac{bytes}{frame} \times 8 \frac{bits}{byte} = 500 \times \frac{k}{T_{tot}} kbps \quad (7)$$

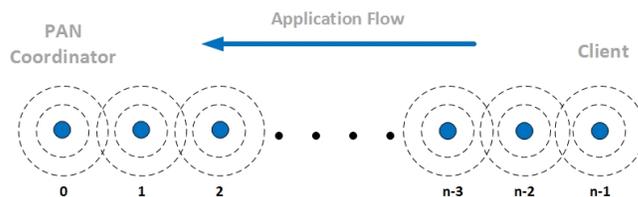


Figure 3. Chain consisting of n sensor nodes.

In order to verify the accuracy of (7), we conduct a number of simulations. Details on the simulated topologies and corresponding results are presented in the following section.

### III. MODEL EVALUATION

Here, the ns-2 simulator is used to verify the accuracy and scaling of the model already discussed. Nodes are configured according to the IEEE 802.15.4 standard, operating at 2.4 GHz with a maximum transmission rate of 250 kbps. The simulation parameters are listed in Table I.

TABLE I. SIMULATION PARAMETERS

IEEE Standard	IEEE 802.15.4
MAC Protocol	Beacon Enabled CSMA - IEEE 802.15.4
Transmission range	20 m
Beacon Order - BO	7
Superframe Order - SO	7
Number of nodes	2 to 100
Application Type	Constant Bit Rate
Offered Traffic Load	10 to 250 kbps
Packet Size	25, 50, 75, 100 bytes
Simulation time	200 sec
Application duration	50 sec
Bit Error Rate (BER)	perfect channel conditions (0)

The WSN topology consists of a sensor nodes chain of increasing length from 2 to 10 nodes, forming a cluster network of static nodes. Each node has an effective transmission range of 20 meters (an interference range of 35 meters) and is located 15 meters away from its immediate neighbors. All nodes operate as Full Function Devices (FFDs), with the first one being the PAN coordinator and the last one serving as the application client. The rest are relay nodes, forwarding received packets from the client to the coordinator.

A CBR flow is applied, running for 50 seconds, four times for each topology, each time sending packets of different size, i.e., 25, 50, 75 and 100 bytes. In the CYBERSENSORS project, data will be collected and forwarded periodically, which makes the choice of CBR data model appropriate. For each configuration, the transmission data rate increases from 10 to 250 kbps with steps of 10 kbps each. Prior to the initiation of the CBR flow and data collection, network simulation runs for a time period of 150 seconds, necessary to reach a stable state in terms of node synchronization and association. Average throughput values are simulated for four different packet sizes in Fig. 4.

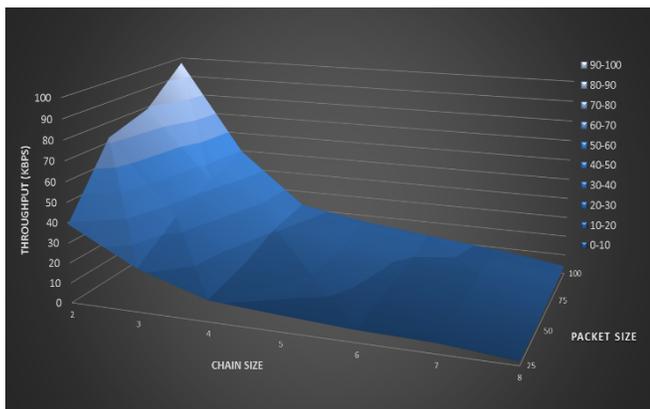


Figure 4. Average throughput (simulated) for a maximum chain length of 8 nodes.

Average throughput for the case of a three-nodes-chain with payload 100 bytes per packet is 49.6 kbps. For the same network size, when packet size is reduced to 25 bytes, the corresponding value degrades to 21.46 kbps. Packet loss levels not explicitly shown here follow the reverse gradient of the curves in Fig. 4, since high throughput values correspond to low packet losses.

Figures 5 and 6 present a comparison between throughput calculated by the numerical analysis model in the previous section and that obtained through simulations for different

network sizes and offered traffic load. In both figures, the deviation between corresponding lines varies from 0.023 to 3.179 kbps (minimum and maximum values, respectively), with an average deviation value of 1.008 kbps.

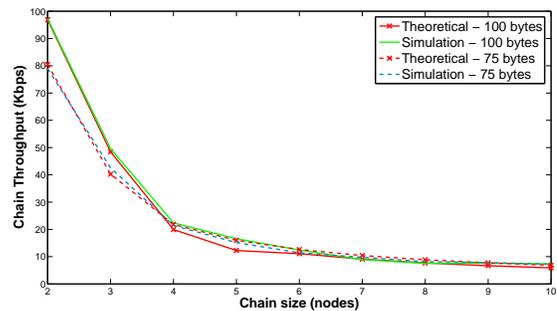


Figure 5. Theoretical vs simulated throughput achieved along a chain of nodes, as a function of the chain length and packet sizes (100 and 75 bytes/packet) and a maximum chain size of 10 nodes.

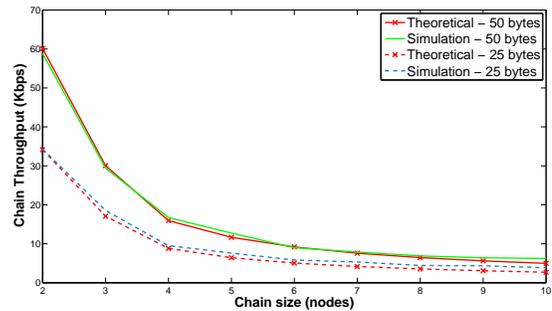


Figure 6. Theoretical vs simulated throughput for 50, 25 bytes/packet.

To evaluate the validity of (7) we also conduct simulations on larger chains of nodes, consisting of up to 100 nodes. Results are shown in Fig. 7 and 8.

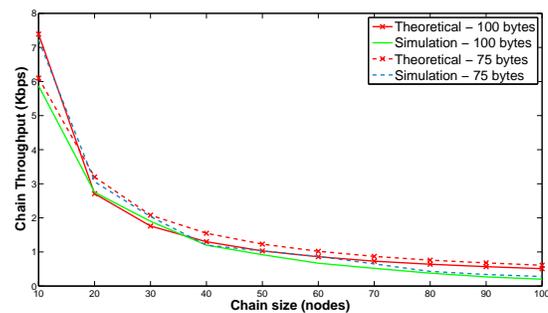


Figure 7. Theoretical vs simulated throughput for 100, 75 bytes/packet for chains of up to 100 nodes.

As the number of nodes increases, the achieved throughput tends to reach very low values. Packet size has a minimum effect to the chain throughput. The proposed model provides a close approximate to the simulated throughput.

To further examine the impact of configuration parameters, we conducted simulations for different BO=SO values, i.e., 5, 6, 8 and 9. Results not presented here, show that there

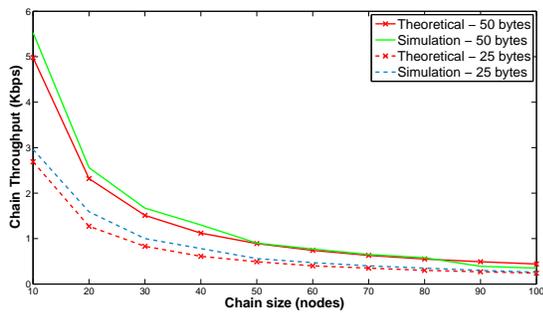


Figure 8. Theoretical vs simulated throughput for 50, 25 bytes/packet.

are no significant differences compared to those presented in this paper. These conclusions further support the validity of our model, since our analysis does not take into account the duration of the active period of a superframe, declared by the SO attribute. Simulation results match with theoretical values even when  $SO < BO$ .

#### IV. CONCLUSION

A key component of the integrated water management system is the chemical sensors subsystem and its monitoring capabilities. In this framework, we present a numerical model for throughput estimation on a chain of IEEE 802.15.4-compliant sensor nodes forming a cluster network. It provides a useful tool within the scope of an integrated water management project. The proposed numerical analysis and corresponding results will be used for the design, deployment and management of the chemical sensors module.

We evaluate our model through a series of simulations in order to check its accuracy and scalability under different packet and network sizes. The comparison between the numerically estimated and simulation values shows that there is a close match in the majority of the examined cases. This approach scales well with respect to network and packet size.

#### ACKNOWLEDGMENT

This work is elaborated through an on-going THALES project (CYBERSENSORS - High Frequency Monitoring System for Integrated Water Resources Management of Rivers). The Project has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Thales. Investing in knowledge society through the European Social Fund.

#### REFERENCES

- [1] D. N. Monekosso and P. Remagnino, "Data reconciliation in a smart home sensor network," *Expert Syst. Appl.*, vol. 40, no. 8, Jun. 2013, pp. 3248–3255.
- [2] J. Wu and H. Qin, "The design of wireless intelligent home system base on zigbee," in *Communication Technology, 2008. ICCT 2008. 11th IEEE International Conference on*, Nov. 2008, pp. 73–76.
- [3] M. Idoudi, H. Elkhorchani, and K. Grayaa, "Performance evaluation of wireless sensor networks based on zigbee technology in smart home," in *Electrical Engineering and Software Applications (ICEESA), 2013 International Conference on*, Mar. 2013, pp. 1–4.

- [4] D. Moraetis et al., "High-frequency monitoring for the identification of hydrological and bio-geochemical processes in a mediterranean river basin," *Journal of Hydrology*, vol. 389, no. 12, Jan. 2010, pp. 127 – 136.
- [5] Y. Kishino, Y. Sakurai, Y. Yanagisawa, T. Suyama, and F. Naya, "Svd-based hierarchical data gathering for environmental monitoring," in *Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication*, ser. *UbiComp '13 Adjunct*. New York, NY, USA: ACM, Sep. 2013, pp. 9–12.
- [6] F. Fernandez-Luque, J. Zapata, and R. Ruiz, "A system for ubiquitous fall monitoring at home via a wireless sensor network," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, Aug. 2010, pp. 2246–2249.
- [7] J. Ko et al., "Wireless sensor networks for healthcare," *Proceedings of the IEEE*, vol. 98, no. 11, Nov. 2010, pp. 1947–1960.
- [8] H. Yan, H. Huo, Y. Xu, and M. Gidlund, "Wireless sensor network based e-health system - implementation and experimental results," *Consumer Electronics, IEEE Transactions on*, vol. 56, no. 4, Nov. 2010, pp. 2288–2295.
- [9] K. Moirogiorgou et al., "High frequency monitoring system for integrated water resources management of rivers," in *Improving Efficiency of Water Systems in a Changing natural and financial Environment (EWaS-MED International Conference)*, Thessaloniki, Greece, Apr. 2013, pp. 1–6.
- [10] "Ieee standard for local and metropolitan area networks—part 15.4: Low-rate wireless personal area networks (lr-wpans)," *IEEE Std 802.15.4-2011 (Revision of IEEE Std 802.15.4-2006)*, 2011, pp. 1–314.
- [11] "Ieee standard for information technology— local and metropolitan area networks— specific requirements— part 11: Wireless lan medium access control (mac) and physical layer (phy) specifications amendment 8: Ieee 802.11 wireless network management," *IEEE Std 802.11v-2011 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, IEEE Std 802.11w-2009, IEEE Std 802.11n-2009, IEEE Std 802.11p-2010, and IEEE Std 802.11z-2010)*, 2011, pp. 1–433.
- [12] H. Holma and A. Toskala, *HSDPA/HSUPA for UMTS*. John Wiley, 2006.
- [13] M. Kohvakka, M. Kuorilehto, M. Hännikäinen, and T. D. Hämäläinen, "Performance analysis of ieee 802.15.4 and zigbee for large-scale wireless sensor network applications," in *Proceedings of the 3rd ACM International Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor and Ubiquitous Networks*, ser. *PE-WASUN '06*. New York, NY, USA: ACM, Oct. 2006, pp. 48–57.
- [14] C. K. Singh, A. Kumar, and P. M. Ameer, "Performance evaluation of an ieee 802.15.4 sensor network with a star topology," *Wirel. Netw.*, vol. 14, no. 4, Aug. 2008, pp. 543–568.
- [15] <http://www.isi.edu/nsnam/ns/>.
- [16] F. Gebali, *Analysis of Computer and Communication Networks*. Springer US, 2008.
- [17] P. Di Marco, P. Park, C. Fischione, and K. Johansson, "Analytical modelling of ieee 802.15.4 for multi-hop networks with heterogeneous traffic and hidden terminals," in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, Dec. 2010, pp. 1–6.
- [18] Y. Liu and Y. Jiang, *Stochastic Network Calculus*. London, UK: Springer, 2008.
- [19] J.-Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*. Berlin, Heidelberg: Springer-Verlag, 2001.
- [20] A. Koubaa, M. Alves, and E. Tovar, "Modeling and worst-case dimensioning of cluster-tree wireless sensor networks," in *Real-Time Systems Symposium, 2006. RTSS '06. 27th IEEE International*, Dec. 2006, pp. 412–421.

# Performance Enhancement of Heterogeneous Networks via Dense Clusterization and Higher Order Modulation

Ahmed Elhamy

Intel Corporation, Wireless Communications Lab  
Middle East Mobile Innovation Center (MEMIC)  
Cairo, Egypt  
ahmed.e@aucegypt.com

Ahmed S. Ibrahim

Electronics and Electrical Communications Department  
Cairo University  
Cairo, Egypt  
asibrahim@ieee.org

**Abstract**—Heterogeneous Networks are introduced in LTE-Advanced in order to fulfill the demanding necessity for more network capacity by the deployment of small cell nodes. In this paper, we study the dense deployment of small cells in hotspot to cover the users' data requests. The aim is to determine the optimum number of outdoor nodes per hotspot to achieve highest average user throughput. The results indicate that there is a range of optimum number of small cell nodes depending on the density of data traffic and the trade-off between available capacity and inter small cell interference. Introducing the small cells layer results in better channel conditions for some users that are close to small cells nodes. Thus, this encourages the usage of higher order modulation schemes like 256QAM. Adopting 256QAM infers that impairments caused by transmitter and receiver circuitry should be considered while evaluating the feasibility of 256QAM for outdoor small cells in heterogeneous networks. Results show that impairments have considerable negative effect on 256QAM throughput gains compared impairments-free system.

**Keywords**—Heterogeneous Networks; Higher Order Modulation; Outdoor Small Cells; 256QAM; Transmitter & Receiver Impairments.

## I. INTRODUCTION

Small Cells (SCs) and Heterogeneous Networks (HetNets) are considered one of the LTE-Advanced research topics that grasped interest in the recent time [1]. It is expected that the data rates requested by mobile subscribers will increase significantly in the coming years [2]. In addition to that, LTE-Advanced will need to satisfy new usage models that involve a giant hump in some services especially mobile data and video streaming. HetNets appears as a promising solution for LTE-Advanced to provide higher data rates and quality of service in the areas of high density of users.

HetNets are access networks consisting of multiple operating layers with different characteristics. The main layer is the Macro-layer that consists of Macrocell Nodes (MCNs) with high transmission power; and it is responsible for the coverage in an entire site. While the SCs layer, which is based on low-power nodes, is deployed in areas of high throughput requirements known as Hotspots.

The SCs layer nodes can be categorized based on density of deployment to sparsely dropped SCs or densely dropped SCs in hotspot areas [1]. SCs nodes also differ in range of coverage ranging from small apartments /offices (Femto Cells) supporting Closed Subscriber Group (CSG)

functionality [3][4] to larger areas like halls and airports (Pico Cells). Another dimension of classification is the deployment environment; since SC nodes can be deployed indoors or outdoors at lower heights compared to MCNs.

The deployment of SCs results in interference problems for User Equipment (UEs) in both layers [4][5] that should be suppressed by different interference mitigation algorithms, such as, ICIC, eICIC and FeICIC [6][7] and power control techniques [3][4].

In this paper, we focus on boosting UE throughput by enhancing the performance of HetNets with outdoor SCs. The enhancement is achieved using two aspects. The first aspect is increasing the density of clusterization of SCs in hotspots. The second aspect is the application of higher order modulations schemes like 256QAM for higher data rates at user terminals.

Dense clusterization of SCs is used at hotspots to provide closer serving nodes, i.e., better channel conditions, to more UEs for better network capacity and reducing traffic load on Macro layer. The disadvantage of the high density deployment of SCs is the non-negligible inter-SC interference. This harms Signal to Interference and Noise Ratio (SINR) between UEs and SC nodes resulting in throttled UE throughput. Thus, maintaining an optimum density of SC deployment is required to taper interference and to maximize throughput for UEs.

Being closer to UEs, less propagation losses to be suffered and better channel conditions and SINR are maintained between UEs and SCs nodes. Thus, higher performance could be achieved by introducing higher order modulation schemes like 256QAM to be an available option on the SC layer for UEs served by SCs. We study the gain achieved by 256QAM scheme in an outdoor HetNet environment and the negative effect of transmitter and receiver impairments on these gains.

The rest of the paper is organized as follows: Section II presents simulation environment and parameters. In Section III, the effect of the SCs density on user throughput is discussed. We introduce 256QAM modulation scheme and transmitter and receiver impairments modeling in Section IV. Section V provides the results of performance evaluation of supporting higher order modulation in the existence of transmitter and receiver impairments. Finally, Section VI concludes the paper.

## II. SIMULATION PARAMETERS

The simulated network layout is based on 7 hexagonal sites where each site is divided into three sectors as shown in Figure 1. Each sector has one hotspot area of 70m radius called cluster with 67% of UEs dropped in it and the rest of UEs are uniformly dropped in sector. The cluster has (N) SC nodes dropped uniformly within a radius of 50m from cluster center [8]. To model real cases, 80% of UEs are assumed to be indoor UEs suffering outdoor to indoor losses [8].

The Macro layer and SC layer operate in non-co-channel mode (no inter-layer interference). MCNs operate at center frequency of 2GHz. While on the other side, SCs operate at center frequency of 3.5 GHz. Bandwidth of 10 MHz is available at both center frequencies. Communication is fully downlink and stochastic channel models were used for macro layer and SCs layer [9].

Data traffic is modeled using bursty FTP traffic model 1 [10] as a closer traffic model to real traffic. In this model, one or more UEs are assumed to request data download of size 0.5 MB from the serving node at random time instants of a Poisson distribution (Exponential inter-request time periods). Simulation parameters are described in Tables I, II, III.

Simulations were held on a MATLAB-based System Level Simulator (SLS). The SLS models the propagation losses and the fast fading channel for the urban macrocell (UMa) and urban microcell (UMi) environments. Then, UEs are associated with the serving nodes based on Reference Signal Received Quality (RSRQ) which is calculated by

$$RSRQ = (N_{RB} * RSRP) / (RSSI) \quad (1)$$

where RSRP stands for Reference Signal Received Power,  $N_{RB}$  is the number of resource blocks and RSSI is the Received Signal Strength Indicator including received power from serving and interfering nodes and noise power [11].

RSRQ is used since it takes interference into consideration which is a decisive factor when HetNet layers operate at different frequencies and the interference profile differs between the macro layer and SC layer. RSRP (received power) as a basis for UE association is deceiving in the case of non-co-channel mode operation since you may get high received power from a serving node but at the same time suffer strong interference and deteriorated SINR. UE association process starts when UE request data download. It is worth noting that SC nodes that do not send data to any UEs, at association instant, are not considered when calculating interference part in RSRQ. This makes association decision depending on instantaneous interference profile derived from the instantaneous traffic in the network instead of assuming worst case one in which all nodes are assumed sending data and causing interference. Interference caused due to CRS signaling is also taken into consideration.

Then, SLS generates the channel matrices and the interference covariance matrices used in modeling received signal. The received signal by  $k^{\text{th}}$  UE ( $y_k$ ) is given by

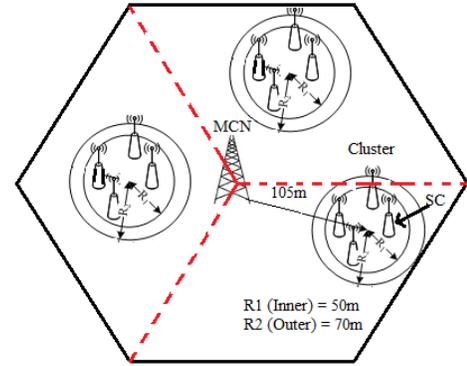


Figure 1. Network Layout for a Hexagonal Site

TABLE I. LAYOUT PARAMETERS

Parameter	Value
Deployment Grid	Hexagonal
Number of Sites / Cells	7 Sites / 21 Cells
Number of Clusters per Cell	1
Number of SCs per Cluster	N = [1:10]
Number of UEs per Cell	60
Percentage of UEs in Cluster	67%
Percentage of Indoor UEs	80%

TABLE II. SYSTEM PARAMETERS

Parameter	Value
Bandwidth (Macro/SCs)	10 MHz / 10 MHz
Carrier Frequency (Macro/SCs)	2 GHz / 3.5 GHz
Modulation	Up to 64QAM
Traffic Modeling	Non Full Buffer ( $\lambda=6,10$ )
File Size	0.5 MB
Scheduling	Proportional Fair with Outer-Loop Link Adaptation 0.15 dB
Receiver	Interference Aware Receiver (MMSE-IRC)
Cell Association	RSRQ

TABLE III. LINK & CHANNEL PARAMETERS

Parameter	Value
Macro Channel Model	ITU-UMa [9]
SCs Channel Model	ITU-UMi [9]
Antenna Configuration	2Tx X 2Rx , Cross Polarized
MIMO Mode	SU-MIMO with Adaptive Rank
Channel Estimation	Perfect
CRS Interference	Modeled – Alternative 2 [12]
UE Feedback	Code Book Based

$$y_k = \sqrt{\alpha_k} H_k W_k x_k + \sum_{i=1}^I (\sqrt{\alpha_i} H_i W_i x_i) + n \quad (2)$$

where ( $H_k, H_i$ ) represent the channel between  $k^{\text{th}}$  UE and the serving node and any of the (I) interfering nodes, respectively. ( $W_k$ ) is the precoding matrix by the serving node for the  $k^{\text{th}}$  UE. ( $W_i$ ) is the precoding matrix by the interfering nodes for other UEs. ( $x_k$ ) is the transmitted signals for  $k^{\text{th}}$  UE. On the other hand, ( $x_i$ ) is the transmitted signals for other UEs by interfering nodes. ( $\alpha_k$ ) represents the received power from the serving node and ( $\alpha_i$ ) represents the

received power from the  $i^{\text{th}}$  interfering node. Finally,  $n$  is the zero mean additive white Gaussian noise of the channel.

Then modulation order is chosen based on the instantaneous SINR ( $\gamma$ ) calculated by

$$\gamma = \frac{\alpha_k \left\| \sum_{n=1}^{N_t} H_k W_k(:,n) \right\|^2}{\sum_{i=1}^I \left( \alpha_i \left\| \sum_{n=1}^{N_t} H_i W_i(:,n) \right\|^2 \right) + \sigma_n^2} \quad (3)$$

Afterwards, throughput is calculated by summing Transport Block Sizes (TBSs) (The higher the modulation order, the bigger the TBS sent to the UE) transferred on the downlink excluding dropped blocks to consider block error rate.

### III. DENSE CLUSTERIZATION OF SMALL CELLS

Hotspot areas of the macrocell are congested with high density of UEs of huge request of data. Such requests are satisfied by deploying a cluster of SCs to provide an acceptable quality of service for these UEs.

The tradeoff appears upon the decision of the density of SCs deployed per hotspot. Increasing the SCs will provide more resources for UEs. The advantages of this approach are the decrease in the waiting or blocking probability and the system will usually not suffer resources starvation or high latency. On the other hand, increasing the number of SCs in hotspot leads to high interference among SC (inter-SC interference) and a reduction in the received power at UEs.

The resources needed to keep the system functional under given traffic conditions enforce a lower bound on the number of SCs to be deployed. While, the inter-SC interference caused by dense deployment of SCs enforces an upper bound constraint on the number of SCs to be deployed in a hotspot area. In addition to that, other factors may play a role in the tradeoff like the cost and the power consumption.

In this section, our target is to define an optimum density of SCs in a hotspot between the two bounds. At this optimum point, highest UE throughput is achieved due to tolerable inter-SC interference and the existence of enough resources to fulfill UEs requests.

We held system level simulations and set the environment as described in Section II. The number of SCs per cluster is varied in the range from 1 to 10 in order to find the optimum density of SCs per cluster to serve hotspot UEs. The optimum density of SC nodes in a cluster is achieved when all UEs (MCN UEs and SC UEs) have their average throughput maximized considering different traffic loads. All UEs are considered since the ratio of association is not constant when the numbers of SCs per cluster is changed due to the change in interference profile considered in RSRQ association.

To achieve this target, average UE throughput is measured for all UEs at different number of SCs per cluster and at different non full buffer traffic loads ( $\lambda = 6, 10, 16$  user requests for file download per second per geographical cell).

By applying this criterion of throughput maximization, the number of SCs for good performance depends on traffic load. For low traffic load ( $\lambda = 6$ ), the maximum throughput occurs when hotspot has 4 SCs as shown in Table IV.

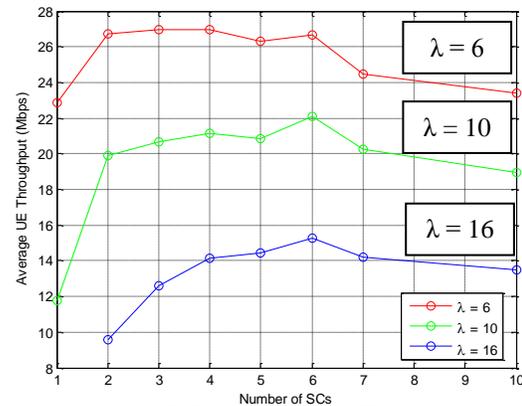


Figure 2. Mean UE Throughput for all UEs vs. Number of SCs

TABLE IV. DENSE CLUSTERIZATION OF SMALL CELLS RESULTS

N SCs	All UEs Throughput (Mbps) $\lambda = 6$	All UEs Throughput (Mbps) $\lambda = 10$	All UEs Throughput (Mbps) $\lambda = 16$
1	22.87	11.74	N/A
2	26.74	19.90	9.56
3	26.96	20.70	12.62
4	<b>26.98</b>	21.13	14.17
5	26.31	20.85	14.45
6	26.67	<b>22.10</b>	<b>15.26</b>
7	24.48	20.23	14.18
10	23.41	18.93	13.51
Mean Resource Utilization (%)	8.58%	15.6%	22.5%

The throughput provided by 2 or 3 SCs is also close to the maximum so it may be a good option of cutting cost without big loss in performance ( $< 1\%$ ). At medium traffic load ( $\lambda = 10$ ), we can notice the traverse in the peak throughput to 6 SCs per hotspot. This is expected with the increase of traffic where the need of resources becomes more urgent than need for a good SINR. This conclusion is confirmed for heavy traffic ( $\lambda = 16$ ) with 6 SCs per hotspot is the optimum choice for SC density in this case. For heavy traffic, low number of SCs provide very low throughput like the case of 2 SCs or cannot handle the UE requests like the case of 1 SC.

Looking at the extremes, with only 1SC per hotspot, system cannot handle heavy traffic and it suffers lack of resources at low and medium traffic cases. While with 10 SCs per hotspot, the inter-SC interference increases and becomes very dominant and outweighs the capacity added to the network by adding SCs.

Figure 2 presents the average UE throughput at different number of SCs per cluster at different traffic loads. The optimum number of deployed SC per hotspot varies with the traffic load. With the increase in traffic load, we can note that there is a need for more SCs to provide more resources to serve UEs while the overall performance of the network decreases. At low traffic loads, resources of small number of SCs are enough to fulfill UEs demands and the dominant factor of performance is the inter-SC interference that needs to be reduced. At a given constant traffic load, the

optimum density is the one that achieves a balance between inter-SC interference and available resources for serving UEs. When performance is very close for a range of SC densities, like the case of low traffic where 2-6 SCs can provide almost same performance, the choice becomes mainly dependent on cost and the least density is the most attractive choice.

#### IV. HIGHER ORDER MODULATION & TRANSCIEVER IMPAIRMENTS

##### A. 256QAM

In LTE and LTE-Advanced releases (Rel.8-11), QPSK, 16QAM and 64QAM have been the main Modulation and Coding Schemes (MCSs) [13]. With the introduction of HetNets and SCs, UEs associated with SC nodes were shown to have better channel characteristics and lower propagation losses. Hence, higher order modulations schemes like 256QAM can be utilized to improve performance [14][15]. 256 QAM constellation points can be defined as:

$$S_{xy} = \sqrt{\frac{1}{170}}(a_x + j b_y) \quad (4)$$

Where

$$\{a_x, b_y\} = \begin{bmatrix} -15,15 & -13,15 & \dots & 15,15 \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ -15,-15 & -13,-15 & \dots & 15,-15 \end{bmatrix} \quad (5)$$

We used link level simulations to produce an LTE physical layer abstraction table with 10 additional MCSs of multiple coding rates and their corresponding TBSs to be inserted in the SLS. These MCSs are being picked when reported SINR in feedback exceeds 20 dB [15]. Figure 3 shows the spectral efficiencies of the added MCSs to support 256QAM in the SLS which is higher compared to all MCSs of lower modulation order [15].

##### B. Transmitter Impairments

Transmitter includes many circuit parts that cause imperfections in the transmitted symbols. The main sources of these imperfections are the transmitter filter and the clipping and non-linearities of the power amplifier [16]. Thus, there is a deviation of the real transmitted symbol compared to ideal transmitted symbol. The percentage of this deviation with respect to transmission power is known as the Transmitter Error Vector Magnitude (TX-EVM). These imperfections are usually modeled as a zero mean additive white Gaussian noise at the transmitter with variance  $\sigma_{tx}^2$  [16].

$$\sigma_{tx}^2 = \epsilon_{tx} * P_t \quad (6)$$

where,  $P_t$  is the average transmitted power and  $\epsilon_{tx}$  is (TX-EVM%)<sup>2</sup>

Since TX-EVM noise is applied at the transmitter side, it experiences the fast fading channel like the transmitted signals.

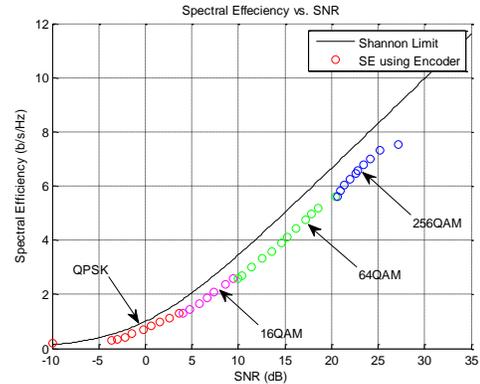


Figure 3. Spectral Efficiency of 256QAM MCS

##### C. Receiver Impairments

Receiver circuitry causes imperfections in symbols decoding as well. The receiver impairments causing factors are mainly the receiver local oscillator phase noise, receiver dynamic range, I/Q imbalance, carrier leakage and carrier frequency offset [16]. The receiver impairments are expressed in the value of Receiver Error Vector Magnitude (RX-EVM). RX-EVM is the percentage of deviation of the real received symbols compared to the ideal received symbol assuming perfect receiver. It is a noise that does not carry channel fading characteristics but it is proportional to the average received power. RX-EVM is modeled similar to TX-EVM as a zero mean additive white Gaussian noise at the receiver with variance  $\sigma_{rx}^2$  [16]

$$\sigma_{rx}^2 = \epsilon_{rx} * P_r \quad (7)$$

where,  $P_r$  is the average received power and  $\epsilon_{rx}$  is (RX-EVM%)<sup>2</sup>

Both TX-EVM and RX-EVM have their negative effect on the final received signal ( $y$ ) at a given UE:

$$y_k = \sqrt{\alpha_k} H_k W_k (x_k + n_{tx}) + \sum_{i=1}^I (\sqrt{\alpha_i} H_i W_i x_i) + n + \sqrt{\alpha_k} n_{rx} \quad (8)$$

The SINR ( $\gamma$ ) is reduced by adding to the interference and noise part in the denominator as shown

$$\gamma = \frac{\alpha_k \left\| \sum_{n=1}^{N_t} H_k W_k (:,n) \right\|^2}{\sum_{i=1}^I \left( \alpha_i \left\| \sum_{n=1}^{N_t} H_i W_i (:,n) \right\|^2 \right) + \sigma_n^2 + \sigma_{n_{tx}}^2 \left\| \sum_{n=1}^{N_t} H_k W_k (:,n) \right\|^2 + \alpha_k \sigma_{n_{rx}}^2} \quad (9)$$

$n_{tx}$  and  $n_{rx}$  are the noise modeling transceiver impairments.

Utilizing higher order modulations such as 256QAM, tolerable imperfections at the transmitter and receiver are reduced since the distance between constellation points are shorter. For example, in 64QAM modulation, TX-EVM is assumed to be 8% [8]. While for 256QAM modulation, TX-EVM is assumed in the range (3% - 6%) [8]. There is no determined value for RX-EVM but in most of research they are assumed in the same range of values of TX-EVM or less (1.5% - 4%) [8].

Transceiver impairments limit the performance of the system so they should be considered while evaluating the throughput gains due to enabling higher order modulation schemes such as 256QAM.

## V. PERFORMANCE EVALUATION RESULTS

To evaluate the performance enhancement due to the application of 256QAM, a HetNet with one cluster of 4 outdoor SC nodes is simulated in the SLS with the same simulation parameters described in section II. 256QAM modulation is provided by SCs only. MCNs do not need to support 256QAM since macro channel conditions do not allow notable utilization of higher order modulation and 64QAM is sufficient.

The downlink data traffic is modeled using a bursty FTP traffic model 1. The performance metric for evaluation is the average per user throughput for SC associated UEs (Two thirds of UEs in most cases) who are candidates to utilize 256QAM and the average per user throughput for all UEs. The throughput was calculated in two cases, Case 1 with no impairments at transmitter and receiver. In Case 2, both transmitter and receiver impairments are taken into consideration. EVM percentages and traffic parameters are described in Table V. The throughput with 256QAM supported is compared against the throughput with only up to 64QAM supported to quantify the enhancement in all cases in terms of percentage gain at given utilization of resources (Amount of exploited Resource Blocks (RBs) out of all available RBs over the whole simulation time).

Instantaneous SINR CDFs for Cases 1, 2 are shown in Figure 4 and Figure 5, respectively. There is a good potential for 256QAM in Case 1 since about 22-27% of file transfers have enough SINR ( $\text{SINR} > 20\text{dB}$  [15]) and can utilize 256QAM at all simulated traffic rates. It also indicates that with the increase of traffic, SINR decreases and hence less file transfers can occur with 256QAM. For Case 2, the percentage of UEs who have promising channel conditions for file transfer using 256QAM decreases to about 21-23%.

In Case 1 (with no impairments), support of 256QAM by the SCs layer nodes results in average per user throughput gains for UEs served by SCs ranging from 6% to 9% at SC resource utilizations in the range from 6% to 11%. It is worth noting that, as traffic load decreases, throughput and gains tend to increase as shown in Table VI by comparing gains for ( $\lambda = 14$ ) to lower values.

In Case 2 (with both transmitter and receiver impairments), the gains due to the support of 256QAM are reduced. They are ranging from 2% to 3% for SC associated UEs as shown in Table VII. This is expected due to the increase of the modeled deviation in transmitted and received symbols by having impairments at both transmitter and receiver. This is reflected in terms of more noise impacting the system as can be found in (8). So, SINR  $\gamma$  decreases and throughput degrades as can be deduced from (9). Effect of impairments can be seen in Figure 5 where instantaneous SINR is capped at 30 dB [8].

This degradation can also be noticed from comparing absolute values of throughput at different traffic load levels. The absolute throughput at any given traffic load decreases when TX, RX impairments are added. Simulations also shown that impairments also decreases the chances of using 256QAM (256QAM Utilization Ratio) compared to no impairments case as shown in Table VIII

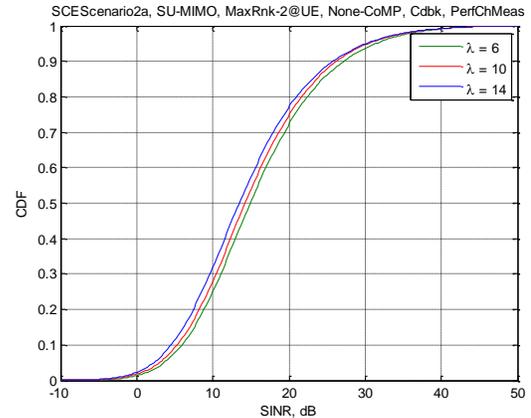


Figure 4. Instantaneous SINR for SCs UEs in Case 1

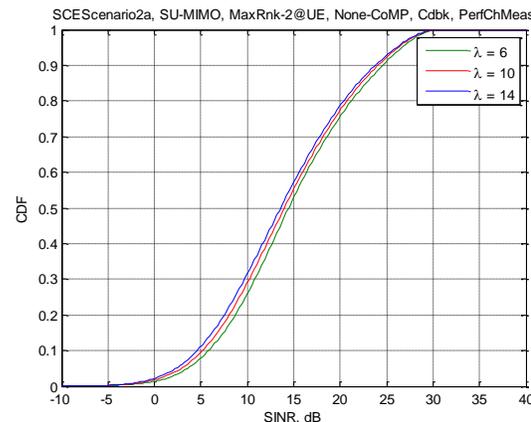


Figure 5. Instantaneous SINR for SCs UEs in Case 2

TABLE V. PERFORMANCE EVALUATION PARAMETERS

Parameter	Value
Traffic Modeling	FTP Traffic Model 1 ( $\lambda = 6, 10, 14$ )
TX-EVM	8% for Macrocell 3% for SCs
RX-EVM	1.5%
Number of SCs	4 SCs/Cluster (1 Cluster/Cell)

TABLE VI. 256QAM PERFORMANCE EVALUATION RESULTS (NO IMPAIRMENTS)

User Arrival Rate ( $\lambda$ ) (Sec <sup>-1</sup> )	Small Cell UEs		All UEs	
	Average User Throughput in Mbps (Gain %)	% Resource Utilization	Average User Throughput in Mbps (Gain %)	% Resource Utilization
6	27.45 (5.56%)	5.95%	28.17 (4.39%)	7.43%
10	22.61 (8.75%)	10.86%	22.75 (7.67%)	13.10%
14	18.02 (7.24%)	17.05%	17.43 (5.50%)	20.08%

TABLE VII. 256QAM PERFORMANCE EVALUATION RESULTS (TX-EVM = 3% & RX-EVM = 1.5%)

User Arrival Rate ( $\lambda$ ) (Sec <sup>-1</sup> )	Small Cell UEs		All UEs	
	Average User Throughput in Mbps (Gain %)	% Resource Utilization	Average User Throughput in Mbps (Gain %)	% Resource Utilization
6	26.58 (2.79%)	6.12%	26.42 (1.60%)	7.63%
10	21.51 (3.02%)	11.20%	21.37 (3.78%)	13.62%
14	16.90 (2.67%)	17.46%	16.22 (1.77%)	20.67%

TABLE VIII. 256QAM UTILIZATION RATIO

$\lambda$ (Sec <sup>-1</sup> )	Case 1 No Impairments	Case 2 With Tx & Rx Impairments
6	15.19%	12.62%
10	15.08%	12.05%
14	13.60%	10.67%

From a practical point of view, 256QAM as a higher order modulation technique is promising for SCs with low density of users (especially indoor femtocells). With moving outdoors and increasing UE density, the gains decrease.

## VI. CONCLUSION

In this paper, a HetNet with outdoor SCs layer was simulated to study two main aspects. The first aspect is the effect of the dense clusterization of SCs in a hotspot of 70m radius on throughput and performance. We found that the optimum number of SCs for deployment for tolerable level of inter-SC interference and maximized performance depends on traffic. For low traffic, it is 2-4 SCs while we need 6 SCs for medium and high traffic. It was found that with the increase in traffic rate, more SCs are needed to provide more resources and handle UEs requests. On the other hand, at low traffic rates, inter-SC interference is dominant and the needed number of SC decreases. Another factor affects decision when the performance is constant over a range of cluster densities which is cost and least density in the range is preferred.

The second aspect was to evaluate the performance gains resulting from supporting 256QAM for outdoor SC layer in HetNets. It was found that about 6%-9% average throughput gains per UE could be achieved for SC UEs. While, introducing imperfections at the transmitter and receiver sides leads to a decrease in throughput and performance gains of 256QAM compared to traditional 64QAM. Only 2%-3% increase in throughput occur in this case.

## ACKNOWLEDGMENT

The authors would like to thank Shady Elbassiouny, Ahmed Darwish, Dr. Hani Elgebaly, Dr. Alexei Davydov and Dr. Jong-Kae Fwu, Intel Wireless Communications lab, for their cooperation and constructive discussions.

## REFERENCES

- [1] 3GPP TR36.932 v.12.1.0, "Scenarios and requirements for small cell enhancements for E-UTRA and E-UTRAN (Release 12)," 2013.
- [2] "Cisco visual networking index: Global mobile data traffic forecast update, 2012-2017," February 2013, pp.10.
- [3] A. S. Ibrahim, A. Darwish, E. A. Makled, and H. El Gebaly, "Adaptive interference mitigation techniques for femtocells," IEEE Wireless Communications and Networking Conference, WCNC, Apr. 2012, pp. 1218 - 1223.
- [4] E. Makled, A. S. Ibrahim, A. Darwish, and H. El Gebaly, "Nonunanimous power inter-cell interference coordination in heterogeneous networks," IEEE Vehicular Technology Conference (VTC'12-Spring), May 2012, pp. 1 - 6.
- [5] M. Shoeb, A. S. Ibrahim, and H. El Gebaly, "Coordinated multi-point algorithms in pico based heterogeneous network," IEEE Advanced Communication Technology (ICACT), Feb. 2012, pp. 37 - 40.

- [6] D. Kimura and H. Seki, "Fujitsu: Inter-Cell interference coordination (ICIC) Technology," Jan. 2012, pp. 89-94.
- [7] D. Tse and P. Viswanath, "Fundamentals of Wireless Communication", Cambridge University Press, 2005.
- [8] 3GPP TR36.872, "Small cell enhancements for E-UTRA and E-UTRAN - Physical layer aspects (Release 12)", Dec. 2013.
- [9] ITU-R M.2135-1, "Guidelines for evaluation of radio interface technologies for IMT-Advanced," 2009.
- [10] 3GPP TR36.814 v.9.0.0, "Further advancements for E-UTRA physical layer aspects (Release 9)," 2010.
- [11] 3GPP TS136.214 v.11.1.0, "Physical layer measurements (Release 11)," 2013.
- [12] R1-112856, "Summary of ad hoc session on FeICIC simulation assumptions", NTT DOCOMO.
- [13] 3GPP TS136.213 v.11.2.0, "Physical layer procedures (Release 11)," 2013.
- [14] E. Lähetkangas, K. Pajukoski, E. Tirola, J. Hämäläinen, and Z. Zheng, "On the performance of LTE-Advanced MIMO: How to set and reach beyond 4G targets," IEEE European Wireless, Apr. 2012, pp. 1-6.
- [15] S. O. Elbassiouny and A. S. Ibrahim, "Link level performance evaluation for higher order modulation," IEEE IWCMC, submitted and available for review, Aug. 2014.
- [16] B. Goransson, S. Grant, E. Larsson, and Z. Feng, "Effect of transmitter and receiver impairments on the performance of MIMO in HSDPA," IEEE Signal Processing Advances in Wireless Communications (SPAWC), Jul. 2008, pp. 496-500.

# Using GSPNs for Performance Evaluation of Networks with Repeated Calls and Different Vacation Policies

Nawel Gharbi

Computer Science Department

University of Sciences and Technology, USTHB

Algiers, Algeria

Email: ngharbi@wissal.dz

**Abstract**—This paper deals with retrial systems where servers are subject to random vacations. So far, these systems were analyzed only by queueing theory and almost works assuming that the service station consists of one server and the customers source is infinite. In this paper, we give a detailed performance analysis of finite-source multiserver networks with repeated calls of blocked customers and multiple or single vacations of servers or all station, using Generalized Stochastic Petri nets. We show how this high level stochastic model allows us to cope with the complexity of such networks involving the simultaneous presence of retrials and vacations, and how stationary performance indices can be expressed as a function of Petri net elements.

**Keywords**—Repeated calls; Finite-source; Vacation policies; Generalized Stochastic Petri nets; Modeling and Performance measures.

## I. INTRODUCTION

Models with repeated calls describe operation of many computer networks and telecommunication systems, e.g., call centers, cellular mobile networks [1][2][3][4] and wireless sensor networks [5]. Systems with repeated attempts are characterized by the following feature: When an arriving customer finds all servers (resources) busy or unavailable, is not put in a queue, but joins a virtual pool of blocked customers called orbit, and will repeat the request to try again to reach the servers after a random delay. Significant references reveal the non-negligible impact of repeated calls, which arise due to a blocking in a system with limited capacity resources or due to impatience of users. There has been a rapid growth in the literature on the queueing systems with repeated attempts (also called retrial queues). For a recent summary of the fundamental methods, results and applications on this topic, the reader is referred to [6][7] and [8].

In this paper, we consider multiserver retrial systems in which each server sometimes takes a vacation, i.e., becomes unavailable to the primary and repeated calls for a random period of time. These vacation periods are usually introduced in order to exploit the idle time of the servers for other secondary jobs as: servicing customers of another system, inspection tasks and preventive maintenance actions which are mainly doing to prevent the risk of failure, to preserve

the sanity of the system, to provide a high reliability and to improve the quality of service. Similarly, the servers breakdowns which may occur randomly, and the repair periods, may be regarded as servers vacations.

A wide class of policies for governing the vacation mechanism, have been discussed in the literature, namely the multiple vacation policy and the single vacation policy. Other studies have considered synchronous vacations of some servers or all the station servers (station vacation). On the other hand, multi-server vacation models were mainly studied in the past decade. Zhang et al. [9][10] studied the multi-server models with either single vacation or multiple vacations. Later, Lin et al. [11] analyzed the multi-server model with working vacations. Ke et al. [12] studied the optimal threshold policies in a finite buffer multi-server vacation model with unreliable servers. Recently, Ke et al. [13] consider a multi-server queueing system with multi-threshold vacation policy and servers breakdowns. Excellent surveys on the vacation models have been reported by Doshi [14], Takagi [15], Tian et al. [16], and recently, by Ke et al. [17].

The main reason for the growing interests in multiple-server vacation models is because they can realistically represent some service/manufacturing systems and computer/telecommunication networks. However, all these works on multi-server vacation queueing models, assume that the customers source is infinite and do not take into account the repeated calls of blocked customers.

In retrial systems with vacations, customers who arrive while all servers are busy or on vacation, have to join the orbit to repeat their call after a random period. Thus, there is a natural interest in the study of this class of models, which has been used in concrete applications as digital cellular mobile networks [18], local area networks with nonpersistent CSMA/CD protocols [19], with star topology [20] and so on. However, almost works combining retrial and vacation phenomenon, assume that the service station consists of one single server and the customers source is infinite [20][21][22]. On the other hand, in all the works cited above, the retrial systems with vacations are analyzed only by the queueing theory.

In this paper, we propose the applicability of Generalized Stochastic Petri nets formalism (GSPNs) for modeling and performance evaluation of networks with repeated calls of blocked customers and servers vacations. To this end, we consider different vacation policies, namely the single and multiple vacations of servers or all the service station.

The paper is organized as follows: First, we describe the systems under study. In Section 3, we present the GSPN models describing multiserver retrial systems with station and server vacations mechanisms and under multiple and single vacation policy. Performance indices are given in Section 4. Next, several numerical examples are presented with some comments and discussions. Finally, we give a conclusion.

## II. DESCRIPTION OF RETRIAL NETWORKS WITH DIFFERENT VACATION POLICIES

In the analysis of retrial systems with vacations, it is usually assumed that the customers source is infinite. However, in many practical situations, it is very important to take into account the fact that the rate of generation of new primary calls decreases as the number of customers in the system increases. This can be done with the help of finite-source retrial models where each customer generates its own flow of primary demands.

In this paper, we consider retrial systems with finite source (population), that is, we assume that a finite number  $K$  of potential customers generate the so called quasi-random input of primary calls with rate  $\lambda$ . Each customer can be in three states: generating a primary call (free), sending repeated calls (in orbit) or under service by one of the servers.

If a customer is free at time  $t$ , it can generate a primary request for service in any interval  $(t, t+dt)$  with probability  $(K-n)\lambda dt + o(dt)$  as  $dt \rightarrow 0$ , where  $n$  is the number of customers in the system. Each customer requires to be served by one and only one server.

The service station consists of  $c$  ( $c \geq 1$ ) homogeneous and parallel servers. Each server can be idle, busy or on vacation. If one of the servers is idle at the moment of the arrival of a call, then the service starts. The requests are assigned to the free servers randomly and without any priority order. The service times are independent, identic and exponentially distributed with rate  $\mu$ . After service, the customer becomes free, so it can generate a new primary call, and the server becomes idle.

We consider the two vacation mechanisms: *server vacation* and *station vacation*. For the first one, which is encountered even more often in practice, each server is an independent working unit, and it can take its own vacation independently of other servers states. In the model with station vacation mechanism, ALL the servers take vacations simultaneously. That is, whenever the system is empty, all the station leaves the system for a vacation, and returns

when the vacation is completed. So, station vacation is group vacation for all servers. This occurs in practice, for example, when a system consists of several interconnected machines that are inseparable, or when all the machines are run by a single operator. In such situations, the whole station has to be treated as a single entity for vacation. Hence, if the system (or the operator who runs the system) is used for a secondary task when it becomes empty (or available), all the servers (the operator) will then be utilized to perform a secondary task. During this amount of time, the servers are unavailable to serve any primary or repeated call and this is equivalent to taking a station vacation.

The *exhaustive service discipline* is considered here. That is, each free server (or all station) can take a vacation only if the system is empty at either a service completion or at the end of a vacation, and only at these epochs. On the other hand, upon completing a vacation, the server returns to the idle state and starts to serve customers, if any, till the system becomes empty. Otherwise, if the server (or the station) at the moment of returning from vacation, finds the system empty, it takes one of the two actions:

- Under the *multiple vacation policy*, the server (station) shall leave immediately for another vacation and continues in this manner until he finds at least one customer (not being served) in the system upon returning from a vacation.
- Under the *single vacation policy*, the server (station) should wait until serving one call at least before commencing another vacation.

The vacation times of all servers (or station) are assumed to be independent and exponentially distributed with rate  $\theta$ .

At the moment of the arrival of a call, if all the servers are busy or on vacation, the customer joins the orbit to repeat his demand after an exponential time with parameter  $\nu$ .

As usual, we assume that the interarrival periods, service times, vacation times and retrial times are mutually independent.

## III. GSPN MODELS OF MULTISERVER RETRIAL SYSTEMS WITH VACATIONS

In this section, we present our approach for modeling finite-source multiserver retrial systems with station and server vacations, under multiple and single vacation policies using the generalized stochastic Petri nets model.

A GSPN is a directed graph that consists of places (drawn as circles), timed transitions (drawn as rectangles) which describe the execution of time consuming activities and immediate transitions (drawn as thin bars) that model actions whose duration is negligible, with respect to the time scale of the problem. This class of transitions has priority over timed transitions and fire in zero time once they are enabled.

Formally, a GSPN [24] is an eight-tuple  $(P, T, W^-, W^+, W^h, \pi, M_0, \theta)$  where :

- $P = \{P_1, P_2, \dots, P_n\}$  is the set of places;



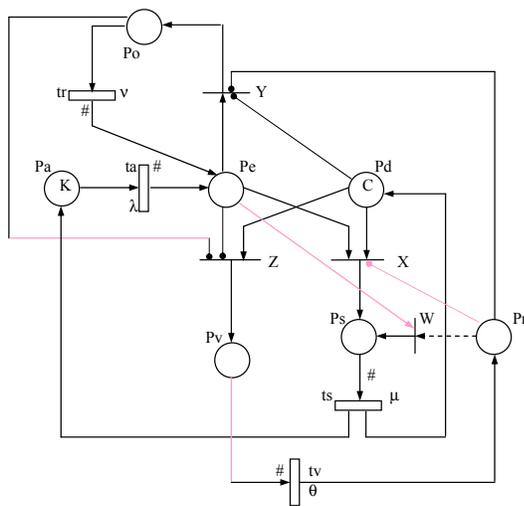


Figure 2. GSPN model of retrial systems with single vacations of servers

transition  $Z$  and transition  $t_v$  to place  $P_d$  equals  $c$  (rather than 1), because the  $c$  servers of the station take a vacation together. So, if the place  $P_d$  contains  $c$  idle servers, the orbit ( $P_o$ ) is empty and there is no arrival to the place  $P_e$ , the immediate transition  $Z$  fires, which represents the begin of the station vacation time. At the end of this period (after a mean delay equals  $1/\theta$ ),  $c$  tokens corresponding to the  $c$  servers of the station will be deposited in  $P_d$ .

### C. Retrial systems with single vacations of servers

This model corresponds to systems where each server is an independent working unit. The single vacation policy means that at the end of a vacation period, even if the system is empty, the server is obliged to wait until serving one call at least, before commencing another vacation.

Fig. 2 shows the GSPN model describing the above system.

In the previous models with multiple vacations, the place  $P_d$  contains all the free servers. Hence, at the end of a service or vacation period, the server returns to the idle state represented by the place  $P_d$ . However, in the model with single vacations given in Fig. 2, at a service completion, the server joins the place  $P_d$  which contains the servers having served at least one call since the last vacation period. So, they can serve other calls if any (firing of transition  $X$ ). Otherwise, they can take a vacation after the firing of the immediate transition  $Z$ . However, at the end of a vacation period, the server joins the place  $P_r$  which represents the servers having just finished a vacation. Hence, the servers of  $P_r$  are obliged to serve at least one call after the firing of the immediate transition  $W$  to join the place  $P_d$ , where they can commence another single vacation.

Initially, all customers are free, the orbit is empty and the  $c$  servers are available to serve the calls or to take a vacation.

At the arrival of a primary or repeated request to the place  $P_e$ , several alternatives are possible:

- If the place  $P_r$  of servers just returning from vacation, contains at least one server, the immediate transition  $W$  fires and the service of the arriving call starts.
- If the place  $P_r$  is empty and the place  $P_d$  contains at least one free server, the immediate transition  $X$  fires and the service period starts.
- If the two places  $P_d$  and  $P_r$  are empty which represents the fact that all the servers are busy or on vacation, the immediate transition  $Y$  fires and a token will be deposited in the place  $P_o$ . So, the customer joins the orbit.

### D. Retrial systems with single vacations of the station

The GSPN modeling systems with single vacations of the station is the same as the model given in Fig. 2, in which the multiplicity of the arc connecting the place  $P_d$  to transition  $Z$  and transition  $t_v$  to place  $P_r$  equals  $c$  (rather than 1), because the  $c$  servers of the station take a vacation together. At the end of this period,  $c$  tokens corresponding to the  $c$  servers of the station will be deposited in  $P_r$ . Hence, the station can't take another vacation until each server serves at least one call.

## IV. PERFORMANCE MEASURES

The aim of this section is to derive the formulas of the most important stationary performance indices. As, all the proposed models are bounded and the initial marking is a home state, the underlying continuous time Markov chains are ergodic for the different vacation policies. Hence, the steady-state probability distribution vector  $\pi$  exists and can be obtained as the solution of the linear system of equations  $\pi \cdot Q = 0$  with the normalization condition  $\sum_i \pi_i = 1$ , where  $\pi_i$  denotes the steady-state probability that the process is in state  $M_i$  and  $Q$  is the transition rates matrix. Having the probabilities vector  $\pi$ , several stationary performance indices of small cell wireless networks with different vacation policies can be derived as follows. In these formulas,  $M_i(p)$  denotes the number of tokens in place  $p$  in marking  $M_i$ ,  $A$  the set of reachable tangible markings, and  $A(t)$  is the set of tangible markings reachable by transition  $t$  and  $E(t)$  is the set of markings where the transition  $t$  is enabled.

- The mean number of customers in orbit ( $n_o$ ):

$$n_o = \sum_{i: M_i \in RS} M_i(P_o) \cdot \pi_i \quad (1)$$

- The mean number of busy servers ( $n_s$ ):

$$n_s = \sum_{i: M_i \in RS} M_i(P_s) \cdot \pi_i \quad (2)$$

- The mean number of customers in the system ( $n$ ):

$$n = n_s + n_o \quad (3)$$

- The mean number of servers on vacation ( $n_v$ ):

$$n_v = \sum_{i:M_i \in RS} M_i(P_v) \cdot \pi_i \quad (4)$$

- The mean number of idle servers ( $n_f$ ):

$$n_f = c - (n_s + n_v) \quad (5)$$

$$= \begin{cases} \sum_{i:M_i \in RS} M_i(P_d) \cdot \pi_i, \\ \text{in multiple vacations,} \\ \sum_{i:M_i \in RS} [M_i(P_d) + M_i(P_r)] \cdot \pi_i, \\ \text{in single vacations.} \end{cases} \quad (6)$$

- The mean rate of generation of primary calls ( $\bar{\lambda}$ ):

$$\bar{\lambda} = \sum_{i:M_i \in E(t_a)} M_i(P_a) \cdot \lambda \cdot \pi_i \quad (7)$$

- The mean rate of generation of repeated calls ( $\bar{\nu}$ ):

$$\bar{\nu} = \sum_{i:M_i \in E(t_r)} M_i(P_o) \cdot \nu \cdot \pi_i \quad (8)$$

- The mean rate of service ( $\bar{\mu}$ ):

$$\bar{\mu} = \sum_{i:M_i \in E(t_s)} M_i(P_s) \cdot \mu \cdot \pi_i \quad (9)$$

- The mean rate of vacation ( $\bar{\tau}$ ):

$$\bar{\tau} = \sum_{i:M_i \in E(t_v)} M_i(P_v) \cdot \theta \cdot \pi_i \quad (10)$$

- The blocking probability of a primary call ( $B_p$ ):

$$B_p = \begin{cases} \frac{\sum_{i:M_i \in RS} \sum_{j=1}^K j \cdot \lambda \cdot \text{Prob}[M_i(P_a)=j \& M_i(P_d)=0]}{\bar{\lambda}}, \\ \text{in multiple vacations,} \\ \frac{\sum_{i:M_i \in RS} \sum_{j=1}^K j \cdot \lambda \cdot P^{(i,j)}}{\bar{\lambda}}, \\ \text{in single vacations.} \end{cases} \quad (11)$$

where:

$$P^{(i,j)} = \text{Prob}[M_i(P_a) = j \& M_i(P_d) = 0 \& M_i(P_r) = 0].$$

- The blocking probability of a repeated call ( $B_r$ ):

$$B_r = \begin{cases} \frac{\sum_{i:M_i \in A} \sum_{j=1}^K j \cdot \nu \cdot \text{Prob}[M_i(P_o)=j \& M_i(P_d)=0]}{\bar{\nu}}, \\ \text{in multiple vacations,} \\ \frac{\sum_{i:M_i \in A} \sum_{j=1}^K j \cdot \nu \cdot P^{(i,j)}}{\bar{\nu}}, \\ \text{in single vacations.} \end{cases} \quad (12)$$

where:

$$P^{(i,j)} = \text{Prob}[M_i(P_o) = j \& M_i(P_d) = 0 \& M_i(P_r) = 0].$$

- The blocking probability ( $B$ ):

$$B = B_p + B_r \quad (13)$$

- The admission probability ( $A$ ):

$$A = 1 - B \quad (14)$$

- Utilization of  $s$  servers ( $U_s$ ): ( $1 \leq s \leq c$ )

$$U_s = \sum_{i:M_i(P_s) \geq s} \pi_i \quad (15)$$

- Vacation of  $s$  servers ( $V_s$ ): ( $1 \leq s \leq c$ )

$$V_s = \sum_{i:M_i(P_v) \geq s} \pi_i \quad (16)$$

- Availability of  $s$  servers ( $A_s$ ): ( $1 \leq s \leq c$ )

$$A_s = 1 - \sum_{i:M_i(P_s) + M_i(P_v) \geq s} \pi_i \quad (17)$$

- The mean waiting time ( $\bar{W}$ ):

$$\bar{W} = n_o / \bar{\lambda} \quad (18)$$

- The mean response time ( $\bar{R}$ ):

$$\bar{R} = (n_o + n_s) / \bar{\lambda} \quad (19)$$

## V. VALIDATION OF RESULTS

In this section, we consider some numerical results to validate the proposed models and also to show the influence of system parameters and vacation policies on the performance measures of multiserver retrial systems. The numerical results were established using the GreatSPN tool.

In Table 1, some experimental results are collected when the servers vacation rate and the station vacation rate are very large. The results were validated by the Pascal program given in the book of Falin and Templeton [25] for the analysis of multiserver retrial queues without vacations. From this table, we can see that the corresponding performance measures are very close to the case without vacation and to each other with server or station vacation policy with very high vacation rate.

Define the parameter  $\rho = N\lambda/\mu$ , which is the largest offered load in the system. Table 2 shows the variation of the mean response time with  $\rho$ , for the single and multiple vacation policies, when the service station consists of one server and the retrial rate is very high. From this table, we can see that the numerical results are very close to those obtained by Trivedi [23] for single server queueing systems with vacations and without retrials, since the retrial rate is very large.

Table I  
VALIDATION OF RESULTS IN MULTISERVER RETRIAL CASE WITHOUT VACATIONS

	Model without vacation [25]	Model with servers vacation	Model with station vacation
Number of servers	4	4	4
Size of source	20	20	20
Primary call generation rate	0.1	0.1	0.1
Service rate	1	1	1
Retrial rate	1.2	1.2	1.2
Vacation rate	-	1e+25	1e+25
Mean number of busy servers	1.800 748	1.800 768	1.800 758
Mean number of customers of repeated calls	0.191 771	0.191 788	0.191 786
Mean rate of generation of primary calls	1.800 748	1.800 744	1.800 746
Mean waiting time	1.106 495	1.106 518	1.106 510

Table II  
MEAN RESPONSE TIME WITH  $N = 50$ ,  $\mu = 1$ ,  $\theta = 0.5$ ,  $c = 1$

$\rho$	Models without retrials [23]		Models with $\nu = 1e + 25$	
	Multiple vacations	Single vacations	Multiple vacations	Single vacations
0.1	3.107	1.494	3.106 810	1.493 581
0.3	3.391	2.370	3.390 962	2.370 404
0.5	3.834	3.172	3.833 990	3.172 221
0.7	4.592	4.152	4.592 591	4.152 760
0.9	6.000	5.718	6.000 657	5.719 090

## VI. CONCLUSION

In this paper, we proposed a technique that allows modeling and analyzing finite-source multiserver retrial systems with different vacation policies using GSPNs. The novelty of the investigation is essentially the combination of multiplicity of servers with the simultaneous presence of repeated calls and vacations, which make the system rather complicated.

The flexibility of GSPNs modeling approach allowed us a simple construction of detailed and compact models for these systems. Moreover, it made it possible to verify many qualitative properties of interest by inspection of the reachability graph. From a performance point of view, the proposed approach offers a rich means of expressing interesting performance indices as a function of the Petri net elements.

Finally, many retrial and vacation systems problems and their solutions can be simplified using the stochastic Petri nets modeling approach with all the methods and tools developed within this framework.

## REFERENCES

- [1] J. R. Artalejo and M. J. Lopez-Herrero, *Cellular mobile networks with repeated calls operating in random environment*, Computers & Operations Research 37, no. 7, 2010, pp. 1158-1166.
- [2] T. V. Do, *A new computational algorithm for retrial queues to cellular mobile systems with guard channels*, Computers & Indus. Engin. 59, 2010, pp. 865-872.
- [3] N. Gharbi, L. Mokdad, and J. Ben-Othman, *Performance and Reliability Analysis of Small Cell Networks with Retrials and Different Breakdowns Disciplines: A Computational Approach*, The 5th IEEE International Workshop on Performance Evaluation of Communications in Distributed Systems and Web based Service Architectures (PEDISWESA'2013), Split, Croatia, July 7-10, 2013.
- [4] N. Gharbi and L. Mokdad, *Performance Evaluation of Telecommunication Systems with Repeated Attempts and Two Servers Classes*, International Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'2013), Toronto, Canada, July 7-10, 2013.
- [5] P. Wuechner, J. Sztrik, and H. De Meer, *Modeling Wireless Sensor Networks Using Finite-Source Retrial Queues with Unreliable Orbit*, Proc. of the Workshop on Perf. Eval. of Computer and Communication Systems (PERFORM'2010), vol. 6821 of LNCS Publisher: Springer-Verlag, 2011.
- [6] G. I. Falin and J. G. C. Templeton, *Retrial Queues*, Chapman and Hall, London, 1997.
- [7] J. R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems: A Computational Approach*, Springer, Berlin, 2008.
- [8] J. R. Artalejo, *Accessible bibliography on retrial queues: Progress in 2000-2009*, Mathematical and Computer Modelling 51, 2010, pp. 1071-1081.
- [9] Z. G. Zhang and N. Tian, *Analysis of queueing systems with synchronous single vacation for some servers*, Queueing Systems: Theory and Applications, vol. 45, no. 2, 2003, pp. 161-175.
- [10] Z. G. Zhang and N. Tian, *Analysis of queueing systems with synchronous vacations of partial servers*, Performance Evaluation, vol. 52, no. 4, 2003, pp. 269-282.

- [11] C. H. Lin and J. C. Ke, *Multi-server system with single working vacation*, Applied mathematical Modelling, vol. 33, 2009, pp. 2967-2977.
- [12] J. C. Ke, C. H. Lin, J. Y. Yang, and Z. G. Zhang, "Optimal (d,c) vacation policy for a finite buffer M/M/c queue with unreliable servers and repairs", Applied Mathematical Modelling, vol. 33, 2009, pp. 3949-3962.
- [13] J. C. Ke, C. H. Wu, and Z. G. Zhang, "A Note on a Multi-server Queue with Vacations of Multiple Groups of Servers", Quality Technology and Quantitative Management, vol. 10, no. 4, 2013, pp. 513-525.
- [14] B. T. Doshi, "Queueing systems with vacations: a survey", Queueing Systems, vol. 1, no. 1, 1986, pp. 29-66.
- [15] H. Takagi, Queueing Analysis: A Foundation of Performance Evaluation, Volume I, Vacation and Priority Systems, Part I, North-Holland, Amsterdam, 1991.
- [16] N. Tian and Z. G. Zhang, Vacation Queueing Models: Theory and Applications, Springer, New York, 2006.
- [17] J. C. Ke, C. H. Wu, and Z. G. Zhang, "Recent Developments in Vacation Queueing Models: A Short Survey", International Journal of Operations Research, vol. 7, no. 4, 2010, pp. 3-8.
- [18] S. J. Kwon, "Performance analysis of CDPD Sleep mode for power conservation in mobile end systems", IEICE Trans. Commun. E84, 2001.
- [19] H. Li and T. Yang, "A single server retrial queue with server vacations and a finite number of input sources", European Journal of Operational Research 85, 1995, pp. 149-160.
- [20] G. K. Janssens, "The quasi-random input queueing system with repeated attempts as a model for collision-avoidance star local area network", IEEE Transactions on Communications 45, 1997, pp. 360-364.
- [21] G. Choudhury, "A two phase batch arrival retrial queueing system with Bernoulli vacation schedule", Applied Mathematics and Computation 188, 2007, pp. 1455-1466.
- [22] Z. Wenhui, "Analysis of a single-server retrial queue with FCFS orbit and Bernoulli vacation", Applied Mathematics and Computation 161, 2005, pp. 353-364.
- [23] K. S. Trivedi and O. C. Ibe, "Stochastic Petri net analysis of finite-population vacation queueing systems", Queueing Systems 8, 1991, pp. 111-128.
- [24] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis, Modelling with Generalized Stochastic Petri Nets, John Wiley and Sons, New York, 1995.
- [25] G. I. Falin and J. G. C. Templeton, Retrial Queues, Chapman and Hall, London, 1997.

# Level Crossing Rate of MRC Receiver Over $k$ - $\mu$ Multipath Fading Environment

Dragana Krstić, Ilija Temelkovski, Srdjan Maričić

Department of Telecommunications,  
Faculty of Electronic Engineering,  
University of Niš  
Niš, Serbia  
dragana.krstic@elfak.ni.ac.rs

Dragan Radenković, Vladeta Milenković

Department of Measurements,  
Faculty of Electronic Engineering,  
University of Niš  
Niš, Serbia

**Abstract**—In this paper, the wireless system with maximal ratio combining (MRC) diversity receiver operating over independent identical  $k$ - $\mu$  multipath fading environment is analyzed. The closed form expression for average level crossing rate of MRC receiver output signal is delivered. The average level crossing rate is calculated as value of the first derivative of MRC receiver output signal envelope. The obtained expression can be used for evaluation of average fade duration of proposed wireless system. Numerical results are presented graphically to show influence of  $k$ - $\mu$  multipath fading parameters on average level crossing rate (LCR) of considered MRC receiver output signal.

**Keywords**- average level crossing rate; MRC receiver;  $k$ - $\mu$  fading.

## I. INTRODUCTION

Multipath fading degrades the system performance and it limits the system capacity. Received signal experiences fading resulting in signal envelope variation. There are more distributions that can be used to describe signal envelope variation in fading channels, which are dependent on propagation environment and communication scenario. The most frequently used statistical models for description of signal envelope variation are Rayleigh, Rician, Nakagami- $m$ ,  $\alpha$ - $\mu$  and  $k$ - $\mu$  [1] [2].

Rayleigh distribution can be used to describe small scale signal envelope variation in linear non line-of-sight multipath fading channels. In linear line-of-sight multipath fading environments, signal envelope variation can be described with Rician distribution. Rician distribution has Rice factor  $k$ . The Rice factor  $k$  is related to the ratio of dominant component's power and scattering component's power. In multipath fading environments with two or more clusters signal envelope variation can be analyzed with Nakagami- $m$  distribution. Nakagami- $m$  distribution has parameter  $m$ . The parameter  $m$  is related to the fading severity and to the number of clusters in multipath fading propagation environment.

By setting  $m=1$ , Nakagami- $m$  distribution reduces to Rayleigh distribution. The one sided Gaussian distribution is obtained from Nakagami- $m$  distribution for  $m=0.5$ . When parameter  $m$  goes to infinite, Nakagami- $m$  fading reduces to no fading case. As parameter  $m$  decreases, the fading severity increases.

The  $\alpha$ - $\mu$  fading can be applied for analyzing small scale signal envelope variation in nonlinear multipath fading environment [3]. The  $\alpha$ - $\mu$  fading is general fading model. The Rayleigh, Weibull and Nakagami- $m$  distributions can be derived from  $\alpha$ - $\mu$  distribution [4].

The Weibull distribution can be obtained from the  $\alpha$ - $\mu$  distribution by setting  $\mu = 1$ . From the Weibull distribution, by setting  $\alpha = 2$ , the Rayleigh distribution is obtained; (then the Rayleigh distribution can be obtained from  $\alpha$ - $\mu$  distribution by setting  $\alpha=2$  and  $\mu=1$ ).

For  $\alpha = 2$ , the  $\alpha$ - $\mu$  distribution reduces to Nakagami- $m$  distribution.

Now, from the Nakagami- $m$  distribution, by setting  $\mu=1$ , the Rayleigh distribution is obtained.

Still from the Nakagami- $m$  distribution, the one-sided Gaussian distribution is obtained using as parameter  $\mu = 1/2$ .

The  $k$ - $\mu$  distribution can be used to describe small scale signal envelope variation in linear line-of-sight multipath fading propagation environments [5]. The  $k$ - $\mu$  distribution has two parameters. The parameter  $k$  is related to ratio of dominant component's power and scattering component's power. The parameter  $\mu$  is related to the number of clusters in propagation environments.

The  $k$ - $\mu$  distribution is also general distribution. From  $k$ - $\mu$  distribution the Rayleigh, Rician and Nakagami- $m$  distributions can be derived as special cases.

By setting for  $k=0$ ,  $k$ - $\mu$  distribution approximates Nakagami- $m$  distribution. For  $\mu=1$ , from  $k$ - $\mu$  distribution can be derived Rician distribution and for  $\mu=1$  and  $k=0$ ,  $k$ - $\mu$  distribution reduces to Rayleigh distribution [6].

There are several combining techniques which can be used to reduce  $k$ - $\mu$  multipath fading effects on level crossing rate of wireless system depend on complexity restriction put on communication system and quality of service required from communication system [7].

The most frequently applied combining techniques are maximal ratio combining (MRC), equal gain combining (EGC) and selection combining (SC). The MRC diversity technique provides the best performance and it is the most complex for practical implementation [1].

In telecommunications, maximal-ratio combining (MRC) is a method of diversity combining in which the signals from each channel are added together, the gain of each channel is made proportional to the root mean square (rms) signal level and inversely proportional to the mean square noise

level in that channel. The different proportionality constants are used for each channel. It is also known as ratio-squared combining and predetection combining. Actually, with Maximal-ratio combining, the diversity branches are weighted by their respective complex fading gains and combined. The MRC is the optimum combiner for independent AWGN channels. MRC can restore a signal to its original shape [8].

The first and second order performance measure can be evaluated for wireless communication system. The first order performance measures are: the outage probability, the bit error probability and the system capacity. The second order performance measures are: the average level crossing rate of output signal envelope and the average fade duration of wireless communication system [9]. The average level crossing rate can be calculated as average value of the first derivative of received output signal and the average fade duration can be obtained as ratio of outage probability and average level crossing rate.

The rest of this paper is organized as follows. In section II, it will be spoken about related works. Section III describes the derivation of level crossing rate of MRC output signal. Numerical results are presented in Section IV. Section V gives last details and concludes the paper.

## II. RELATED WORKS

There are more works in open technical literature considering the second order performance measures of wireless communication system operating over multipath fading channels. In [10], wireless communication system with SIR based dual branches SC receiver operating over Rician multipath fading environment in the presence of cochannel interference subjected to Rayleigh multipath fading is considered. The average level crossing rate and the average fade duration are derived for proposed system. The average level crossing rate and average fade duration of SC receiver operating over multipath fading channel are evaluated in [11] and [12] respectively. The performance analysis of selection diversity over exponentially correlated  $\alpha$ - $\mu$  fading environment is done in [13].

In [6], an exact closed-form expression for the phase-envelope joint distribution of the  $k$ - $\mu$  fading environment, a general fading model that includes the Rice and the Nakagami- $m$  models as special cases is derived. The derived joint statistics are obtained for both, Rice and Nakagami- $m$  cases.

In paper [14], the ratio of product of two random variables and random variable is considered. The product of two random variables in the numerator of the ratio can represent desired signal envelope subjected to two multipath fading. The random variable in denominator of the ratio can represent cochannel interference signal envelope affected to multipath fading. The results obtained in the paper can be used in performance analysis of wireless communication system operating over multipath fading channels in the presence of cochannel interference which suffer to multipath fading.

The outage performance and symbol error rate analysis of L-Branch MRC for  $k$ - $\mu$  and  $\eta$ - $\mu$  fading are given in [15].

The first order system performance analysis of  $L$ -branch MRC for  $k$ - $\mu$  fading is done by Milišić, Hamza and Hadžialić. They made outage performance, outage and Symbol Error Probability performance and BEP/SEP and outage performance analysis in [16], [17] and [18], respectively.

Analysis of channel capacity per unit bandwidth of  $L$ -branch MRC receiver operating over  $k$ - $\mu$  fading channels for two adaptive transmission schemes is presented in [19]. Using proposed system model, the optimal power, rate adaptation and constant transmit power policies are analyzed. The expressions for capacity evaluation are derived in the terms of finite sums and the effects of diversity order and fading parameters on the channel capacity for given techniques are considered and numerically presented.

The closed-form expressions for the level crossing rate and average fade duration of  $k$ - $\mu$  distributed fading signal envelope is presented in [20]. The proposed equations are validated by reduction to known Rice, Rayleigh and Nakagami- $m$  distributions as special cases. They are also compared with measured data and shown that provide good agreement.

There are many new studies in the area such as [21]. In this report, a dual-hop decode-amplify-forward (DAF) transmission system over Nakagami- $m$  fading channel is studied. The DAF relay system is a hybrid of decode-and-forward and amplify-and-forward relay systems that show the benefits of both decode-and-forward and amplify-and-forward relay systems and is also called hybrid relay system or hybrid DAF relay system. Signal-to-noise ratios and BERs for various system models with varying number of transmit and receive antennas have been discussed.

In this paper, the wireless communication system with MRC diversity receiver operating over multipath fading channel will be analyzed. The received signal is subjected to  $k$ - $\mu$  multipath fading. MRC diversity receiver is used to reduce multipath fading effects to outage probability and average level crossing rate of proposed system. The closed form expressions for average level crossing rate and average fade duration will be derived. The probability density function and cumulative distribution function of MRC output signal envelope will also be calculated as expressions in closed form. To the best author knowledge the average level crossing rate of MRC output signal envelope in the presence of  $k$ - $\mu$  multipath fading is not reported in open technical literature. The results obtained in this paper can be used in designing and analyzing of wireless communication system operating over  $k$ - $\mu$  multipath fading environments.

## III. LEVEL CROSSING RATE OF MRC OUTPUT SIGNAL

The wireless communication system with MRC receiver operating over  $k$ - $\mu$  multipath fading environments is considered. The  $k$ - $\mu$  multipath fading is presented at the inputs of MRC receiver. The MRC is applied to reduce  $k$ - $\mu$  fading effects on system performance. The  $k$ - $\mu$  multipath fading is identical and independent.

Assuming that thermal noise power is equal among branches, the squared MRC receiver signal envelope can be

obtained as sum of squared signal envelopes of inputs of MRC:

$$z^2 = \sum_{i=1}^L y_i^2 \quad (1)$$

where  $y_i$  is  $k$ - $\mu$  distributed signal envelope which can be given as

$$y_i^2 = y_{i1}^2 + y_{i2}^2 + \dots + y_{i2\mu}^2 \quad (2)$$

where  $y_{ik}$ ,  $k=1, 2, \dots, 2\mu$ , are Gaussian random variables with average  $A_{ik}$  and variance  $\sigma^2$ . The probability density function of  $y_{ik}$  is

$$p_{y_{ik}}(y_{ik}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_{ik}-A_{ik})^2}{2\sigma^2}} \quad (3)$$

The first derivative of MRC output signal  $z$  is:

$$\dot{z} = \frac{1}{z} \sum_{i=1}^L y_i \dot{y}_i \quad (4)$$

where

$$y_i \dot{y}_i = y_{i1} \dot{y}_{i1} + y_{i2} \dot{y}_{i2} + \dots + y_{i2\mu} \dot{y}_{i2\mu} \quad (5)$$

After substituting expression (5) into (4), the  $\dot{z}$  can be written in the form:

$$\dot{z} = \frac{1}{z} \sum_{i=1}^L (y_{i1} \dot{y}_{i1} + y_{i2} \dot{y}_{i2} + \dots + y_{i2\mu} \dot{y}_{i2\mu}) \quad (6)$$

The first derivative of Gaussian random variable is Gaussian random variable. The linear transformation of Gaussian random variable is also Gaussian random variable. Therefore, random variable  $\dot{z}$  has conditional distribution:

$$p_{\dot{z}}(\dot{z}/z, y_{ik}) = \frac{1}{\sqrt{2\pi}\sigma_{\dot{z}}} e^{-\frac{(\dot{z}-\bar{\dot{z}})^2}{2\sigma_{\dot{z}}^2}} \quad (7)$$

where  $\bar{\dot{z}}$  and  $\sigma_{\dot{z}}$  are average value and variance of  $\dot{z}$ , respectively.

The average value of  $\dot{z}$  is

$$\bar{\dot{z}} = \frac{1}{2} \sum_{i=1}^L (y_{i1} \dot{y}_{i1} + y_{i2} \dot{y}_{i2} + \dots + y_{i2\mu} \dot{y}_{i2\mu}) = 0 \quad (8)$$

since

$$\bar{\dot{y}}_{i1} = \bar{\dot{y}}_{i2} = \dots = \bar{\dot{y}}_{i2\mu} = 0 \quad (9)$$

The variance of  $\dot{z}$  is

$$\sigma_{\dot{z}}^2 = \frac{1}{z^2} \sum_{i=1}^L (y_{i1}^2 \sigma_{\dot{y}_{i1}}^2 + y_{i2}^2 \sigma_{\dot{y}_{i2}}^2 + \dots + y_{i2\mu}^2 \sigma_{\dot{y}_{i2\mu}}^2) \quad (10)$$

$$\sigma_{\dot{y}_{i1}}^2 = \sigma_{\dot{y}_{i2}}^2 = \dots = \sigma_{\dot{y}_{i2\mu}}^2 = 2\pi^2 \sigma^2 f_m^2 = \pi^2 f_m^2 \frac{\Omega}{\mu} \quad (11)$$

where  $f_m$  is maximal Doppler frequency,  $\Omega$  is power of  $k$ - $\mu$  random variable and  $\mu$  is severity of  $k$ - $\mu$  fading.

The variance of the first derivative of  $z$  becomes

$$\begin{aligned} \sigma_{\dot{z}}^2 &= \frac{1}{z^2} \pi^2 f_m^2 \frac{\Omega}{\mu} \sum_{i=1}^L (y_{i1}^2 + y_{i2}^2 + \dots + y_{i2\mu}^2) = \\ &= \frac{1}{z^2} \pi^2 f_m^2 \frac{\Omega}{\mu} z^2 = \pi^2 f_m^2 \frac{\Omega}{\mu} \end{aligned} \quad (12)$$

The joint probability density function of  $z$  and  $\dot{z}$  is

$$p_{z\dot{z}}(z\dot{z}) = p_z(\dot{z}/z) \cdot p_z(z) = p_{\dot{z}}(\dot{z}) \cdot p_z(z) \quad (13)$$

where random variable  $z$  follows  $k$ - $\mu$  distribution

$$\begin{aligned} p_z(z) &= \frac{2\mu L(k+1)^{\frac{L\mu+1}{2}}}{k^{\frac{L\mu-1}{2}} e^{L\mu k} (L\Omega)^{L\mu+1}} z^{L\mu} e^{-\frac{L\mu(k+1)}{L\Omega} z^2} \cdot \\ &\cdot I_{L\mu-1}\left(\frac{2L\mu\sqrt{k(k+1)}}{L\Omega}\right) = \\ &= \frac{2\mu L(k+1)^{\frac{L\mu+1}{2}}}{k^{\frac{L\mu-1}{2}} e^{L\mu k} (L\Omega)^{L\mu+1}} z^{L\mu} e^{-\frac{\mu(k+1)}{\Omega} z^2} \cdot \\ &\cdot I_{L\mu-1}\left(\frac{2\mu\sqrt{k(k+1)}}{\Omega}\right) \end{aligned} \quad (14)$$

The level crossing rate of MRC output signal can be evaluated as the average value of the first derivative of MRC output signal:

$$\begin{aligned} N_2 &= \int_0^{\infty} d\dot{z} \dot{z} p_{z\dot{z}}(z\dot{z}) = \int_0^{\infty} d\dot{z} \dot{z} p_{\dot{z}}(\dot{z}) \cdot p_z(z) = \\ &= p_z(z) \int_0^{\infty} d\dot{z} p_{\dot{z}}(\dot{z}) \cdot \dot{z} = p_z(z) \frac{1}{\sqrt{2\pi}} \sigma_{\dot{z}} = \\ &= \frac{1}{\sqrt{2\pi}} \frac{2\mu L(k+1)^{\frac{L\mu+1}{2}}}{k^{\frac{L\mu-1}{2}} e^{L\mu k} (L\Omega)^{L\mu+1}} z^{L\mu} e^{-\frac{\mu(k+1)}{\Omega} z^2} \cdot \\ &\cdot I_{L\mu-1}\left(\frac{2\mu\sqrt{k(k+1)}}{\Omega} z\right) \cdot \pi f_m \left(\frac{\Omega}{\mu}\right)^{1/2} \end{aligned} \quad (15)$$

Normalized  $N_2$  is

$$\begin{aligned} \frac{N_2}{f_m} &= \sqrt{\frac{2\pi}{2}} \frac{2L\sqrt{\mu}(k+1)^{\frac{L\mu+1}{2}}}{k^{\frac{L\mu-1}{2}} e^{L\mu k} L^{\frac{L\mu+1}{2}} \Omega^{\frac{L\mu+1}{2}}} z^{L\mu} \cdot \\ &\cdot e^{-\frac{\mu(k+1)}{\Omega} z^2} I_{L\mu-1}\left(\frac{2\mu\sqrt{k(k+1)}}{\Omega} z\right) \end{aligned} \quad (16)$$

IV. NUMERICAL RESULTS

In Fig. 1, the level crossing rate of MRC output signal versus MRC output signal envelope is presented for different values of Rice factor  $k$ , the number of clusters in propagation environment  $m$ , power of MRC output signal and the number of inputs of MRC combiner.

The level crossing rate of MRC output signal increases as envelope  $z$  increases for lower values of envelope  $z$  and the level crossing rate decreases as envelope  $z$  increases for higher values of  $z$ . The average level crossing rate decreases as MRC output signal envelope power increases. The power of MRC output signal  $z$  has greater influence on average level crossing rate for lower values. The influence of parameters on average level crossing rate decreases as signal envelope increases.

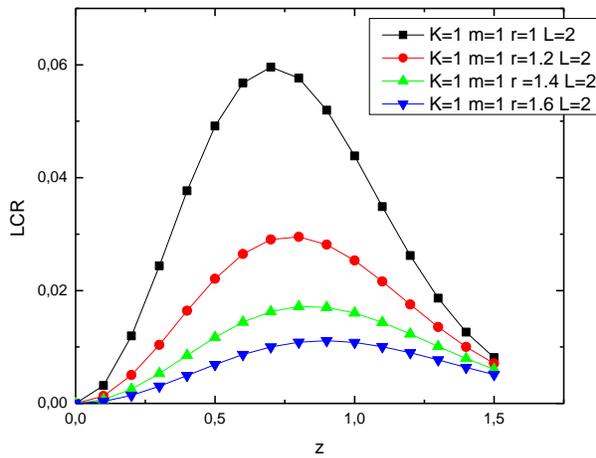


Figure 1. The level crossing rate (LCR) of MRC output signal versus MRC output signal envelope  $z$

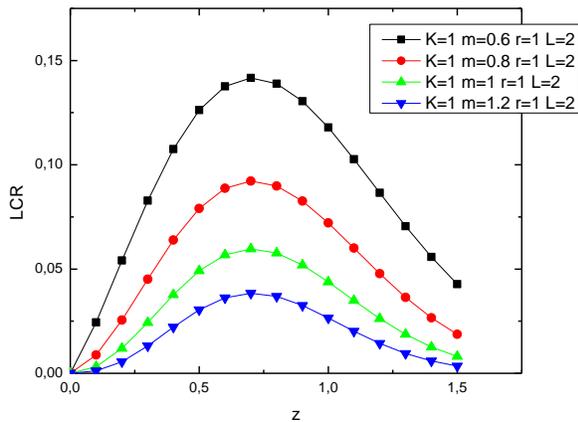


Figure 2. The level crossing rate (LCR) versus MRC output signal envelope  $z$

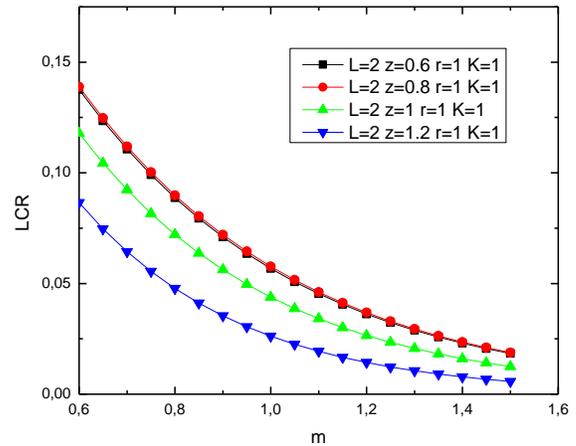


Figure 3. The level crossing rate (LCR) of MRC output signal versus number of clusters in propagation environment,  $m$

In Fig. 2, the level crossing rate of MRC output signal versus MRC output signal envelope is presented for different values of parameter  $m$ . The average level crossing rate decreases as parameter  $m$  increases. The system performances are better for greater values of parameter  $m$ .

The level crossing rate of MRC output signal versus number of clusters in propagation environment,  $m$ , is shown in Fig. 3, for different values of MRC output signal envelope and number of inputs of MRC combiner, power of MRC output signal and Rice factor  $k$ . The average level crossing rate decreases as parameter  $m$  increases. The average LCR of MRC output signal envelope also decreases with increasing of MRC output signal envelope  $z$ . The system performances are better for greater values of parameter  $m$  and MRC output signal envelope.

The outage probability of wireless system increases as average level crossing rate increases.

V. CONCLUSION

The wireless communication system with MRC diversity receiver operating over multipath fading channel is considered in this paper. Received signal is subjected to  $k$ - $\mu$  multipath fading environments. The  $k$ - $\mu$  distribution can be used to describe small scale signal envelope variation in linear non line-of-sight multipath fading environment with two or more clusters. The MRC diversity receiver is used to mitigate  $k$ - $\mu$  multipath fading effects on system performance. The MRC receiver provides the best performance and has the highest implementation complexity. The closed form expressions for average level crossing rate of MRC receiver output signal and average fade duration of proposed system are evaluated. The average level crossing rate is calculated as the average value of the first derivative of MRC output signal and average fade duration is calculated as the ratio of outage probability and average level crossing rate.

The outage probability and the bit error probability are the first order performance measures and the average level

crossing rate of output signal envelope and the average fade duration are the second order performance measures.

Numerical results are presented graphically to show the influence of  $k$ - $\mu$  multipath fading parameters on level crossing rate. The expression for average level crossing rate of MRC receiver output signal has four parameters. The parameter  $\mu$  is fading severity. The average level crossing rate decreases as the parameter  $\mu$  increases. The parameter  $\mu$  has greater influence on average level crossing rate for lower values of parameter  $\mu$ . The parameter  $k$  is defined as a ratio of dominant components power and scattering components power. The average level crossing rate decreases as the parameter  $k$  increases. The influence of parameter  $k$  on average level crossing rate is bigger for lower values of parameter  $\mu$ .

$L$  is the number of branches of MRC receiver. The average level crossing rate decreases as parameter  $L$  (number of branches) increases. The parameter  $\Omega$  is average squared value of signal envelope at the input of the receiver. The average fade duration increases as parameter  $\Omega$  increases.

The obtained results in this paper can be used for analyzing and designing of wireless communication system with MRC receiver operating over  $k$ - $\mu$  multipath fading environments.

#### ACKNOWLEDGMENT

This paper has been partially funded by the Ministry of Education, Science and Technological Development of Republic of Serbia under projects III-44006 and TR-33035.

#### REFERENCES

- [1] G. Stuber, Principles of Mobile Communication, Boston: Kluwer Academic Publishers; 2000.
- [2] P. M. Shankar, Fading and Shadowing in Wireless Systems, Springer, Dec 7, 2011 - Technology & Engineering”, DOI 10.1007/978-1-4614-0367-8
- [3] M. D. Yacoub, “The  $\alpha$ - $\mu$  Distribution: A Physical Fading Model for the Stacy Distribution”, IEEE Transactions on Vehicular Technology, vol. 56, no. 1, Jan. 2007, pp. 27-34.
- [4] M. D. Yacoub, “The  $\alpha$ - $\mu$  distribution: a general fading distribution”. 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2002, Lisbon, Portugal, Print ISBN: 0-7803-7589-0, DOI: 10.1109/PIMRC.2002.1047298, vol.2, September 15-18, 2002, pp. 629 – 633.
- [5] M. D. Yacoub, ”The  $k$ - $\mu$  Distribution”, <http://www.eletrica.ufpr.br/anais/sbrt/SBrT19/001000000008700059.pdf>
- [6] U. S. Dias and M. D. Yacoub, “The  $k$ -  $\mu$  Phase Envelope joint Distribution, IEEE Transactions on Communications, no. 01, 2010; 58, pp. 40-45. DOI:10.1109/TCOMM.2010.01.080175
- [7] S. Panic, D. Stefanovic, I. M. Petrovic, M. C. Stefanovic, J. A. Anastasov, and D. S. Krstic, “Second-order statistics of selection macro-diversity system operating over Gamma shadowed  $\kappa$ - $\mu$  fading channels”. EURASIP J. Wireless Comm. and Networking, 2011: 151, 2011.
- [8] [http://en.wikipedia.org/wiki/Maximal-ratio\\_combining](http://en.wikipedia.org/wiki/Maximal-ratio_combining)
- [9] M. Stefanovic, B. Miric, P. Spalevic, and S. Panic, “Second Order Statistic Analysis of Selection Macro-Diversity Combining over Gama Shadowed Rayleigh Fading Channels,” Scientific Publications of the State University of Novi Pazar, SER. A: Applied Math. Inform. and Mech, vol. 1, No. 1, 2009, pp. 1-9
- [10] A. S. Panajotovic, M. C. Stefanovic, D. Li, Draca. and N. M. Sekulovic, “Average Level Crossing Rate of Dual Selection Diversity in Correlated Rician Fading with Rayleigh Cochannel Interference”, IEEE Communications Letters”, 14(7), 2010, pp. 605-607.
- [11] A. S. Panajotovic, N. M. Sekulovic, M. C. Stefanovic, and D. Li, Draca, “Average Level Crossing Rate of Dual Selection Diversity over Correlated Unbalanced Nakagami-m Fading Channels in the Presence of Cochannel Interference”, IEEE Communications Letters 16(5), 2012, pp. 691-693.
- [12] A. S. Panajotovic, M. C. Stefanovic, D. Li, Draca, N. M. Sekulovic, and I. M. Petrovic, “Average fade duration of dual selection diversity over correlated Rician fading channels in the presence of cochannel interferences”, EUROCON 2011, pp. 1-4.
- [13] Z. Popovic, S. Panic, J. A. Anastasov, P. C. Spalevic, and M. C. Stefanovic. “Performance analysis of selection diversity over exponentially correlated  $\alpha$ -  $\mu$  fading environment”, Int. J. Communication Systems 24(7), 2011, pp. 925-937.
- [14] D. Krstic, M. Stefanovic, N. Simic, and A. Stevanovic. “The Level Crossing Rate of the Ratio of Product of Two  $k$ -  $\mu$  Random Variables and  $k$ -  $\mu$  Random Variable”, 13<sup>th</sup> WSEAS International Conference on Electric Power Systems, High Volages, Electric Machine, POWER’13, Chania, Crete Island, Greece, ISBN 978-960-474-329-2, Aug. 27-29, 2013, pp. 153-158.
- [15] M. Milišić, M. Hamza, and M. Hadžialić, “Outage Performance and Symbol Error Rate Analysis of L-Branch Maximal-Ratio Combiner for  $k$ - $\mu$  and  $\eta$ - $\mu$  Fading”, Computer and Information Science » Computer Science and Engineering » "Vehicular Technologies: Increasing Connectivity". book edited by Miguel Almeida. ISBN 978-953-307-223-4, Published: April 11, 2011 under CC BY-NC-SA 3.0 license, pp. 333-358.
- [16] M. Milišić, M. Hamza, and M. Hadžialić, “Outage Performance of L-branch Maximal-Ratio Combiner for Generalized  $k$ - $\mu$  Fading”, IEEE International Conf. on Vehicular Technology, VTC Spring 2008, Singapore 11-14 May 2008, pp. 325-329.
- [17] M. Milišić, M. Hamza, and M. Hadžialić, “Outage and Symbol Error Probability Performance of L-branch Maximal-Ratio Combiner for Generalized  $\kappa$ - $\mu$  Fading”, 50th International Symposium ELMAR, ISSN: 1334-2630, Print ISBN: 978-1-4244-3364-3, Zadar, Croatia, 10-12 September 2008, pp. 231-236.
- [18] M. Milišić, M. Hamza, and M. Hadžialić, “BEP/SEP and Outage Performance Analysis of L-Branch Maximal-Ratio Combiner for  $k$ - $\mu$  Fading”, Int. J. Digital Multimedia Broadcasting 2009, doi:10.1155/2009/573404.
- [19] M. Stefanović, S. Panić, J. Anastasov, A. Cvetković, and Z. Popović, “Capacity Evaluation for Maximal Ratio Combining over  $\kappa$ - $\mu$  Fading Channels”, Serbian Journal of Electrical Engineering, Vol. 8, No. 2, May 2011, pp. 221-228.
- [20] S. L. Cotton and W. G. Scanlon, “Higher-order Statistics for  $k$ -  $\mu$  Distribution”, Electronics Letters, no 2, 2007, DOI:10.1049/el:20072372
- [21] S. G. Bansal and J. H. Abawajy, “Performance Analysis of Two-Hop Decode-Amplify-Forward Relayed System in different Fading Conditions”, Wireless Communications and Mobile Computing, DOI: 10.1002/wcm.2370, ISSN: 1530-8669, 2013.

## Statistical Analysis and Re-generation of the Radiated Disturbances from Multiple Noise Sources

Suna Choi and Jong-wa Kwon  
 Radio Technology Research Department  
 Electronics and Telecommunications Research Institute  
 Daejeon, Republic of Korea  
 sunachoi@etri.re.kr

**Abstract**—A re-generation method of radiated disturbances from multiple noise sources using statistical approach is suggested. The radiated disturbances from multiple noise sources consisting of an electric board and a Plasma Display Panel (PDP) TV are investigated. The frequency spectrum and the Amplitude Probability Distribution (APD) of radiated noises are measured. The radiated magnetic field from each noise source is measured below 30MHz, and then APD measurement is progressed at the frequencies where the noise sources radiate high disturbances in common. A sequence of noise pulses is observed from PDP TV, and the behavior of the APD of the electric board noise is similar to the Gaussian noise. The parameters of the disturbances are extracted using the Middleton's class A noise model and approximate empirical method. From the extracted parameters, random noise data is re-generated and radiated using a signal generator and antenna. APD comparison of the measured multiple noises and re-generated noises is performed. The results show that the re-generated noises are well matched to measured noises. The re-generated noise can be applied to the performance evaluation of communication systems against non-Gaussian multiple noises circumstances.

**Keywords**—multiple noise sources; radiated disturbance; amplitude probability distributio; electric boar; plasma display panel.

### I. INTRODUCTION

As the number of electric devices used in real situation is increased, problems of radiated interferences from multiple noise sources are more serious. In general, the multiple noises are assumed as Additive White Gaussian Noise (AWGN) [1] environments and most communication systems are designed to achieve optimal performance in AWGN environment. However, there are some applications where non-Gaussian noise or impulsive noise is dominant. Therefore, the performance evaluations of communications systems against the real multiple noise circumstance are necessary.

Since applying the performance evaluation in real situation has many difficulties, it is importance to re-generate the multiple noises in laboratory circumstance. The statistical analysis of multiple noises is required because the noise condition is not constant but various according to the time in real circumstance. Amplitude Probability Distribution (APD)

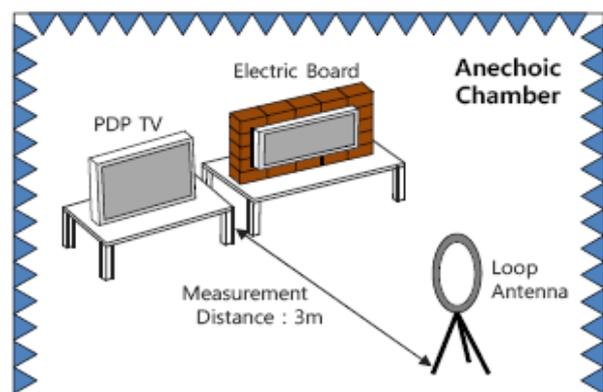


Figure 1. Measurement set-up

measurement gives probabilistic information about the disturbances, which has higher sensitivity and repeatability than existing methods, such as peak or quasi-peak measurements [2]-[4].

In this paper, a re-generation method of radiated disturbances from multiple noise sources using statistical approach is suggested. An electric board and a PDP TV are selected as noise sources for realizing an example of a real multiple noise circumstance. The spectra and APDs of radiated noises are measured according to the measurement methods from international standards of the special international committee on radio interference (CISPR) [5]. The statistical characteristics are analyzed using the Middleton's Class A noise model [6], which has been widely adopted due to the excellent agreement with measurement data. The parameters of the radiated noises are extracted using the approximate empirical method [7]-[8]. Using the extracted parameters, random noise data which has the statistically identical characteristics with the multiple noises is generated via a simulator. Then, the generated random noise is radiated in the air using a signal generator and an antenna.

Applying the suggested noise re-generation approach, radio noises which are statistically identical to the measured multiple noises can be radiated. Therefore, the performance evaluation of devices against the multiple noises in real circumstance can be performed easily in laboratory circumstance.

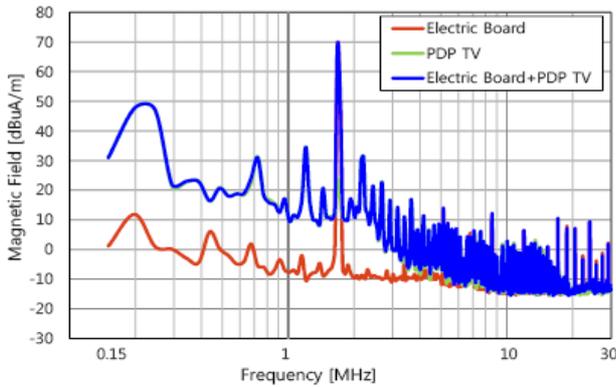


Figure 2. Disturbance spectra from noise sources

## II. MEASUREMENT

### A. Measurement set-up

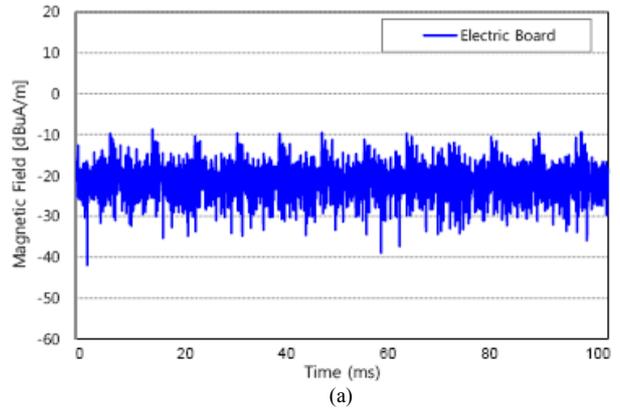
The special international committee on radio interference (CISPR) regulates the measurement methods of radiated disturbance. CISPR11[9] and CISPR16-1-4 [10] provide the measurement method of the magnetic field below 30MHz using an Open Area Test Site (OATS) method, and CISPR15 [11] provides another measurement method below 30MHz using a loop antenna system .

We performed the measurements of the radiated noises from multiple sources in an anechoic chamber according to the measurement method of OATS. As shown in Fig. 1, the antenna is a loop with 0.6m diameter, and the separation distance between the antenna and multiple sources is 3m. The electric board and PDP TV composing the multiple noise sources are placed side by side. The radiated noises are measured in terms of magnetic field strength in dB (uA/m).

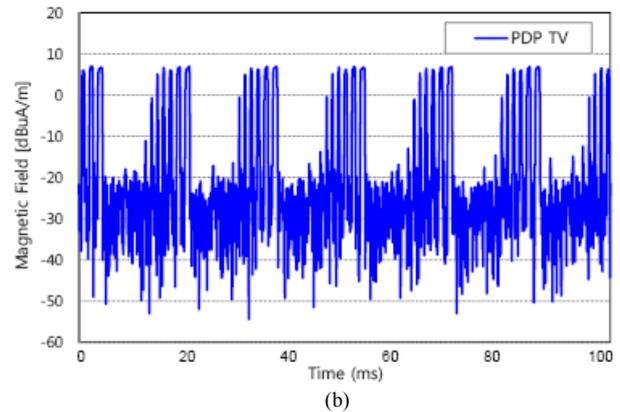
### B. Radiated magnetic field from 150k to 30MHz

Measurements of radiated magnetic fields of a multiple noises are performed in the peak mode with 9 kHz resolution bandwidth (RBW). The radiated disturbance is entirely measured from 150 kHz to 30 MHz, when each of the electric board and PDP TV is operated and then both of them are activated simultaneously. As presented in fig. 2, high radiated disturbance from electric board is shown at the fundamental frequency of 1.71MHz and the harmonic frequencies. Several high disturbances shown between 5MHz and 8MHz is supposed to come from inside circuits composing the electric board. The PDP TV shows more dense disturbances at the fundamental frequency of 240kHz and the harmonic frequencies. The disturbances from multiple noise sources show the combined shape of the two noises.

To investigate the fluctuation of noises according to the time, the radiated noise is measured with zero frequency range. Fig. 3 shows an example of each noise fluctuation from the electric board and PDP TV measured at 5.15MHz.



(a)



(b)

Figure 3. Noise fluctuations at 5.15MHz (a) Electric Board (b) PDP TV

The disturbances from electric board show relatively slight variation according to the time, whereas the noise from PDP TV shows high fluctuation periodically.

### C. APD measurement at the common frequencies

APD measurement method gives probabilistic information about the disturbance. The ordinate axis of the APD curve displays the probability that the disturbance envelope exceeds the abscissa level. When there is no impulsive input, the APD curve is gradually declined from the left top to the right side of bottom. When impulses are received, the APD shows drastic extension to the right side in the middle of the curve and the peak value of impulse appears on the abscissa of APD.

To investigate the effect of radiated disturbance from multiple noise sources, APD measurement is made at frequencies where both of the electric board and the PDP TV generate high disturbances. From the disturbance spectra of fig. 2, frequencies of 3.43, 5.15, 6.13 and 7.11MHz are selected. At each frequency, APD measurements are performed for 120 second using R&S ESU EMI receiver. Fig. 4 shows the APD measurement results of radiated noises from each of the electric board and PDP TV and both noise sources.

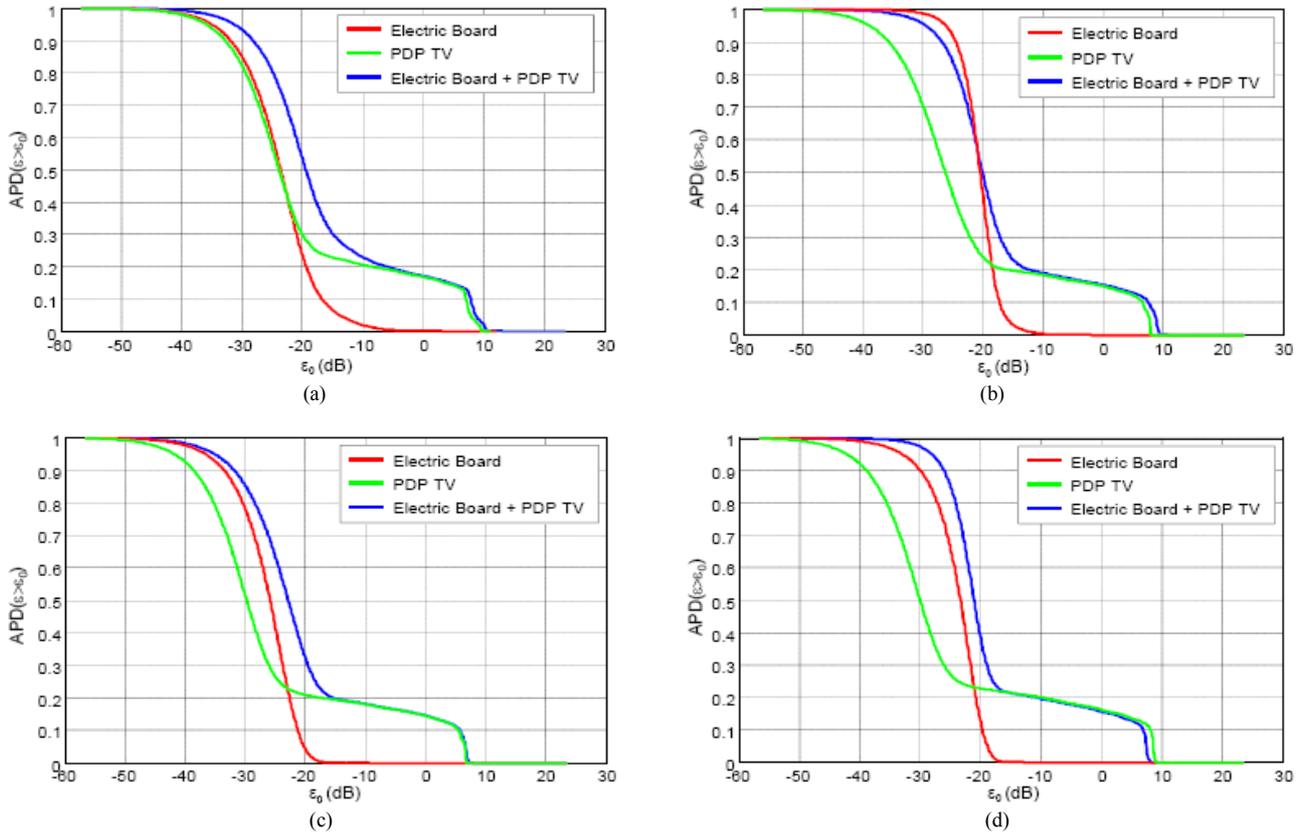


Figure 4. APD measurements of multiple noises (a) At 3.43MHz (b) At 5.15MHz (c) At 6.13MHz (d) At 7.11MHz

At all the measured frequencies, the APD graphs of electric board show gradual decrease. On the other hand, the APDs of PDP TV are gradually decreased and extended in the middle of curve. The APD shapes of multiple noises are similar to those of the PDP TV; however, the levels in the left side are higher than those of PDP TV due to the effects of noise from electric board. Uniquely, the APD of multiple noises at 5.15MHz is less than that of electric board in some range. It is considered that the noises from electric board and PDP TV are cancelled out by the each other.

### III. EXTRACTION OF NOISE PARAMETERS

The characteristics of multiple noises can be investigated by applying Middleton's Class A noise model. The model is appropriate to describe a highly impulsive interference and

TABLE I. PARAMETERS OF THE MULTIPLE NOISES

Frequency [MHz]	Electric Board		PDP TV		Electric Board + PDP TV	
	$A$	$\Gamma$	$A$	$\Gamma$	$A$	$\Gamma$
3.43	0	0.007	0.22	0.0048	0.23	0.0125
5.15	0	0.01	0.195	0.0028	0.2	0.011
6.13	0	0.038	0.19	0.0014	0.19	0.006
7.11	0	0.0063	0.215	0.0013	0.21	0.007

show good agreement with measurement data. It is defined by two parameters of  $A$  and  $\Gamma$ .  $A$  is the impulsive index which is the product of the mean number of disturbance events per second and the mean length of a disturbance.  $\Gamma$  is the Gaussian factor, as the ratio of mean power of the Gaussian component and that of impulsive non-Gaussian component.

Approximate empirical method is applied to estimate noise parameters.  $A$  is approximately extracted at the point where the sharp rise in disturbance versus probability occurs.  $\Gamma$  is extracted at the point where the straight-line starts to bend in probability versus disturbance. The smaller is  $A$ , the properties of the disturbance are dominated by the characteristic of typical, individual impulses and the disturbance shows the non-Gaussian properties. However, as  $A$  is close to zero, the noise exhibits a Gaussian property. The smaller is  $\Gamma$ , the intensity of non-Gaussian component is stronger related to the Gaussian component of the disturbance.

Table I shows the noise parameters extracted from the APD measurement, when each of the electric board and PDP TV is operated and then both of them are activated simultaneously. At all frequencies, the values of  $A$  parameter of electric board are nearly 0. This means the disturbances from the electric board are non-impulsive Gaussian noises. When investigating the  $A$  parameters extracted from the PDP

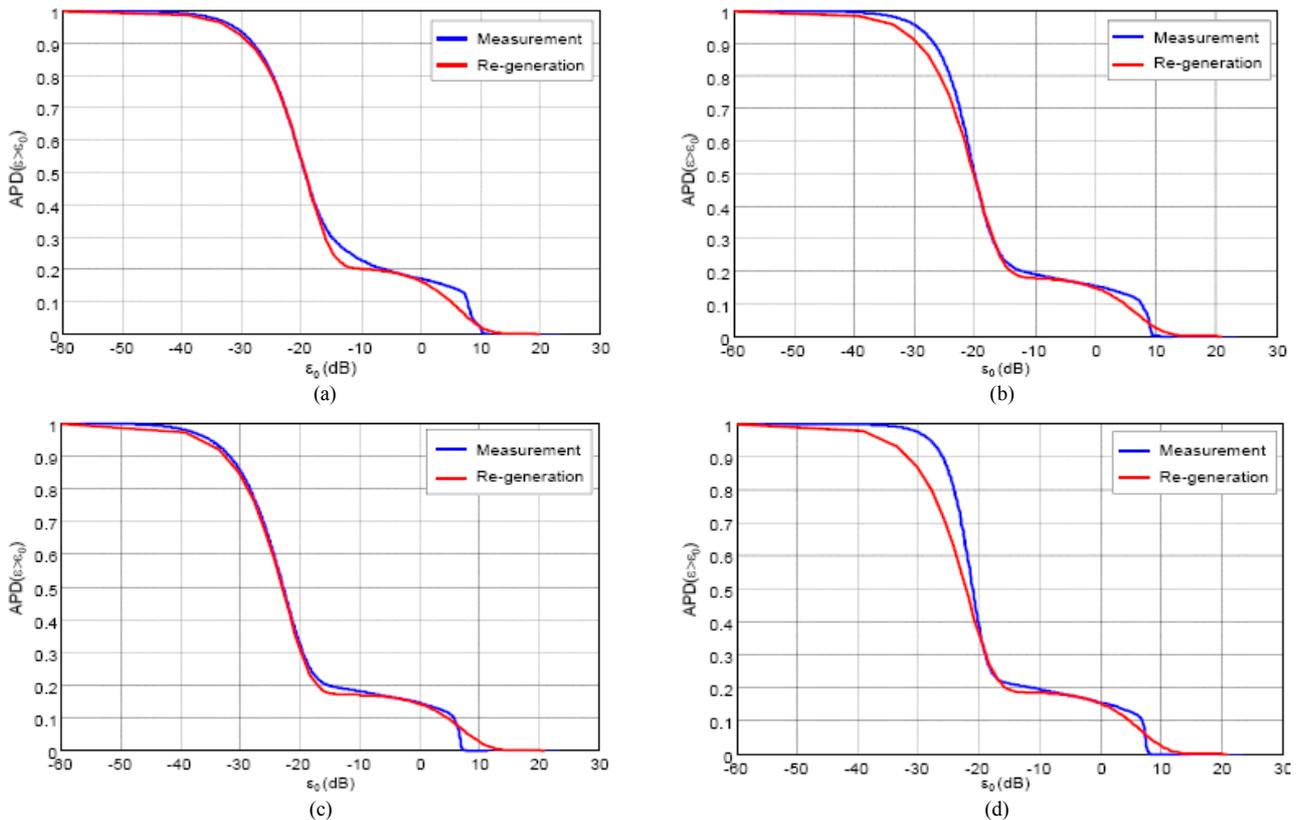


Figure 5. APD Comparison of measured multiple noises and re-generated noises (a) At 3.43MHz (b) At 5.15MHz (c) At 6.13MHz (d) At 7.11MHz

TV and multiple noise sources, they show similar values at each frequency. It means the impulsive components of multiple noise sources mostly come from the PDP TV, whereas,  $\Gamma$  parameters of the PDP TV and multiple sources have different values at each frequency. The  $\Gamma$  values of multiple noise sources are higher than those of PDP TV. It is because the non-impulsive noises from electric board are applied to the multiple noises.

#### IV. NOISE RE-GENERATION

Using the extracted noise parameters, multiple noises are re-generated. The statistical functions are determined by the parameter values of  $A$  and  $\Gamma$ , and the random noise data following the determined statistical functions are generated using a simulator tool. For each measured frequency, one million points of random data are generated. The data is downloaded to a signal generator and radiated through an antenna. The re-generated noise is measured by another antenna. Fig. 5 shows the comparisons of APD graphs of measured multiple noises and re-generated noises at each frequency. Even though there is slight difference between two APD graphs at the left side at 7.11MHz, the APD shapes of re-generated noises are generally well matched to that of measured noises.

#### V. CONCLUSIONS

The re-generation method of radiated disturbances from multiple noise sources using statistical approach was suggested. The statistical characteristics of radiated disturbance from multiple noise sources are analyzed through the APD measurement. The radiated noises from electric board and PDP TV are measured from 150 kHz to 30 MHz, and then frequencies where both noise sources radiate disturbances in common are selected. At each frequency, the APD is measured and then the characteristic parameters of the noises are extracted using the Middleton's class  $A$  noise model and approximate empirical method. Random noise data is re-generated from the extracted parameters and radiated using a signal generator and antenna. APD Comparison of the measure multiple noises and re-generated noises shows that re-generated noises reproduce the measured multiple noises with the similar statistical characteristics. The re-generated noise can be applied to the performance evaluation of electric devices against the multiple noises.

#### ACKNOWLEDGMENT

This research was funded by the MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013

REFERENCES

- [1] B. Sklar, *Digital Communications.*, 1988 :Prentice-Hall
- [2] CISPR 16-1-1, Specification for radio disturbance and immunity measuring apparatus and methods Part 1-1: Radio disturbance and immunity measuring apparatus – Measuring apparatus, Edition 3.1., Nov. 2010
- [3] CISPR 16-2-3, Specification for radio disturbance and immunity measuring apparatus and methods Part 2-3: Methods of measurement of disturbances and immunity – Radiated disturbance measurements, Edition 3.1, Aug, 2010, Jun. 2009
- [4] K. Wiklundh, “The relation between the amplitude probability distribution of an interfering Signal and its impact on digital radio receivers,” *IEEE Transaction on Electromagnetic Compatibility*, vol. 48, no. 3, Aug. 2006, pp. 537-544
- [5] S. Choi, Y. Shim, and S. Park, “APD analysis of the radiated disturbance from the plasma display panel TVs,” *International Conference on ICT convergence*, Oct. 2012, pp. 170-174
- [6] D. Middleton, “Statistical physical models of electromagnetic interference,” *IEEE transaction on Electromagnetic Compatibility*, vol. EM-19, no. 3, Aug. 1977, pp. 106-127
- [7] D. Middleton, “Procedures for determining the parameters of the first order canonical models of class A and class B electromagnetic interference,” *IEEE transaction on Electromagnetic Compatibility*, vol. EM-21, no. 2, Aug. 1979, pp. 190-208
- [8] L. A. Berry, “Understanding Middleton’s canonical formula for class A noise,” *IEEE transaction on Electromagnetic Compatibility*, vol. EM-23, no. 3, Nov. 1981, pp. 337-344
- [9] CISPR 11, Industrial, scientific and medical equipment – Radio-frequency disturbance characteristics – Limits and methods of measurement, Edition 5.1, May 2010
- [10] CISPR 16-1-4, Specification for radio disturbance and immunity measuring apparatus and methods - Antennas and test sites for radiated disturbance measurements, Ed.3.0, Apr. 2010
- [11] CISPR 15, Limits and methods of measurement of radio disturbance characteristic of electrical lightning, Ed. 7.0, Dec. 2005

# Modeling and Performance Evaluation of Scheduling Algorithms For Downlink LTE cellular Network

Bechir Nsiri, Mallouki Nasreddine, Mahmoud Ammar, Walid Hakimi, Mhatli Sofien  
 National Engineering School of Tunis, Sys'Com Laboratory, ENIT  
 Tunis, Tunisia  
 bechirnsiri@gmail.com, mallouki\_nasreddine@yahoo.fr, mahmoud.ammar@enit.rnu.tn,  
 Walid.Hakimi@enit.rn.tn, sofien\_mhatli@yahoo.fr

**Abstract**— Long Term Evolution is standardized by the 3rd Generation Partnership Project to have wider channels up to 20MHz, with low latency and packet optimized radio access technology. The peak data rate envisaged for LTE is 100 Mbps in downlink and 50 Mbps in the uplink. The 3GPP has chosen the OFDMA as the radio access technology due to his simple implementation in receiver and spectral efficiency. To enhance system's data rate and ensure quality of service, the Radio Resource Management Scheduling Mechanisms plays a very crucial components to guarantee the Quality of Service performance for different services. In this paper we modeled and evaluated the performance of Round Robin, Proportional Fairness and Max Rate scheduling algorithms. The performances are compared in term in throughput and fairness index for this scheduler.

**Keywords**- Scheduling; Fairness; Max Rate; Ressource Block.

## I. INTRODUCTION

The Long Term Evolution (LTE) is standardized by the 3GPP in Release 8, as the successor of the Universal Mobile Telecommunication System (UMTS), in order to ensure a high speed data transmission with mobility for mobile communication. The radio access technology chosen for LTE system is the Orthogonal Frequency Division Multiple Access (OFDMA), in both Time Division Duplexing (TDD) and Frequency Division Duplexing (FDD), because of the high degree of flexibility in the allocation of radio resources to the Users Equipments (UEs) and his robustness to the selectivity of multipath channels [1][2]. LTE is capable of supporting different transmission band of spectrum allocation (Multiple Channel Bandwidth), ranging from 1.4 Mhz to 20 Mhz, for both paired and unpaired bands. The high peak transmission rate reaches the LTE system is 100 Mbps in downlink (DL) and 50 Mbps in uplink (UL). To achieve the performance objectives, LTE employs the several enabling technologies which include Hybrid Automatic Repeat Request (HARQ) technical and different MIMO transmission methods are deployed [3] [4].

LTE technology presents a very challenging multiuser problem: Several User Equipments (UEs) in the same geographic area require high data rates in a finite bandwidth with low latency. Multiple access techniques allow UEs to share the available bandwidth by allocating to each UE a

fraction of the total system resources. The strong motivation beyond the resource allocation algorithms for scheduling is the improvement of system performance by increasing the spectral efficiency at the wireless interface and consequently enhancing the system capacity. Other constraints such as fairness must also be improved. Hence, it is important to find away to performance effective trade-off between efficiency and fairness. To develop an efficient scheduler to reach this trade-off, several factors must be taken into account such as: Signal-to-Interference-plus-Noise Ratio (SINR), packet delays, buffer status (queues length and packet delays), and type of service, fairness, channel conditions and complexity (time and computing).

In this paper, we study and compare the different scheduling algorithms for downlink LTE system and we discuss the factors which mentioned earlier for several proposed resources allocation schemes. This paper is organized as follow: in Section II, we describe the LTE downlink scheduling mechanism and in Section III, we evaluate and compare the algorithms performance.

## II. LTE DOWNLINK SCHEDULING

The air interface of LTE technology is based on OFDMA and SC-FDMA in the downlink and Uplink respectively to deliver the flexibility and increase data rate without additional bandwidth or increase transmit power. The base station (eNodeB) is the entity responsible for controlling the air interface between the network and user equipments. The data transmission in LTE system is organized as physical resources which are represented by a time-frequency resource grid consisting of Resources Blocks (RBs) which has a duration of 0.5 ms and a bandwidth of 180 KHz (12 subcarriers spaced with 15 KHz). It is a straight forward to see that each RB has  $12 \times 7 = 84$  resource elements in the case of normal cyclic prefix and  $12 \times 6 = 72$  resource elements in the case of extended cyclic prefix.

The scheduler entity have a role to assigns resources blocks every TTI, based on the channel condition feedback received from User Equipment in the form of Channel Quality Indicator (CQI) send by the UEs to the eNodeB, to indicate the data rate supported by the downlink channel. Every value of CQI, index in the range 1 to 15, corresponds to the highest Modulation and Coding Scheme (MCS) and the amount of redundancy included [12]. The corresponding

bit rate per bandwidth is standardized by 3GPP and is shown in TABLE I.

TABLE I. CQI TABLE

CQI Index	Modulation	Code rate X1024	Efficiency
0	No transmission		
1	QPSK	78	0.1523
2	QPSK	120	0.2344
3	QPSK	193	0.3770
4	QPSK	308	0.6016
5	QPSK	449	0.8770
6	QPSK	602	1.1758
7	16QAM	378	1.4766
8	16QAM	490	1.9141
9	16QAM	616	2.4063
10	64QAM	466	2.7305
11	64QAM	567	3.3223
12	64QAM	666	3.9023
13	64QAM	772	4.5234
14	64QAM	873	5.1152
15	64QAM	948	5.5547

In LTE system, the resource allocation is done in time and frequency domain. In time domain, the downlink channel is divided into frame of 10ms each consists of 10 subframes of 1 ms each referred to as Transmission Time Interval (TTI). In frequency domain, the available system bandwidth is divided into sub-channels of 180 KHz, comprising of 12 consecutive equally spaced OFDM sub-carriers of 15 KHz each. A time-frequency radio resource spanning over 0.5 ms slots in the time domain and over 180 KHz sub-channel in the frequency domain is called Resource Block (RB) [5]. The number of resource blocks in the available bandwidth is called Resource grid. Resource Element (RE) represents one OFDM subcarrier during one OFDM symbol interval. The number of RBs in a resource grid depends on the size of the bandwidth. The LTE operates in the bandwidth of 1.4 MHz up to 20 MHz, with number of RBs ranging from 6 to 100 respectively [5] [6].

#### A. LTE scheduling Mechanisms

In wireless communications, specifically in LTE system, each user return a value of CQI to eNodeB every TTI corresponding to the channel state of the user  $i$  and the mean data rate supported by the channel at the time slit  $t$ . The scheduler is responsible for assigning the RBs in time and frequency domain resources to the different UEs under the CQI-received as a feedback from the UE by the BS. Every 1ms the assignment of resources could change depending upon various factors including CQI for each user. In order to perform channel-aware packet scheduling, each eNodeB need to have the knowledge of Channel State Information (CSI) for each user, for all the RBs in the available bandwidth. The CSIs are derived based on channel gain, interference conditions and SINR estimation errors. In this paper, we discuss the major scheduling algorithms that are

used by the LTE downlink schedulers, they are, Round Robin (RR), Proportional Fairness (PFS) and Max-Rate algorithms. These scheduling algorithms are described in the next section [7]. The Figure 1 describes the packet scheduling strategies.

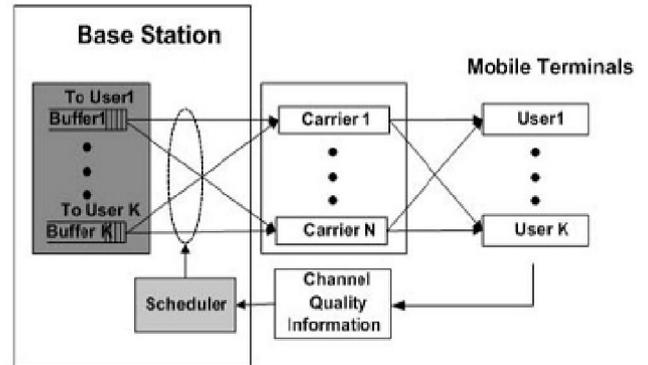


Figure 1. LTE Scheduling Scenario.

The LTE downlink scheduler is designed to ensure high Quality of service (QoS), maximization of system capacity, reducing complexity and ensures fairness between all active users. Then, scheduling algorithms should be capable to exploit the channel variation condition with maintaining fairness between the users flows.

#### B. LTE scheduling Algorithms

In LTE system, the scheduling algorithms assume that the eNodeB would receive the CQI feedback, every TTI, as a matrix with dimensions Number\_UEs x RB\_grid\_size. The value of each field in the matrix is the CQI feedback of each user for each RB [6]. The different scheduling algorithms are describes as follow:

- **Round Robin (RR)** is the simplest scheduling algorithm which assigns time interval to each mobile station in equal portion and in order with the same priority. This algorithm is very simple and easy to implement [14].
- **Proportional Fairness Algorithm (PFS)** work as follow. The CQI feedback of user  $k$  in time TTI  $i$  is in term of a requested data rate  $R_{k,n}(t)$ , which design the  $k^{th}$  user's sub-carrier can currently supported. The PFS algorithm keep track of the average throughput  $T_{k,n}(t)$  of each user on every sub-carrier in a past window of length  $t_c$ . The  $t_c$  parameters means the trade-off between fairness and throughput. The larger value of  $t_c$  is  $t_c = \infty$ , in this situation the allocation resources according to PFS algorithm is decided solely by instantaneous SNR, leading to maximum system throughput and poor fairness characteristics. On the other hand, the lower value of  $t_c$  parameter is  $t_c = 1$  in this situation scheduling becomes fair [8] [9]. In the time slot  $t$ , the PFS algorithm transmits at each sub-carrier to the user  $K$  with the largest value of  $J$  calculated as follow:

$$j = \frac{R_{k,n}(t)}{T_{k,n}(t)} \quad (1)$$

The average throughput  $T_{k,n}(t)$  can be updated using an exponentially weighted low-pass filter [13] :

$$T_{k,n}(t+1) = \begin{cases} (1 - \frac{1}{t_c})T_{k,n}(t) + \frac{1}{t_c}R_{k,n}(t) & k = k^*(t) \\ (1 - \frac{1}{t_c})T_{k,n}(t) & k \neq k^*(t) \end{cases} \quad (2)$$

- **Max-Rate** scheduler transmit, every TTI, to the user having the largest SNR, so users that have the fading peak are likely to be scheduled all the time, while other that experience deep fades are not scheduler at all. Max Rate scheduler has to maximize system throughput but it totally ignores fairness. The received SNR of the  $n^{th}$  RB signal of the  $k^{th}$  user at the  $t^{th}$  TTI can be expressed by[8]:

$$SNR_{k,n}(t) = \frac{S_{k,n}(t) H_{k,n}(t)}{N_0 B/N} \quad (3)$$

Where  $S_{k,n}(t)$ ,  $H_{k,n}(t)$  are the allocated transmission power and channel gain on  $n^{th}$  sub-carrier at  $t^{th}$  TTI respectively,  $N_0$  is the power spectral density of AWGN, B is the bandwidth and N is the number of sub-carriers.

The instant data rate of each user is determined and the BS serves each user at this rate. The instant service rate on the  $n^{th}$  sub-carrier at  $t^{th}$  TTI is got by:

$$R_{k,n}(t) = B/N \log_2(1 + SNR) \quad (4)$$

Where,  $R_{k,n}(t)$  is the  $k^{th}$  user transmission rate at  $t^{th}$  time slot,  $B$  is the total bandwidth and  $N$  is the number of sub-carriers [10] [11].

PFS algorithm transmits for each user when its channel is good and at the same time the scheduling algorithm is perfectly fair on the long term. We plot in Fig. 2 the frequency response of three users. Thus, the PFS algorithm schedules a user when its instantaneous channel quality is high relative to its own average channel condition over the time scale  $t_c$ .

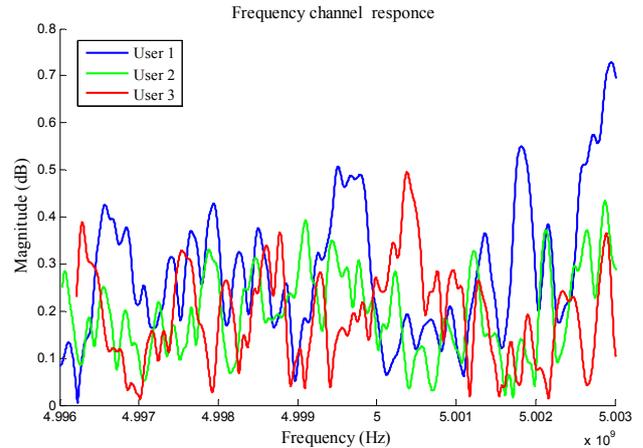


Figure 2. Frequency Channel Response for Three users.

### III. SIMULATION RESULTS AND DISCUSS

#### A. Simulation Parameters

In this section, we will simulate and discuss the performance of the three scheduling algorithms, such as RR, PFS and Max-Rate, over LTE system. The simulations are carried out for frequency-selective channels modeled by ITU for Pedestrian-B (Ped-B) channels. Our simulations are performed for users ranging from 5 to 25, choosing the bandwidth of 5MHz containing 25 RBs and 300 occupied sub-carriers. The simulations parameters used are listed in the TABLE II.

TABLE II. SIMULATION PARAMETERS

Parameters	Value
Channel type	ITU-Pedestrian B
Number of Base station	1
Number of users	5, 10, 20, 30, 40
Scheduling Algorithms	Round Robin Max-Rate Proportional Fairness
Bandwidth (MHz)	5
Transmission mode	SISO
Numbers of subframes	140
Tc parameter	1

#### B. Simulation Results and Discuss

In this section, we present the simulation evaluation of scheduling algorithms. In order to evaluate and find the scheduler disciplines, the performance is measured in term of overall system capacity and fairness index using the three scheduling algorithms. First, we plot the number of allocated RBs for every user over time using each scheduling algorithm.

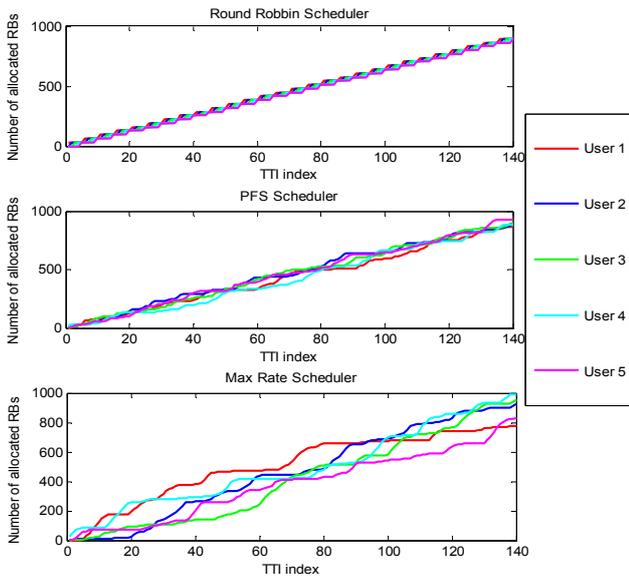


Figure 3. Number of allocated RBs for each user vs. TTI index using RR, PFS and Max Rate schedulers.

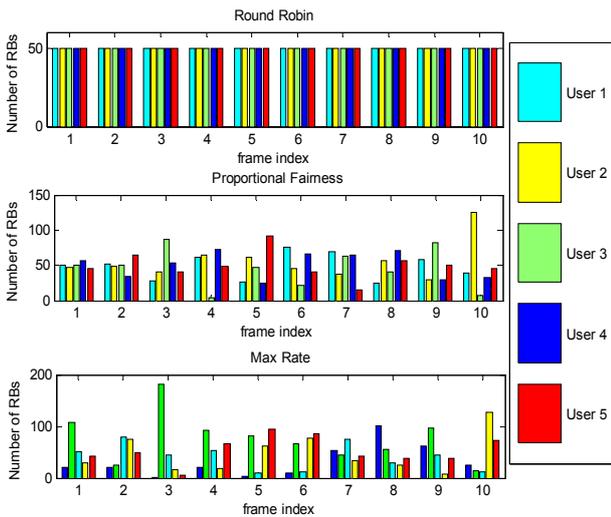


Figure 4. Means allocated RBs per frame for each user using RR, PFS and Max Rate schedulers.

Figures 3 and 4 show the evolution of allocated RBs versus time and means number of RBs per frame, respectively, allocated for each user using RR, PFS and Max-Rate algorithms. From the simulation results, we can see that the RR scheduler delivers fairness for all users with allocates the same number of RBs for each user which have 50 Resources Block each frame. But, Max-Rate scheduler allocates a different number for each user derived from channel quality SNR to maximize average system throughput. The PFS scheduler tries to strike a balance between fairness and achieving the Maximum throughput by allocation almost equal means RBs for each user every TTI.

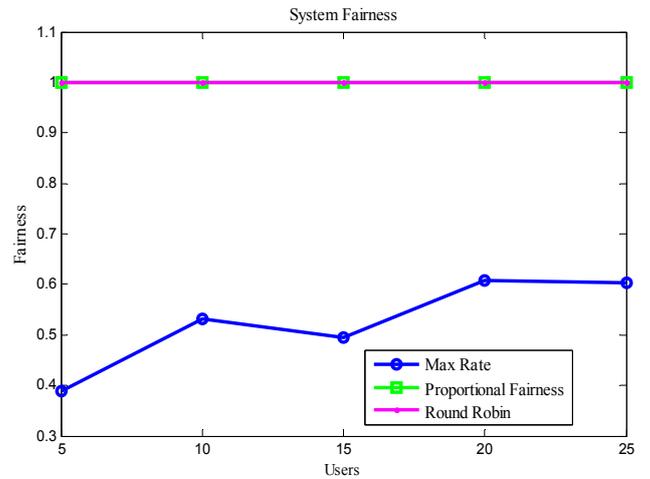


Figure 5. Number of allocated RBs for each user vs. TTI index using RR, PFS and Max Rate schedulers.

The system fairness for scheduling algorithms versus users ranging from 5 to 25 is investigated in Fig. 5. It is also observed that Max rate is the unfairness algorithm and the index fairness not exceed 0.6. This is because the Max rate algorithm allocates the system resources to users who have a strongest channel and serve the users who are demanding service in system. But as we can see, the PFS scheduler has a constant fairness index almost equal to 1 because as we saw in section II, with a low  $t_c$  parameter this algorithm maintenance index fairness without involving system throughput.

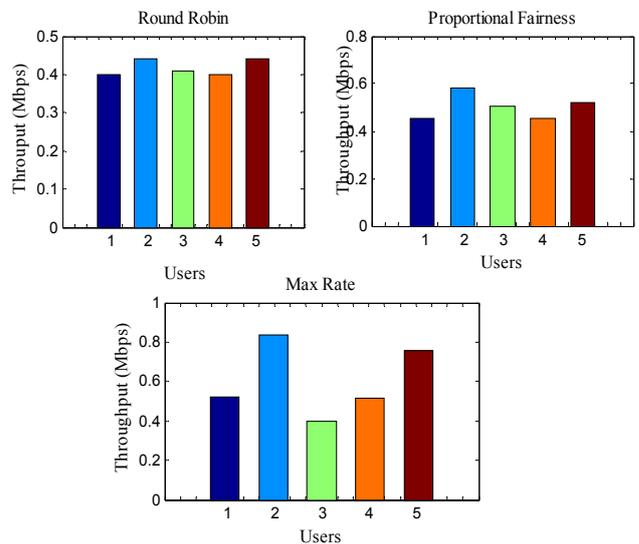


Figure 6. Total Throughput for each user using RR, PFS and Max rate scheduler, 5 Users

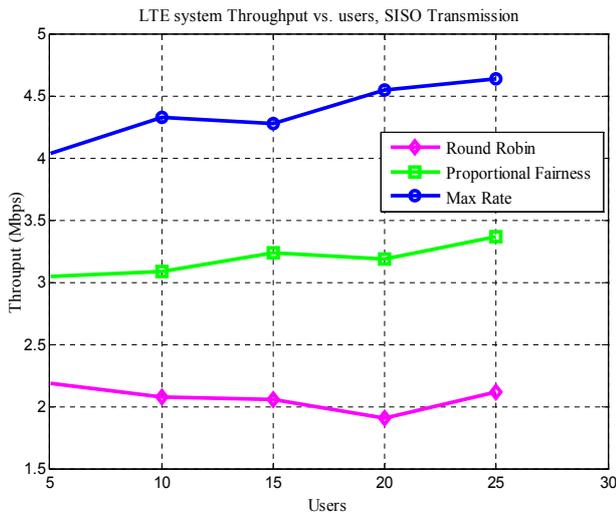


Figure 7. System Throughput versus user using RR, PFS and Max Rate schedulers

In order to compare the different scheduling algorithms, we simulate and plot the average throughput for each user and system throughput versus users ranging from 5 to 25. We can see that the system throughput achieved by RR algorithm reaches the lowest value because this algorithm allocates all sub-carriers to one user at each time slot independently of users' channel response and rate requirements. But Max Rate algorithm, according to system throughput, reaches the best result because this algorithm allocates system resources to users with the strongest channel and it maximizes the system throughput. PFS algorithm exploits the propagation channel condition in order to maximize system capacity without comprising fairness. As we can see in Figs. 6 and 7, PFS algorithm has a good behavior because it reaches a good level of system throughput.

#### IV. CONCLUSION

In this paper, we present the performances of three scheduling algorithms such as Round Robin, Proportional Fairness and Max Throughput in term of fairness and system capacity. We can see that the RR scheduler promotes priority to fairness between all users regardless of system throughput. On the other hand, Max throughput is used to maximize the system capacity without considering the fairness among users. But, from the results obtained, it is also observed that the proportional fairness algorithm performs a compromise between system fairness and throughput. From this result, we can use a mixed between PFS and Max Throughput scheduler to maximize system throughput with guaranteed fairness between users.

#### REFERENCES

[1] 3GPP TS 36.212 Version 10.0.0 Release 10, Evolved Universal Terrestrial Radio Access (E-UTRA): Multiplexing and channel coding, March, 2009.

[2] C. Mehlführer, M. Wrulich, J. Colom Ikuno, D. Bosanska, and M. Rup, "Simulating the Long Term Evolution Physical Layer," in Proc. EUSIPCO 2009, March 2009, pp.1471 – 1478.

[3] M. AMMAR, B. NSIRI, W. HAKIMI and M. ELJAMAI, "A comprehensive study of Open-loop Spatial Multiplexing and Transmit Diversity for Downlink LTE", International Journal of Computer Sciences and Telecommunication (IJCST), volume 5 Issue 2, January 2014, pp. 1-7.

[4] Technical White paper: "Long Term Evolution (LTE): A Technical Overview," by Motorola.

[5] S. M. Chadchan, "A Fair Downlink Scheduling Algorithm for 3GPP LTE Networks", I. J. Computer Network and Information Security, 6, pp 34-41, 2013.

[6] D. Talevski and Li. Gavrilovska, "Novel Scheduling Algorithms for LTE Downlink Transmission", Telfor Journal, Vol. 4, No. 1, February 2012, pp. 45-56.

[7] Dinesh Mannani, "Modeling and Simulation of Scheduling Algorithms in LTE Networks", Bachelor of Science Thesis report, Warsaw, January 2012.

[8] Al. Bahillo Martinez, "Evaluation of multiuser scheduling algorithm in OFDM for different services", Master of Science in Electronics report, June 2006.

[9] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas" IEEE transactions on information theory, vol. 48, no. 6, June 2002.

[10] H. Seo and B. Gi Lee, "A Proportional Fair Power Allocation for Fair and Efficient Multiuser OFDM Systems", School of Electrical Engineering, Seoul National University, April 2004.

[11] E. L. Hahne, "Round-Robin Scheduling and Window Flow Control for Max-Min Fairness in Data Networks" IEEE, technical report, November 1987.

[12] Raymond H. Y. Louie, "Open-Loop Spatial Multiplexing and Diversity Communications in Ad Hoc Networks", Revised version. Submitted to IEEE Tr. Info Theory, 17 september 2010.

[13] Institute of Telecommunications Vienna University of Technology, Austria Gusshausstrasse 25/389, A-1040 Vienna, Austria, "Vienna LTE Simulators System Level Simulator Documentation, v1.7r1119".

[14] C. Mehlführer, J. Colom Ikuno, M. Šimko, S. Schwarz, M. Wrulich and M. Rupp, "The Vienna LTE simulators – Enabling reproducibility in wireless communications research", EURASIP Journal on Advances in Signal Processing 2011, pp 1-14.

# Channel Estimation for Downlink LTE System Based on LAGRANGE Polynomial Interpolation

Mallouki Nasreddine, Nsiri Bechir, Walid Hakimi and Mahmoud Ammar

University of Tunis El Manar, National Engineering School of Tunis, LR99ES21 Lab. Sys' Com, ENIT  
Tunis, Tunisia

Email : {mallouki\_nasreddine@yahoo.fr, bechirsiri@gmail.com, Walid.Hakimi@enit.rnu.tn, mahmoud.ammar@enit.rnu.tn}

**Abstract**—Long Term Evolution (LTE) uses different techniques to achieve high throughput required, such as the HARQ techniques, Multiple-Input Multiple-Output (MIMO) transmission schemes and estimation techniques. In the present paper, research focuses on Downlink LTE channel estimation which is based on the interpolation to estimate channel coefficients. Thus, we propose an adaptive method for polynomial interpolation based on Lagrange polynomial. We perform the Downlink LTE system for Single-Input Single-Output (SISO) and MIMO transmission then compare the obtained results with linear and Sinus Cardinal Interpolations. The simulation results show that the Lagrange method outperforms the linear interpolation in term of Block Error Rate (BLER) and throughput vs. Signal to Noise Ratio (SNR).

**Keywords**—LTE; MIMO; SISO; Interpolation; Lagrange

## I. INTRODUCTION

In modern world, requirement of high data rate communication has become inevitable. Applications such as streaming transmission, video images, and World Wide Web browsing require high speed data transmission with mobility. In order to fulfill these data requirements, the 3rd Generation Partnership Project (3GPP) [1][2] introduced Long Term Evolution (LTE), to provide high speed data rate for mobile communication. The LTE system affords an important effective bit rate and allows increasing system capacity in terms of numbers of simultaneous calls per cell. In addition, it has a low latency compared to 3G/3G+ networks. It offers a theoretical speed of 100 Mbits/s in the Downlink and 50Mbits/s in the Uplink transmission. The LTE uses Orthogonal Frequency Division Modulation (OFDM) and Orthogonal Frequency Division Modulation multiple access technique (OFDMA) in the downlink transmission [3]. The OFDM provides the signal transmitted robustness against the multipath effect and can improve the spectral efficiency of the system [4][5]. On the other hand, the implementation of MIMO system increases channel capacity and decreases the signal fading by sending the same information at the same time through multiple antennas[5]. The combination of these two powerful technologies (MIMO-OFDM) in the LTE system improving thus the spectral efficiency and throughput offered without increasing resources for base bands and power output. To

best exploit the power of MIMO-OFDM technology, it is imperative to manage at best the estimation of the channel coefficients; this operation is ensured by the interpolation of pilots.

In this paper, we represent a polynomial interpolation algorithm using the method of Lagrange [10] which greatly reduces the complexity of the transceiver. The simulation is made on a 'Vehicular A' (Veh A) [11][12] channel through SISO and MIMO system using Least Square equalizer (LS). Section II of this paper give an over view of MIMO-OFDM transmission. In Section III, we present Lagrange interpolation algorithm. Finally, Section IV provides the numerical results.

## II. MIMO-OFDM TRANSMISSION

### A. MIMO OFDM transmissions schemes

In this section, we are going to describe the MIMO OFDM transmission scheme. A modulation block is used to modulate the original binary data symbol using the complex constellation QPSK, 16 QAM or 64 QAM according to the LTE standard [6][7]. Pilot insertion is generated according to the LTE standards, followed by Inverse Fast Fourier Transform operation (IFFT); at the end, a cyclical prefix is inserted to remedy the phenomenon of the Inter Symbol Interference (ISI) and the Inter Sub carriers Interference. Transmission is made through a multipath Fast Fading channel over a multiple antenna system. Multiple antennas can be used in the transmitter and the receiver; consequently, MIMO encoders are needed to increase the spatial diversity or the channel capacity. Applying MIMO allows us to get a diversity gain to remove signal fading or getting a gain in terms of capacity. Generally, there are three types of MIMO receivers, as presented in [1]. At the reception, the cyclical prefix is firstly removed, followed by the Fast Fourier Transform operation (FFT); after the extraction of pilots, parameters of channel is estimated through the block interpolation followed by equalization. The method of interpolation chosen is essential to make the estimation more efficient and to reduce the equalizer complexity.

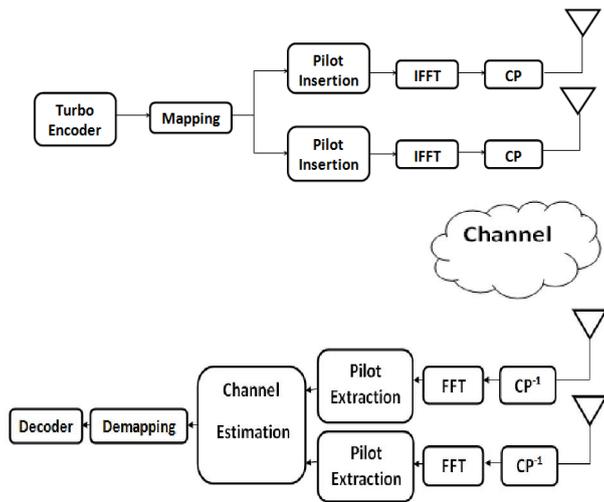


Fig. 1. MIMO-OFDM transmission.

B. Analysis of standard LTE pilot scattering

In the LTE standards, pilots are placed in a well-defined ways to cover up the frequency and time domain. The location of pilots for 2x2 MIMO transmissions scheme in LTE system is shown in the following figures.

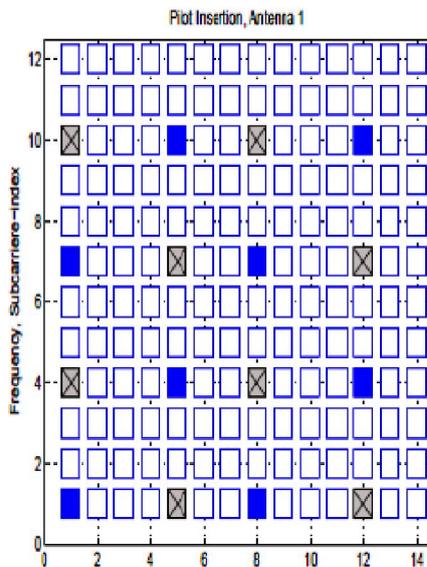


Fig. 2. Pilot structure of Transmitter

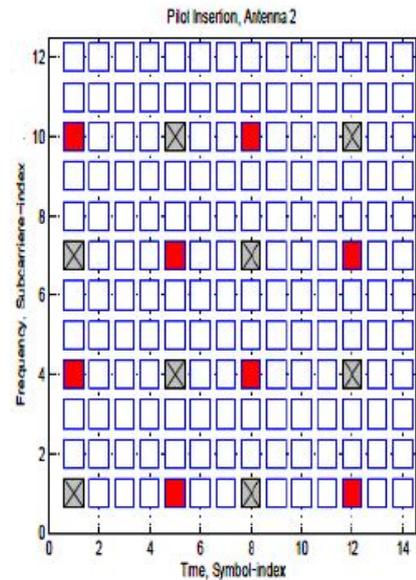


Fig. 3. Pilot structure of Transmitter 2.

It can be seen that, through the first antenna, pilots are disposed in OFDM symbols numbers 1, 5, 8 and 12 while for the second antenna, they are placed in the same OFDM symbols, but in different subcarriers index.

Those positions allow a better coverage of the frequency and time and reduce the risk of interference in reception [3].

III. DESCRIPTION OF THE INTERPOLATION ALGORITHM

A. Linear Interpolation

In linear polynomial interpolation, the channel coefficients are estimated using the linear relationship between two successive pilots.

Linear interpolation is given by the following expression:

$$H_{(k)}^{(i)} = i/d * H_{(k)(p)}^{(i)} + (1 - i/d) * H_{(k)(p+1)}^{(i)} \quad (1)$$

where  $H_{(k)}^{(i)}$  is the channel coefficient to estimate,  $H_{(k)(p)}^{(i)}$  and  $H_{(k)(p+1)}^{(i)}$  two successive pilots,  $i$  is the subcarriers index,  $k$  is the OFDM symbol index,  $p$  is the pilot index and  $d$  is the distance between two pilots [8].

B. Sinus Cardinal Interpolation

Sinus Cardinal (SinC) interpolation is given by the following expression [7]:

$$S(x) = \sum_{k=0}^n S(k) \text{SINC}(x - k) \quad (2)$$

where  $S(k)$  is the pilots,  $k$  is the position of  $y$ ,  $S(x)$  is the SinC interpolation function.

In this work, we use 2 pilots to estimate channel coefficients using SinC interpolation.

The interpolation is represented as follow:

1-Extract received  $y_{(k)(p)}^{(i)}$  pilots from received signal  $y_{(k)}^{(i)}$

2- Calculate the channel coefficients of pilots symbols with Least Square estimator

$$H_{(k)(p)}^{(i)} = y_{(k)(p)}^{(i)} / x_{(k)(p)}^{(i)} \quad (3)$$

3- Estimate  $H_{(k)}^{(i)}$  with SinC interpolation:

$$H_{(k)}^{(i)} = \text{sinc}(x - x_{p_0}) * H_{(k)(p_0)}^{(i)} + \text{sinc}(x - x_{p_1}) \quad (4)$$

C. Lagrange polynomial Interpolation

Lagrange polynomials allow interpolating a set of points by a polynomial which goes exactly through these points. The Lagrange polynomial is given by the following expression [10]

$$P(x) = \sum_{i=0}^n y_i L_i(x) \quad (5)$$

$$L(x) = \prod_{j=0, j \neq i}^n (x - x_j) / (x_i - x_j) \quad (6)$$

Where  $y_i$  the pilots,  $x$  is the position of  $y$ ,  $L$  is the coefficients of Lagrange and  $n$  is the Lagrange polynomial order.

D. Algorithm description

The received signal for MIMO system model consisting of  $N_T$  transmits antennas and  $N_R$  receives antennas can be represented by the following Equation:

$$Y_{(k)}^{(i)} = X_{(k)}^{(i)} H_{(k)}^{(i)} + N \quad (7)$$

where  $Y_{(k)}^{(i)} = [y_0^{Nsc} \dots y_0^{Nsc} \dots y_{N_{OFDM} SYM N_R}^{Nsc}]$  is the received vector,  $H_{(k)}^{(i)}$  is the channel coefficient matrix of the dimensions  $N_T \times N_R$  express the channel gain and  $N = [n_1, n_2, \dots, n_{N_R}]^T$  is the noise vector.

The matrix  $H_{(k)}^{(i)}$  is written as follow [11]:

$$H_{(k)}^{(i)} = \begin{pmatrix} H_{(k)}^{(i)} & h_{(k)1,2}^{(i)} & \dots & h_{(k)1,N_T}^{(i)} \\ h_{(k)2,1}^{(i)} & h_{(k)2,2}^{(i)} & \dots & h_{(k)2,N_T}^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{(k)N_R,1}^{(i)} & h_{(k)N_R,2}^{(i)} & \dots & h_{(k)N_R,N_T}^{(i)} \end{pmatrix} \quad (8)$$

For each reception antennas, after eliminating Cyclical Prefix and Fast Fourier Transform operation, pilots are extracted and then interpolation block is attacked to estimate the parameter  $H_{(k)}^{(i)}$  of the channel. The interpolation operation is necessary for both frequency and time domain.

In the present work, we use a Lagrange polynomial interpolation for frequency domain and linear interpolation

for temporary. The interpolation algorithm is represented as follow:

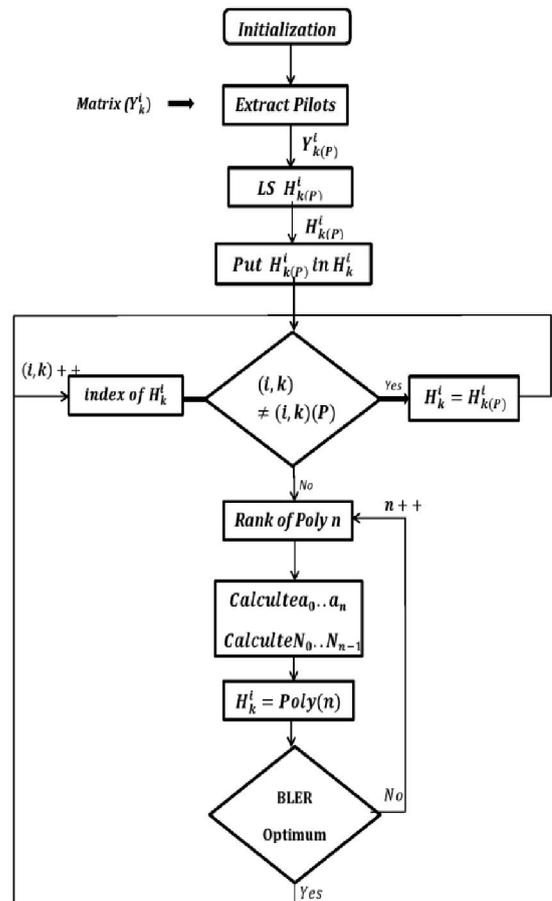


Fig. 4. Algorithm of Interpolation.

- 1- Extract  $y_{(k)(P)}^{(i)}$  pilots from received signal  $y_{(k)}^{(i)}$
- 2- Calculate the channel coefficients of pilots symbols with Least Square estimator

$$H_{(k)(p)}^{(i)} = y_{(k)(p)}^{(i)} / x_{(k)(p)}^{(i)} \quad (9)$$

3- Calculate  $L_0, \dots, L_n$  Coefficients of Lagrange with  $n$  order of Lagrange polynomial and  $p$  index of pilots, we start with  $n=2$ . For example for 12 first coefficients to estimate we use 3 first pilots placed respectively at  $x_{p_0} = 0$ ,  $x_{p_1} = 6$  and  $x_{p_2} = 12$  frequency index

$$L_0 = ((x_i - x_{p_1}) * (x_i - x_{p_2})) / ((x_{p_0} - x_{p_1}) * (x_{p_0} - x_{p_2})) \quad (10)$$

$$L_1 = ((x_i - x_{p_0}) * (x_i - x_{p_2})) / ((x_{p_1} - x_{p_0}) * (x_{p_1} - x_{p_2})) \quad (11)$$

$$L_2 = ((x_i - x_{p_0}) * (x_i - x_{p_1})) / ((x_{p_2} - x_{p_0}) * (x_{p_2} - x_{p_1})) \quad (12)$$

where  $L_0$ ,  $L_1$  and  $L_2$  are the coefficients of Lagrange polynomial of order  $n=3$ ,  $x_i$  is the frequency index of  $H_{(k)}^{(i)}$  to estimate and  $x_{p_0}, x_{p_1}, x_{p_2}$  are the frequency index of first tree pilots.

4-Estimate  $H_{(k)}^{(i)}$  with Lagrange polynomial:

$$H_{(k)}^{(i)} = L_0 * H_{(k)(p)}^{(i)} + L_1 * H_{(k)(p+1)}^{(i)} + L_2 * H_{(k)(p+2)}^{(i)} \quad (13)$$

where  $H_{(k)(p)}^{(i)}$ ,  $H_{(k)(p+1)}^{(i)}$  and  $H_{(k)(p+2)}^{(i)}$  are three successive pilots.

5-Testing the estimation operation performance by incrementing the polynomial of order  $n$  until having optimal performance. For our simulation, the optimal performance was noticed at a value of  $n=5$ .

#### IV. SIMULATIONS RESULTS

Our simulation was performed for LTE downlink transmission through a channel which uses the profile of ITU-Veh A for SISO and MIMO system with use of 16 QAM (CQI=7) constellation. We show simulation results for known channel, Lagrange polynomial interpolation algorithm, SinC interpolation and linear interpolation. All simulations are used over a Least Square equalizer. Simulation results are compared in term of Block Error Rate (BLER) and Throughput vs. SNR. Both Systems are simulated using the parameters shown in TABLE I [11][12].

TABLE I. PARAMETERS SIMULATION

Transmission Bandwidth	1.4 MHz
Carrier Frequency	2.1 GHz
Data Modulation	16QAM (CQI 7)
Channel	ITU-Veh A
Interpolation	Polynomial interpolation OF LAGRANGE

##### A. Simulation results and discussion

To observe the effect of the Lagrange polynomial interpolation compared with linear and SinC interpolation techniques, we simulate and trace the performance of LTE Downlink system in SISO and MIMO transmission over multipath channel (ITU-Veh A) using an LS equalizer. The simulations have been carried out for the 16-QAM (CQI=7). The Block Error Rate (BLER) and throughput vs. SNR results were study. Figures 5 and 6 show Block Error Rate vs. SNR for known channel, Lagrange polynomial, Sinus Cardinal and linear interpolations for both SISO and MIMO transmission.

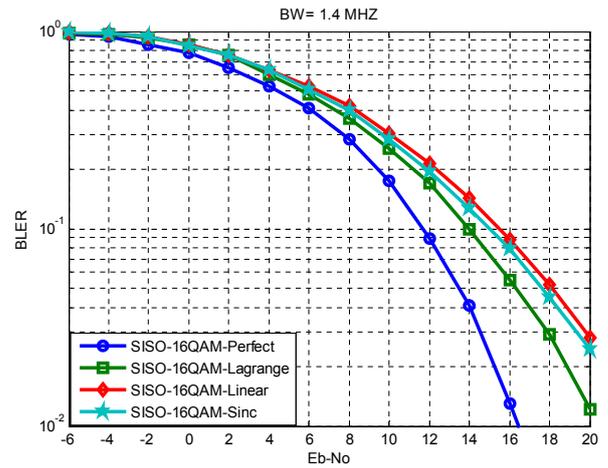


Fig. 5. BLER vs. SNR for SISO Transmission over Veh-A channel, CQI=7.

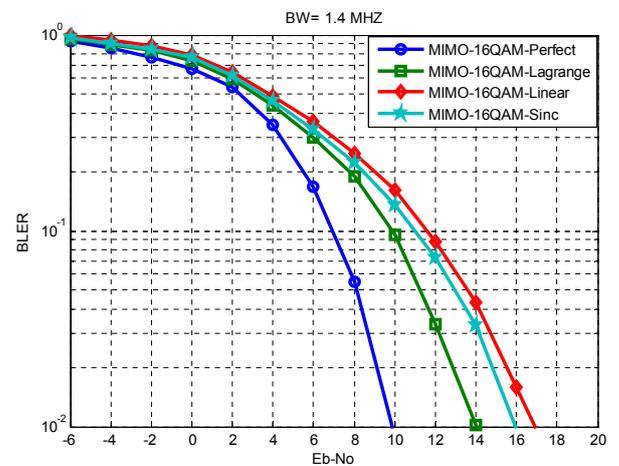


Fig. 6. BLER vs. SNR for MIMO Transmission over Veh-A channel, CQI=7.

The BLER results of LTE downlink transmission for SISO and MIMO transmission scheme is shown in Fig. 5 and Fig. 6. We can see that the Lagrange polynomial interpolation enhances the performance of downlink LTE system by almost 2 dB for  $BLER = 10^{-1}$ .

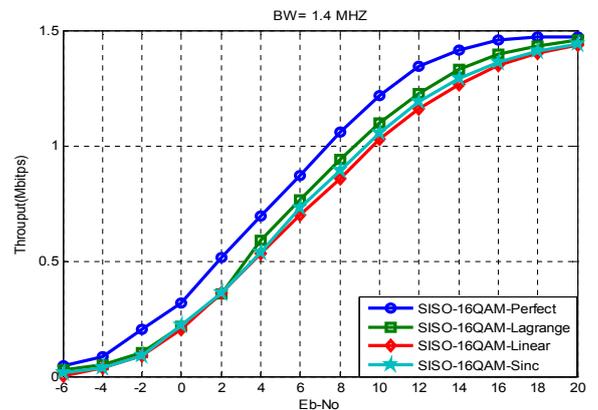


Fig. 7. Throughput vs. SNR for SISO Transmission over Veh-A channel, CQI=7.

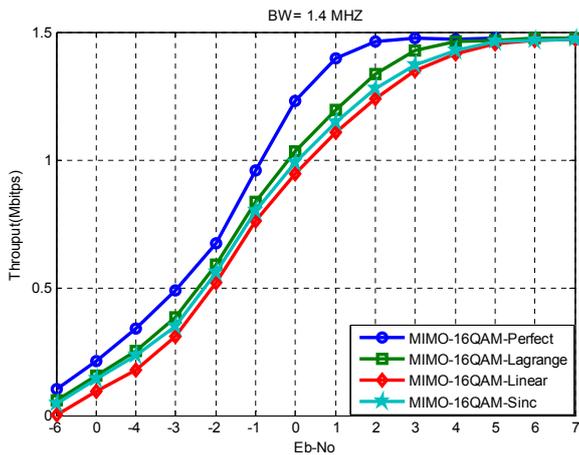


Fig 8 .Throughputvs. SNR for MIMO Transmission over Veh-A channel, CQI=7.

As shown in Figures 7 and 8 the throughput of SISO and MIMO transmission for CQI=7 over Vehicular A channel can be compared. We can note that the suggested algorithm of interpolation improve throughput compared to the linear and Sinus Cardinal interpolation. For example, with throughput=1 MHz we have a gain almost than 1dB for SISO and MIMO systems.

After studying performance, we find that the Lagrange polynomial interpolation offers a significant improvement compared to the linear and Sinus Cardinal Interpolation, as a result of the precision given by using  $n$  pilots to estimate each parameter channel. In fact, the use of  $n$  pilots in estimation of the channel coefficients takes into account the correlation between the pilot subcarriers; which makes this calculation more accurate and thus enhances the system efficiency.

It is obvious that this polynomial interpolation algorithm is more complex than the linear and Sinus Cardinal Interpolation, however it significantly improves system performance especially in the case of a fast fading channel (our case).

On the other hand, this algorithm has the advantage of having an adaptable order of polynomial interpolation  $n$  according to transmission condition. For example, in our case, we use a channel ITU-Veh A in the Bandwidth of 1.4 MHz where we have  $n = 5$  for same Bandwidth but for ITU-PEDISTRIAN-B channel  $n = 4$ .

## V. CONCLUSION

In the present work , adaptive polynomial interpolation algorithm was described in relation with the method of Lagrange for Downlink LTE system. Simulation is achieved through an ITU-Veh A channel with CQI = 7 and for SISO and MIMO systems .We conclude that, despite the complexity of this algorithm (compared to the linear and Sinus Cardinal Interpolation), it offers a considerable improvement of the performance of Downlink LTE system. In effect, using a maximum number of pilots to estimate the parameters of the channel (5 in our case) against two for a linear interpolation optimize considerably the estimation of these parameters.

## REFERENCES

- [1] 3rd Generation Partnership Project, Technical Specification Group Radio Access Network; evolved Universal Terrestrial Radio Access(UTRA): Base Station (BS) radio transmission and reception, pp.22 – 33, TS 36.104, V8.7.0, 2009.
- [2] 3rd Generation Partnership Project, Evolved Universal Terrestrial Radio Access (E-UTRA); User Equipment (UE) radio transmission and reception, pp. 22 – 33, ARIB STD-T63-36.101, V8.4.0, 2008.
- [3] 3rd Generation Partnership Project, Technical Specification Group Radio Access Network; evolved Universal Terrestrial Radio Access (UTRA): Physical Channels and Modulation layer, pp. 55 – 67, TS 36.211, V8.8.0, 2009.S. Caban, Ch. Mehlhübler, M. Rupp, M. Wriulich, “Evolution of HSDPA and LTE”, Ltd. Published 2012 by John Wiley & Sons,.
- [4] S. Sesia, I. Toufik, and M. Baker, LTE – The UMTS Long Term Evolution from Theory to Practice, 1<sup>st</sup> ed, John Wiley and sons, LTD UK;2009
- [5] Technical White paper: “Long Term Evolution (LTE): A Technical Overview,”byMotorola.[http://www.motorolasolutions.com/web/Business/Solutions/Industry%20Solutions/Service%20Providers/Wireless%20Operators/LTE/\\_Document/Static%20Files/6834\\_MotDoc\\_New.pdf](http://www.motorolasolutions.com/web/Business/Solutions/Industry%20Solutions/Service%20Providers/Wireless%20Operators/LTE/_Document/Static%20Files/6834_MotDoc_New.pdf)
- [6] Z. Lin, P. Xiao, B. Vucetic, and M. Sellathurai, “Analysis of receiver algorithms for lte scfdma based uplink systems,” *IEEE Transaction on Wireless Communications*, vol. 9, pp.60–65, 2010.
- [7] J. F. ValenzuelaValdes, M. A. Garcia Fernandez, A. M. Martinez Gonzalez, and D. A.Sanchez- Hernandez, “Evaluation of true polarization diversity for mimo systems,” *IEEE Transaction onAntennas and Propagation*, vol. 58, pp. 2746–2755, 2009..4
- [8] Niru Desai, G. D. Makawana, “Space Diversity for Wireless Communication System– A Review”, *International Journal of Engineering Science and Innovative Technology (IJESIT)* Volume 2, Issue 3, May 2013.University of RENNES internalaggregationUFR MATHEMATICS, (2010).
- [9] Technical White paper, “Sampling Theory and Spline Interpolation”,Openstax cnx,URL: <http://cnx.org/content/m11126/latest/>
- [10] University of RENNES internal aggregation UFR MATHEMATIQUES, (2010)pp. 1 – 3.
- [11] 3GPP, TR25.996, V 11.0.0.0,” Spatial channel Model for Multiple Input Multiple Output (MIMO)”, 2012.

## Reception for Layered STBC Architecture in WLAN Scenario

Piotr Remlein

Chair of Wireless Communications  
Poznan University of Technology  
Poznan, Poland  
e-mail: remlein@et.put.poznan.pl

Hubert Felcyn

Chair of Wireless Communications  
Poznan University of Technology  
Poznan, Poland  
e-mail: hubert.felcyn@gmail.com

**Abstract**—In this paper, reception for multi-stream Orthogonal Frequency Division Multiplexing (OFDM) transmission is analyzed. The system architecture employs Linear Dispersion Space-Time Block Codes (LD-STBC). In the transmitter, a part of spatial streams is Space Time Block Coded (STBC). The LD-STBC-VBLAST OFDM receiver is described and analyzed. The quality of reception for Wireless-LAN (WLAN) transmission with channel type E is investigated using a computer simulation. We present the simulation results for two models of OFDM receivers. Performance of VBLAST and LD-STBC-VBLAST receivers has been compared. The Bit Error Rate (BER) and Packet Error Rate (PER) have been determined for different numbers of spatial streams in use. The results illustrate that the LD-STBC-VBLAST OFDM receiver improves the transmission quality in WLAN scenario.

**Keywords**—multi-antenna transmission; receiver; space time block coding; wireless networks.

### I. INTRODUCTION

In recent years, WLANs have gained on popularity. This is due to the fact that thanks to the advanced technologies they already offer high quality (with low error rate) and high speed transmissions. Simultaneously, constant grow of demand in even higher network throughput and quality transmission are observed. Therefore, insightful research on WLANs is necessary to change the existing standards [1]. As theoretical and practical research carried out lately [2-4] has shown, transmission through multi-path wireless channels may improve the system's capacity if used adequately. According to B  olskei and Paulraj [3] a Multiple Input Multiple Output (MIMO) system enables increasing of a wireless channel's capacity proportionally to the growing number of transmit and receive antennae. A practical implementation of a MIMO system is shown by Wolniansky et al. [5]. It is the so-called Vertical Bell Laboratories Layered Space Time (VBLAST) system, which has a simple structure, yet it offers high spectral efficiency. In VBLAST, a single data stream is divided into several sub-streams transmitted simultaneously by several antennae as a result of which transmission speed may be improved.

Literature suggests many options to form receivers that would receive signals transmitted in MIMO system [2-5]. One of the methods is Maximum Likelihood (ML). This detection method offers the lowest error rate but is rather

difficult to implement. Wolniansky et al. [5] propose MIMO signal detection based on the Zero-Forcing (ZF) criterion. The ZF method is characterized by relatively low computational requirements. However, its weakness is certainly the so-called noise enhancement occurring in the case of minor SNR values. Considerably effective detection algorithms that use the so-called  $QR$  decomposition of channel matrix have been proposed in [6] [7]. Another advantage of MIMO transmission is quality improvement with reference to drop in error rate. This is obtained by using Space Time Block Codes (STBC) [8].

The superior purpose of spatial multiplexing is to maximize data transmission speed while the essence of space-time coding is to ensure high quality resulting from maximizing the diversification. These two advantages offered by MIMOs exclude each other. The so-called Linear Dispersion (LD) method was proposed by Hassibi and Hochwald [9]. The method attempts to use both the aforesaid advantages of MIMO transmission: spatial multiplexing and diversification gain. As test results show [9-11], owing to the method high transmission speed may be obtained with any configuration of antenna systems on both sides of the radio connection with simultaneous code gain. Solutions known for the MIMO transmission, such as the VBLAST [5] algorithm or ZF, may be applied for receiving [4-7].

The LD-STBC-VBLAST method was used by the authors for OFDM transmission in a WLAN system. Simulation results for selected receive algorithms that may be used for WLAN 802.11n MIMO/OFDM system are presented. Performance, in the terms of BER, LD-STBC-VBLAST and VBLAST receivers has been compared. The analyzed system uses a multi-stream transmission in which a part of spatial streams is STBC-coded and a part is transmitted without coding. It was assumed that individual subcarriers are modulated with 2-PSK, 4-PSK or 16-QAM signal. The purpose hereof is to compare the operation of the aforesaid system for two different receivers: LD-STBC-VBLAST, using the LD (Linear Dispersion) algorithm [9] and VBLAST [5] and to check the suitability of the abovementioned receivers for the improvement of data transmission quality in WLAN 802.11n.

The BER and PER were determined for the E type transmission channel model [12]. The simulation referred to transmission through E type WLAN channel because, as test

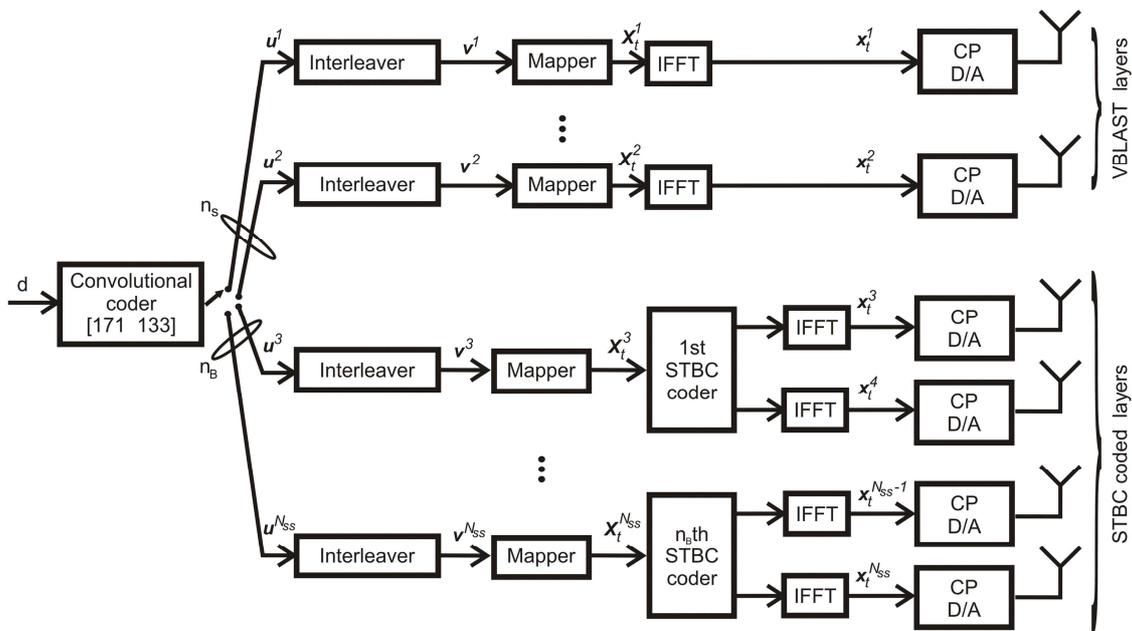


Figure 1. Scheme of the transmit part of analyzed system.

results shown by Kotrys et al. in e.g. [4], the lowest error rate has been obtained in a MIMO transmission using the channel. It was also assumed that the Channel State Information (CSI) is known in the receiver.

This paper is organized as follows: Section 2 describes the simulation model. Section 3 presents the reception algorithms used in researches. Section 4 contains simulation results that have been carried out, and, finally, Section 5 includes a summary and conclusions.

## II. SYSTEM MODEL

In order to assess the quality of operation of the LD-STBC-VBLAST receiver in WLAN, many simulation experiments have been made. We used to the simulation the MATLAB environment. The model of the simulated system enables BER and PER determination. A block diagram of the transmitting part of the simulated system is presented in Fig. 1.

In the transmitter, the information sequence  $d$  is coded by a convolutional encoder [171 133] with rate  $R=1/2$ , used in the 802.11n standard [1]. The coded  $u$  sequence generated by the encoder is divided into  $N_{ss}$  spatial streams  $u^1, \dots, u^{N_{ss}}$ . Three different variants of MIMO transmissions are possible: a non-coded multistream transmission, an STBC-coded stream transmission, a transmission where a part of streams is non-coded and a part is STBC-coded.

Each of the spatial streams is subject to interleaving in blocks reflecting the successively assigned OFDM symbols as per the 802.11n recommendation [1]. Depending on the valence of the applied modulation, the bits of the interleaved sequence  $v$  are adequately grouped and mapped into the elements of 2-PSK, 4-PSK or 16-QAM constellations. Signals  $X_i(k)$  represent the signals transmitted on the  $k$ -subcarrier of the OFDM symbol. Signals that modulate subcarriers within the  $t$ -symbol OFDM form a vector of  $X_t$  signals. Samples of the OFDM symbol in time domain are formed using the Inverse Fast Fourier Transform (IFFT)

algorithm. They make up the  $x_t$  vector. Then, samples of the OFDM symbol are supplemented with a Cyclic Prefix (CP) and transformed from Digital to Analogue (D/A).

To adhere to the 802.11n standard [1], in the tested system, each OFDM symbol uses the 52 sub-carriers to transmit data, 4 subcarriers are used to transmit the so-called pilot signals. OFDM is performed with the use of the 64-point Fourier transform. The duration time of a single OFDM symbol is  $4\mu s$  with the sampling frequency of 20MHz. To avoid the intersymbol interference, the  $0.8\mu s$  cyclic prefix is added. The transmission throughput of the analyzed system depends on the number of spatial streams that were used and valence of modulation applied to each subcarrier of the OFDM signal. A specification of the analyzed system variants is shown in Table I.

TABLE I. ANALYZED SYSTEM VARIANTS

Modulation	Number of spatial streams	Throughput [Mb/s]
2-PSK	2	13
2-PSK	3	19,5
4-PSK	2	26
4-PSK	3	39
16-QAM	2	52
16-QAM	3	78

## III. RECEPTION ALGORITHMS

In the analyzed system, data transmission is performed using two, three or four spatial streams. Correct synchronization and estimation of the channel state in the receiver was assumed. A total of signals transmitted by all transmit antennae (modified as a result of channel passing) reaches each receive antenna. The signal from receive antenna after sampling is transformed, with the Fast Fourier

Transform (FFT), from time domain to frequency domain and subsequently demodulated. Two receiving methods have been analyzed in the paper, VBLAST and LD-STBC-VBLAST receivers.

The VBLAST receive algorithm consists in iterative reduction of the intersymbol interference between signals transmitted by different transmit antennas and may be illustrated as follows [5]:

*Initialization:*

$$\mathbf{W}_1 = \mathbf{H}^+ \quad (1)$$

$$i = 1 \quad (2)$$

*Successive iterations:*

$$k_i = \arg \min_{j \in \{k_1, \dots, k_{t-1}\}} \|(\mathbf{W}_i)_j\| \quad (3)$$

$$y_{k_i} = (\mathbf{W}_i)_{k_i} \mathbf{r}_i \quad (4)$$

$$\hat{a}_{k_i} = Q(y_{k_i}) \quad (5)$$

$$r_{i+1} = r_i - \hat{a}_{k_i} (\mathbf{H})_{k_i} \quad (6)$$

$$\mathbf{W}_{i+1} = \mathbf{H}_{k_i}^+ \quad (7)$$

$$i = i + 1, \quad (8)$$

where:  $\mathbf{r}$  is received signal vector,  $\mathbf{H}^+$  is matrix of Moore-Penrose pseudo-inversion of the channel matrix  $\mathbf{H}$  [1],  $(\mathbf{W}_i)_j$  is its  $j$ -row of matrix  $\mathbf{W}_i$ ,  $Q(\cdot)$  is the function of decision that selects the closest, in terms of Euclid's distance, point from the constellation of signals modulating individual subcarriers,  $(\mathbf{H})_{k_i}$  is the  $k_i$ -column of matrix  $\mathbf{H}$ ,  $\mathbf{H}_{k_i}$  is the matrix obtained through clearing columns  $k_1, \dots, k_i$  of matrix  $\mathbf{H}$  [4][5].

The receive method based on the LD-STBC-VBLAST algorithm has been adopted by Longoria-Gandara et al. [10] to OFDM WLAN transmission. The method is applied in the case where in the MIMO system non-coded streams are transmitted by selected antennae and simultaneously STBC coded streams are transmitted by other antennae.

The transmit part of the LD-STBC-VBLAST system is presented in Fig. 1. The basic idea of the systems is concurrent transmission of spatial streams both non-coded and space-time block coded. Then, it was assumed to denote the system having  $n_S$  of non-coded spatial streams and  $n_B$  of STBC coded spatial streams as  $(n_S, n_B)$ -LD-STBC-VBLAST, for example the description (0,2) denotes the system which use two STBC encoded streams.

In the LD-STBC-VBLAST receiver, the theory of linear dispersion described by Hassibi and Hochwald [9] was used

to demodulate. Therefore, during modulation signals received from both non-coded and STBC coded streams may be treated the same.

Further on, the following designations have been assumed:

$n_A$  – number of antennae in a single STBC stream;

$N_T$  – number of transmit antennae;

$N_R$  – number of receive antennae;

$n_B$  – number of STBC coded streams;

$n_S$  – number of streams not coded with STBC

Table II below shows which signals are transmitted by two individual antennae in subsequent time intervals [11]. This constitutes a description of the time and space coding performed for a given antenna configuration [9].

TABLE II. TRANSMITTED SIGNALS

Non-coded streams		STBC coded streams STBC block: B=1,...,n <sub>B</sub>	
Time	Antenna i=1,...,n <sub>S</sub>	Antenna 1	Antenna 2
T	$S_{i,j}$	$S_{n_A,1}$	$S_{n_A,2}$
t+T	$S_{i,2}$	$-S_{n_A,2}^*$	$S_{n_A,1}^*$

The signal reaching the receiver is presented as follows [10]:

$$\begin{bmatrix} y_1^{(1)} & y_1^{(2)} \\ y_2^{(1)} & y_2^{(2)} \\ \vdots & \vdots \\ y_{N_R}^{(1)} & y_{N_R}^{(2)} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1N_T} \\ h_{21} & h_{22} & \dots & h_{2N_T} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1} & h_{N_R2} & \dots & h_{N_R N_T} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{nC} \\ \mathbf{S}_C \end{bmatrix} + \begin{bmatrix} n_1^{(1)} & n_1^{(2)} \\ n_2^{(1)} & n_2^{(2)} \\ \vdots & \vdots \\ n_{N_R}^{(1)} & n_{N_R}^{(2)} \end{bmatrix} \quad (9)$$

In the above, as well as in the formulas that follow, the below notation has been applied:

- the subscripts signify numbers of relevant antennae;
- the superscripts signify the number of modulation interspace in a given time interval

The transmitted signal is specified as  $\begin{bmatrix} \mathbf{S}_{nC} \\ \mathbf{S}_C \end{bmatrix}$ . It is composed of two separate matrixes of which each describes symbols transmitted in relevant streams: nC – non-coded and C – STBC-coded.

$$\mathbf{S}_{nC} = \begin{bmatrix} S_1^{(1)} & S_1^{(2)} \\ S_2^{(1)} & S_2^{(2)} \\ \vdots & \vdots \\ S_{n_S}^{(1)} & S_{n_S}^{(2)} \end{bmatrix} \quad (10)$$

$$\mathbf{S}_C = \begin{bmatrix} \mathbf{S}_1^C \\ \vdots \\ \mathbf{S}_{n_B}^C \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1^{(1)C} & \mathbf{S}_1^{(2)C} \\ \vdots & \vdots \\ \mathbf{S}_{n_B}^{(1)C} & \mathbf{S}_{n_B}^{(2)C} \end{bmatrix} \quad (11)$$

And where each matrix element (13) is given as [10]:

$$\begin{bmatrix} \mathbf{S}_B^{(1)C} & \mathbf{S}_B^{(2)C} \end{bmatrix} = \begin{bmatrix} s_{n_A,1} & -s_{n_A,2}^* \\ s_{n_A,2} & s_{n_A,1}^* \end{bmatrix}, \quad (12)$$

where  $B=1, \dots, n_B$ .

By applying the LD theory [9] equation (11) may be noted as follows [10]:

$$\begin{bmatrix} y_1^{(1)} \\ y_1^{(2)*} \\ \vdots \\ y_{N_R}^{(1)} \\ y_{N_R}^{(2)*} \end{bmatrix} = [\mathbf{H}_{nc} \quad \mathbf{H}_c] \mathbf{S}_{LD} + \begin{bmatrix} n_1^{(1)} \\ n_1^{(2)*} \\ \vdots \\ n_{N_R}^{(1)} \\ n_{N_R}^{(2)*} \end{bmatrix} \quad (13)$$

In the matrix notation, equation (15) may be noted like this:

$$\mathbf{Y}_{LD} = \mathbf{H}_{LD} \mathbf{S}_{LD} + \mathbf{N}_{LD}, \quad (14)$$

where all matrixes are called LD matrixes.

The matrix of the transmitted signal may also have the form of the LD matrix:

$$\mathbf{S}_{LD} = \begin{bmatrix} \mathbf{S}_{LD}^{nC} \\ \mathbf{S}_{LD}^C \end{bmatrix}, \quad (15)$$

In the receiver, similarly to [10], the so-called QR decomposition of the channel matrix  $\mathbf{H}$  is used.

This decomposition consists in splitting the channel matrix into two matrixes whose product equals the channel matrix:

$$\mathbf{H}_{LD} = \mathbf{Q}_{LD} \mathbf{R}_{LD} \quad (16)$$

Matrix  $\mathbf{Q}$  is a rectangular matrix  $2N_R \times n_{SYM}$ . Whereas matrix  $\mathbf{R}$  is a square, upper triangular matrix  $n_{SYM} \times n_{SYM}$ , where  $n_{SYM}=2(n_S+n_B)$ . A detailed description of the QR decomposition algorithm may be found in [6].

After determining the  $\mathbf{Q}$  and  $\mathbf{R}$  matrix for the  $\mathbf{H}$  channel matrix, linear detection of the received signal takes place in the receiver. The detection algorithm [7] is as follows:

*Initialization:*

$$k = N_T \quad (17)$$

$$\mathbf{w} = \mathbf{Q}^H \mathbf{n} \quad (18)$$

$$\tilde{\mathbf{s}} = \mathbf{Q}^H \mathbf{y} = \mathbf{R} \mathbf{s} + \mathbf{w} \quad (19)$$

*Successive iterations (1):*

$$\hat{s}(k) = Q \left[ \tilde{s}(k, k) / R(k, k) \right] \quad (20)$$

$$k = k - 1 \quad (21)$$

*If k equals 0 – end of algorithm operation*

$$i = k + 1 \quad (22)$$

$$interf(k) = 0 \quad (23)$$

*Successive iterations (2):*

$$interf(k) = interf(k) + R(k, i) \cdot \hat{s}(i) \quad (24)$$

$$i = i + 1 \quad (25)$$

*End of loop (2)*

$$\tilde{\mathbf{s}} = \mathbf{R} \mathbf{s} + \mathbf{w} - \mathbf{interf} \quad (26)$$

or

$$\tilde{\mathbf{s}} = \mathbf{Q}^H \mathbf{y} - \mathbf{interf} \quad (27)$$

*End of loop (1).*

where,  $\mathbf{y}$  is received signal vector,  $\mathbf{s}$  is transmit signal vector,  $\tilde{\mathbf{s}}$  decision statistic for transmit signal,  $\hat{\mathbf{s}}$  estimate for transmit signal,  $\mathbf{Q}^H$  the hermitian transpose of  $\mathbf{Q}$ ,  $\mathbf{n}$  represents the white gaussian noise of variance  $\sigma_n^2$  observed at the  $N_R$  receive antennae while the average transmit power of each antenna is normalized to one.

The presented detection algorithm is based on successive interference reduction. The decisions on transmitted signals  $\tilde{\mathbf{s}}$  are determined allowing for the calculated information on interfering signals (*interf*) coming from other transmit antennae.

#### IV. SIMULATION RESULTS

By means of the computer simulation, we have determined the BER and PER depending on the SNR value. An assumption has been made that transmission takes place in E type WLAN channel [12]. A comparison of the quality of MIMO systems operation using the following two types of receivers has been presented: LD-STBC-VBLAST and VBLAST depending on the number of spatial streams and selected modulations: 2-PSK, 4-PSK, 16-QAM. In the simulations, ideal synchronization has been assumed as well

as that the receiver knows the CSI. The transmitted packets were 1000-byte long. OFDM technique has been applied. The 64-point IFFT/FFT has been implemented, where data is transmitted on 52 subcarriers. Additionally, four subcarriers have been used to transmit pilot signals and 8 subcarriers constituted a protection interval. To assess correctness of operation of the proposed simulation model, a series of tests confirming the results taken from literature [4][5][7] have been performed. Different combinations of parameter setups for the investigated MIMO systems have been simulated. The most representative results have been selected for the presentation.

In Figures 2 and 3, PER and BER curves are illustrated for the VBLAST and LD-STBC-VBLAST receive systems including two and three spatial streams for different number of transmit  $N_t$  and receive  $N_r$  antennae. Transmission in these systems takes place at the speed of 52 and 78Mb/s respectively with the WLAN channel type E [12].

Considering the transmission with two spatial streams (Fig. 2), with PER at  $10^{-3}$ , (0,2) LD-STBC-VBLAST system proved the best properties. Here, transmission takes place using four transmit and receive antennae. The (0,2) LD-STBC-VBLAST system offers 1% PER with about 17 dB. The (1,1) LD-STBC-VBLAST system including three transmit and receive antennae is by approximately 6 dB inferior. The system employing VBLAST receiver, where the number of antennae equals the number of spatial streams for the same level of PER (at  $10^{-3}$ ), is inferior to (1,1) LD-STBC-VBLAST system by 0.2 dB. The (0,2) LD-STBC-VBLAST system offers approximately 7 dB gain comparing to the (1,1) LD-STBC-VBLAST and VBLAST systems with the BER of about  $10^{-4}$ , respectively. The throughput for these systems is equal 52Mb/s.

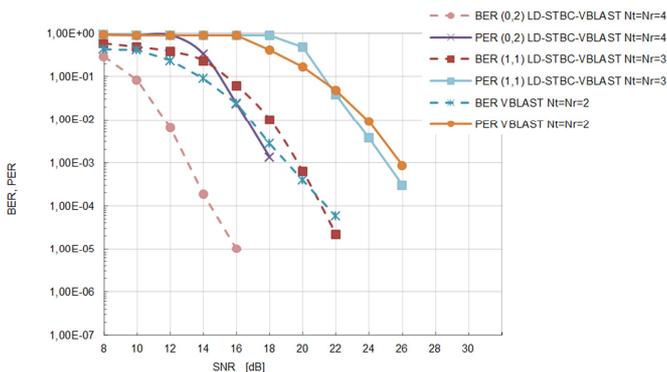


Figure 2. PER and BER for two-spatial-stream-systems, modulation 16QAM, 52Mb/s.

Fig. 3 represents the simulation results for systems with three spatial streams and obtained throughput 78Mb/s. For 1% PER the best results have been noted in the case of (1,2) LD-STBC-VBLAST system including five transmit and receive antennae. This level is obtained when SNR equals 19 dB. The (2,1) LD-STBC-VBLAST systems is inferior by

3.5 dB. It has four transmit and receive antennae. The number of antennae in the system employing VBLAST receiver equals the number of spatial streams and is equal 3. This has proven to perform (PER at  $10^{-4}$ ) poorer than the best presented (1,2) LD-STBC-VBLAST system by 6 dB. The (1,2) LD-STBC-VBLAST system offers approximately 3 dB gain comparing to the (2,1) LD-STBC-VBLAST and BLAST systems with the BER of about  $10^{-4}$ , respectively.

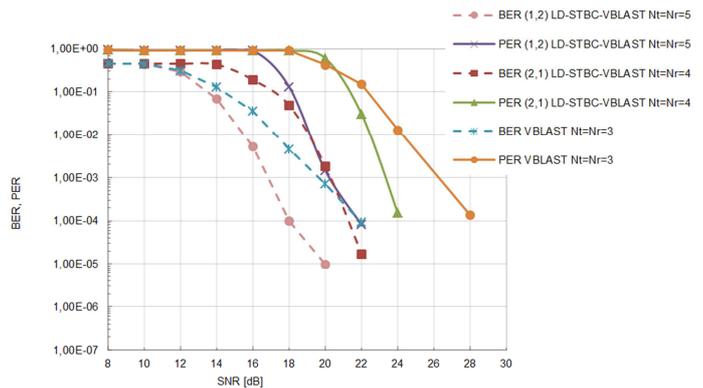


Figure 3. PER and BER for three-spatial-stream-systems, modulation 16QAM, 78Mb/s.

A system that uses LD-STBC-VBLAST receive enables enhancement of transmission speed with coincident quality improvement through application of an additional spatial stream. To improve quality, an STBC coding on additional spatial stream must be used. If the number of spatial streams grows from one to two, a 100% increment of speed is obtained with simultaneous minor improvement in PER for 2-PSK, 4-PSK and 16-QAM modulation. If the number of spatial streams grows from two to three, the speed increment is 50% with 1% improvement in PER by 3 dB for 2-PSK modulation and by 4 dB for modulations 4-PSK and 16-QAM.

## V. CONCLUSION AND FUTURE WORK

This paper presented a proposal of use LD-STBC-VBLAST reception for WLAN systems with a hybrid transmission. A multistream transmission was suggested where a part of spatial streams was STBC coded and a part was transmitted without any codes. The impact of transmit diversification on the quality of transmission has been analyzed. Based on the simulation results it may be clearly observed that the transmit diversification offers better properties of the transmission system. The LD-STBC-VBLAST receiver proves the best results in BER and PER when compared to the system with a VBLAST receiver at the cost of increased number of antennae.

As it results from the performed tests, the method that has been applied (LD-STBC-VBLAST) allows increasing of the transmission speed with no deterioration of the error rate through suitable selection of the transmitted spatial streams.

The system with LD-STBC-VBLAST receiver allows iterative reduction of interference. Therefore, the BER and PER results are considerably better than in the case of the VBLAST receiver system. Given the presented simulation results, we can suppose that the investigated LD-STBC-VBLAST receiver could be successfully used in next wireless networks which are currently being developed.

As a future research task, another reception methods should be examined and complexity evaluation of investigated reception algorithms should be perform.

#### ACKNOWLEDGMENT

This work was partially supported by the Polish National Science Centre under research grant 2011/01/B/ST7/06578.

#### REFERENCES

- [1] 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 5: Enhancements for higher throughput, IEEE Std 802.11n-2009 (Amendment to IEEE Std 802.11-2007 as amended by IEEE Std 802.11k-2008, IEEE Std 802.11r-2008, IEEE Std 802.11y-2008, and IEEE Std 802.11w-2009), pp. c1 –502, 29 2009.
- [2] G. J. Foschini and M. J. Gam, “Layered space-time architecture for wireless communication in a fading environment when using multiple antennae,” *Bell Labs. Syst. Tech. Journal*, vol. 1, 1996, pp. 41-59.
- [3] H. Bölcskei and A. J. Paulraj, “Multiple-input multiple-output (MIMO) wireless systems,” *The Communications Handbook*, CRC Press, 2nd ed., 2002, pp. 90.1-90.14.
- [4] R. Kotrys, M. Krasicki, P. Remlein, and P. Szulakiewicz, “Receiver Algorithms for Multi-stream Data Transmission in WLAN 802.11n Networks,” *Wireless Personal Communications*, vol. 68, 2013, pp. 1583-1594, doi: 10.1007/s11277-012-0540-7
- [5] P.W. Wolniansky, G.J. Faschini, G.D. Golden, and R.A. Valenzuela, “V-BLAST: An architecture for realizing very high data-rates over the wireless channel,” *Proc. IEEE ISSSE-98*, Pisa, Italy, 1998, pp. 295-300.
- [6] D. Wübben, R. Böhnke, J. Rinas, V. Kühn, and K.D. Kammeyer, “Efficient algorithm for decoding layered space-time codes,” *IEE Electronic Letters*, vol. 37, 2001, pp. 1348-1350.
- [7] D. Wübben, R. Böhnke, V. Kühn, and K.D. Kammeyer, “Reduced Complexity MMSE Detection for BLAST Architectures,” *Proc. IEEE GLOBECOM Conf.*, 2003, pp. 2258-2262.
- [8] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, “Space-Time Block Codes from Orthogonal Designs,” *IEEE Transactions on Information Theory*, vol. 45, no. 5, 1999, pp. 1456-1467.
- [9] B. Hassibi and B.M. Hochwald, “High rate codes that are linear in space and time,” *IEEE Transactions of Information Theory*, vol. 48, no. 7, 2002, pp. 1804-1824.
- [10] O. Longoria-Gandara, A. Sanchez-Hernandez, J. Cortez, M. Bazdresch, and R. Parra-Michel, “Linear Dispersion Codes Generation from Hybrid STBC-VBLAST Architectures,” *Proc. of 4th International Conference on Electrical and Electronics Engineering*, 2007, pp. 142-145.
- [11] M. Tianyu and M. Motani, “STBC-VBLAST for MIMO wireless communication systems,” *IEEE International Conference on Communications, ICC*, vol. 4, 2005, pp. 2266- 2270.
- [12] IEEE P802.11 Wireless LANs, TGn Channel Models, IEEE 802.11-03/940r4, May 2004.

# A Wideband Envelope Modulator Design for Envelope-Tracking SiGe Power Amplifier (ET-PA) for Broadband Wireless Applications

Yan Li<sup>+</sup>, Jerry Lopez<sup>\*+</sup>, and Donald Y.C. Lie<sup>+</sup>

RF Micro Devices Inc., Phoenix, AZ, USA

<sup>\*</sup>NoiseFigure Research Inc., Lubbock, TX, USA

<sup>+</sup>Department of Electrical and Computer Engineering

Texas Tech University, Lubbock, TX, USA

e-mail: donald.lie@ttu.edu

**Abstract** — This work presents a design example of a wideband envelope modulator (EM) that can effectively modulate the instantaneous supply voltage for a monolithic SiGe power amplifier to form a highly efficient envelope-tracking power amplifier (ET-PA) system for potential broadband wireless applications. Trade-offs between linearity, switching noise, efficiency and bandwidth have been examined carefully with both experimental data and RF/analog/digital co-simulations. The SiGe ET-PA using our wideband EM was characterized by the WiMAX 64QAM 8.75 MHz signal, showing an overall PAE of 30.5% at output power of 17 dBm with an error vector magnitude (EVM) of 4.4%, while successfully passing the stringent WiMAX spectral mask.

**Keywords:** Broadband wireless; envelope modulator; envelope-tracking (ET); LTE; SiGe; power amplifier (PA); WiMAX

## I. INTRODUCTION

Recently, envelope-tracking (ET) has become a very popular efficiency enhancement technique for RF power amplifiers (PAs) design [1]-[5]. Split-band design concept has been applied to the envelope modulator (EM) for wideband applications [1] [4] [5]. However, compared with narrowband cellular applications (e.g., EDGE), the high peak-to-average power ratio (PAPR or PAR) of the broadband signals (e.g., LTE/WiMAX) require the EM to have wider bandwidth and lower distortion, while still maintaining excellent efficiency. Therefore, careful investigations of the design trade-offs for the wideband EM are still required. In this paper, we will report a design example for a wideband EM paired with a SiGe monolithic PA to form an effective ET-PA system for potential broadband wireless applications. A discrete linear-assisted switching EM is designed to investigate the overall ET-PA system performances by experimental data and RF/analog/digital co-simulation. The effects of bandwidth and switching frequency as two major factors on linearity and efficiency will be studied carefully to make this EM applicable for high PAR wideband applications.

This paper is organized as follows. Section II presents the designs of the common-emitter SiGe PA and linear-assisted switching EM. In Section III, we demonstrate the effective of our designed EM by applying it to the SiGe PA to form an ET-PA system. A WiMAX 64QAM 8.75 MHz signal (PAR of 10.5 dB) will be applied to the ET-PA to

showcase the efficiency and linearity enhancements of the ET technique over traditional fixed-supply PAs.

## II. CIRCUITS DESIGN

### A. Design of Common-Emitter SiGe Power Amplifier

A monolithic 1-stage common-emitter SiGe PA is used here as an example to form an ET-PA system to study the trade-offs for wideband EM design. This PA was designed and fabricated in IBM 7HP 0.18 $\mu$ m SiGe BiCMOS technology [1]-[3]. The simplified schematic and die picture of the PA are shown in Fig. 1. The high-breakdown heterojunction bipolar transistor (HBT) option is used for the PA design with a total emitter-area of 220  $\mu$ m<sup>2</sup> (typical  $BV_{CEO}=4.2$  V;  $BV_{CBO}=12.5$  V). This monolithic SiGe PA was tested on a FR4 PCB. The RF Choke (RFC) inductor was left off-chip to achieve high Q at 2.4 GHz for better power-added-efficiency (PAE). A high-Q bondwire is used as the output tank inductor design, together with more than 4 downbonds (i.e., bondwires at the emitter node) to reduce the ground parasitic inductance for high PAE [3]. No other off-chip elements are needed nor used for the PA input and output matching.

It is important to characterize the PA thoroughly before designing the EM for optimal ET-PA performances, as the collector impedance presented by the PA ( $R_{load}$ ) will affect the efficiency and linearity performance of the EM. Fig. 2 shows the measured PAE vs. output power ( $P_{out}$ ) at different supply voltage  $V_{CC}$  in the continuous wave (CW) mode. For the fixed-supply PA, its PAE reduces rapidly when  $P_{out}$  drops, but the PAE at low  $P_{out}$  can be greatly enhanced by varying  $V_{CC}$  as shown by the dash curve, which shows the idea for an ET-PA operation. The dash curve is obtained at each peak PAE point of different  $V_{CC}$  levels. This characteristic of PAE enhancement indicates that the ET technique can improve the average efficiency of the PA compared with the case of a fixed-supply PA. Fig. 2 also plots the collector impedance presented by the PA to the EM (i.e.,  $R_{load}$ ), which is calculated from the DC supply voltage and the measured DC supply current of the PA at each peak PAE point (please refer to Fig. 4 for the definition of  $R_{load}$  [4]). The  $R_{load}$  presented by the PA varies, dependent on the operating regions of the PA [1]. According to the measurement data shown in Fig. 2,  $R_{load}$  can change roughly from 70  $\Omega$  to 10  $\Omega$  when  $P_{out}$  increases from 8 dBm to 20 dBm.

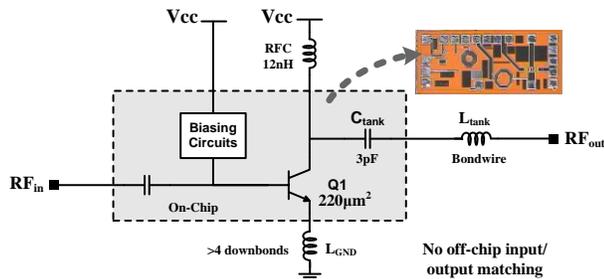


Fig. 1 Simplified schematic and die picture of the 1-stage PA designed and fabricated in IBM 7HP 0.18  $\mu\text{m}$  SiGe BiCMOS technology

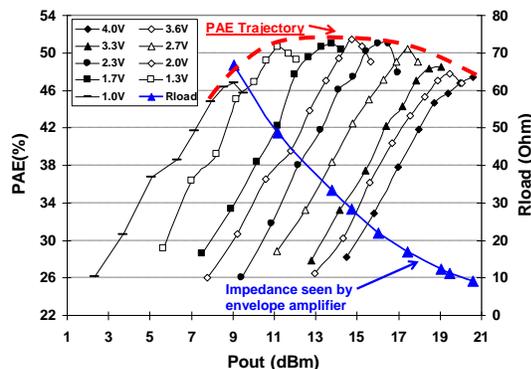


Fig. 2 Measured PAE vs.  $P_{\text{out}}$  of the SiGe PA, and the impedance seen by the EM by changing the PA supply voltage ( $V_{\text{CC}}$ ) biasing in the CW mode

### B. Linear-Assisted Switching Envelope Modulator

A proper EM design is critical to achieve the best overall efficiency and linearity performances for an ET-PA. As reported in [1]-[4], the finite bandwidth and the associated group delay of the EM are large contributors of nonlinearity in an ET-PA. In addition, to take advantage of the efficiency enhancement provided by the ET technique, the EM needs to maintain high efficiency throughout the ET-PA operation. The overall power efficiency of an ET-PA system is the product of the EM efficiency and the PA collector efficiency (CE), which is expressed as:

$$\eta_{\text{ET-PA}} = \eta_{\text{Env.-Mod.}} \cdot \eta_{\text{PA,CE}} \quad (1)$$

Therefore, the EM design targets are high efficiency and wide bandwidth to track the instantaneous input envelope.

#### 1) Split-Band Design of the Envelope Modulator

The envelope signal is extracted from the modulated I/Q (i.e., in-phase/quad-phase) signals from the LTE/WiMAX baseband and then feed into the EM. Such nonlinear transformation will expand the bandwidth of the envelope by a factor of 5-10 compared with the original signal bandwidth [4] [5]. Conventionally, the EM can be implemented in the form of a linear regulator (e.g., a low dropout regulator (LDO) as in [6]), as the linear topology offers wide bandwidth and can be with almost no output ripple. Nonetheless, the power efficiency of linear regulator is very poor when the output voltage level is low [6], making it unsuitable for high PAR signals for 3G/4G

applications. On the other hand, a switching regulator has high power efficiency across a broad range of output voltage, but it produces significant output ripples and its bandwidth is constrained to be a fraction of the switching frequency [7], making it suitable only for narrowband applications such as the North American Digital Cellular (NADC) in [7]. Switching regulators can also be applied to high data-rate systems when a rather high switching frequency is employed [8], but the high switching frequency inevitably causes high switching loss that limits the power efficiency (e.g.,  $\sim 76\%$  maximum for WCDMA in [8]) and also can degrade ET-PA linearity considerably, which can often defeat the purpose of using switching regulators.

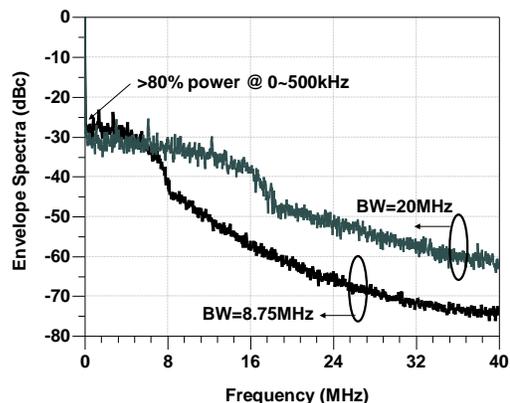


Fig. 3 Simulated envelope spectra of WiMAX 8.75/20 MHz signals.

Many recent reports on the wideband EM design for ET-PA have combined the advantages of a wideband linear regulator and a high efficiency switching regulator in various ways [9]-[11]. Fig. 3 shows the simulated envelope spectra of WiMAX 8.75 MHz and 20 MHz signals (PAR of  $\sim 10.2$  dB). An important characteristic of the envelope spectrum is that  $\sim 80\%$  of envelope power resides from DC to several kHz, while over 99% of the envelope power resides within DC to 8MHz for the 8.75 MHz signal, and within DC to 20 MHz for the 20 MHz signal, respectively. Such a characteristic of the envelope spectrum implies that a “split-band” EM design (or often called “linear-assisted switching” structure) can achieve a high efficiency over a wide bandwidth [4]. The split-band EM consists of a wideband linear stage (therefore of low efficiency) and a high efficiency narrowband switching stage. This split-band design lessens the requirements of the switching stage, since the fast transients of the envelope signal will be taken care of by the wideband linear stage, while the switching stage will handle DC and the slow moving signals with high efficiency. The efficiency of the entire EM ( $\eta_{\text{env.-mod.}}$ ) is a combination of the switching stage efficiency ( $\eta_{\text{SW}}$ ) and the linear stage efficiency ( $\eta_{\text{lin}}$ ), as expressed by:

$$\frac{1}{\eta_{\text{env.-mod.}}} = \frac{\alpha}{\eta_{\text{SW}}} + \frac{1-\alpha}{\eta_{\text{lin}}} \quad (2)$$

, where  $\alpha$  is the ratio of the output power from the switching stage to the total output power of the EM [4] [12].

## 2) Discrete Envelope Modulator Design

The linear-assisted switching EM is designed by using commercial-of-the-shelf (COTS) components to investigate the overall efficiency and linearity trade-off in an ET-PA system. Fig. 4 shows the circuit implementation of the discrete EM using an operational amplifier (Op-Amp) as the linear stage and a buck converter as the switching stage. The buck converter supplies the slow slew-rate load current ( $I_{SW}$ ) that contributes to the majority of the load current ( $I_{load}$ ) to ensure high efficiency, while the wideband linear Op-Amp stage operates in a feedback mode to track the high slew-rate current ( $I_{in}$ ). Additionally, the ripples caused by the buck converter will be attenuated and/or filtered by the linear Op-Amp. The smooth transition between the switching stage and the linear stage is realized by a hysteretic current feedback control. The hysteretic current feedback control consists of a current sensing resistor  $R_{sense}$  that senses the output current of the linear stage and a hysteresis comparator to control the buck converter. The value of the sensing resistor  $R_{sense}$  is chosen to be  $1\ \Omega$  in this case, as it needs to be much smaller than  $R_{load}$  (i.e., the load impedance presented by the PA) to achieve high efficiency.

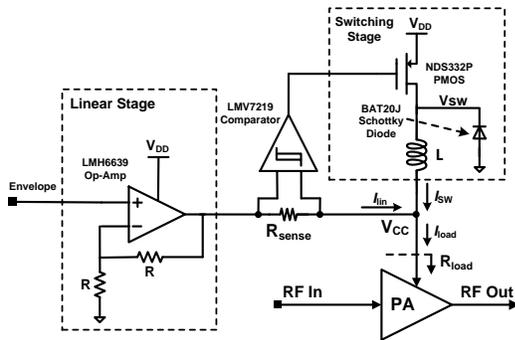


Fig. 4 Schematic of the linear-assisted switching envelope modulator (EM) designed by using COTS components

### C. Efficiency and Nonlinearities of Envelope Modulator

Although many reported on the efficiency of the EM design in the literature [1] [2] [4] [13], its linearity and switching noise trade-offs vs. efficiency have not been studied as rigorously, which are especially critical for wideband signals. In this section, the nonlinearities of the discrete linear-assisted switching EM will be characterized. The switching noise and the bandwidth of the EM are two major factors that cause distortions to the output envelope signal. Understanding the effects of the switching frequency and bandwidth limitation of the EM helps to optimize both efficiency and linearity of the overall ET-PA.

#### 1) Bandwidth of the Envelope Modulator

Previous works suggest that for a good linearity performance, the linear stage (i.e., the Op-Amp) should have sufficient bandwidth to track the high frequency contents of the envelope signal with high fidelity [1] [4] [5]. In addition, once the linear Op-Amp stage is used to assist the switching buck converter, it should have sufficient

bandwidth to suppress the switching ripples/noise. The switching ripples beyond the bandwidth of the linear stage can distort the envelope signal, and be mixed with the modulated carrier in the PA to cause large spurious noise at the PA output, potentially degrading the system linearity.

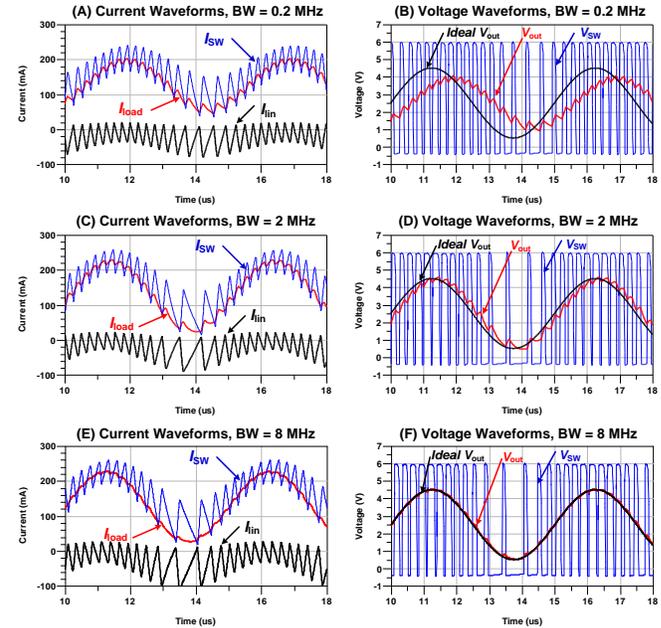


Fig. 5 Simulated (A, C, E) current and (B, D, F) voltage waveforms of the EM; the behavior model is used for the Op-Amp with different 1-dB bandwidths, while realistic SPICE models are used for other blocks of the EM. Input voltage =  $1.25 + \sin(2\pi \cdot 200\text{kHz} \cdot t)$  V,  $L = 4.7\ \mu\text{H}$ ,  $R_{load} = 20\ \Omega$ .

To investigate the effect of the EM bandwidth, the commercial available Op-Amp (i.e., LMH6639 in Fig 4) is replaced with an Op-Amp behavior model provided by Agilent's ADS, such that the bandwidth of the Op-Amp can be changed manually in the SPICE simulations. The realistic SPICE models are still used for other blocks of the EM. Fig. 5 shows the simulated current and voltage waveforms of the EM using different 1-dB bandwidths of the Op-Amp at an input wave of  $1.25 + \sin(2\pi \cdot 200\text{kHz} \cdot t)$  V. Here, the 1-dB bandwidth is defined as the frequency where the gain response decreases by 1 dB. As shown in Fig. 5(A), (C) and (E), the output current of the switching stage ( $I_{SW}$ ) has large ripples on the waveforms, which need to be suppressed or cancelled by the output current of the linear stage ( $I_{in}$ ) to reproduce an accurate load current waveform ( $I_{load}$ ). When the 1-dB bandwidth of the Op-Amp is set as 0.2 MHz, the output voltage ( $V_{out}$ ) of the EM exhibits not only the switching ripples but also with some attenuation (Fig. 5(B)). When the 1-dB bandwidth of the Op-Amp is set as 2 MHz, the output voltage waveform  $V_{out}$  can follow the input voltage without attenuation, but the switching ripples still cannot be suppressed (Fig. 5(D)). When the 1-dB bandwidth of the Op-Amp is set as 8 MHz,  $V_{out}$  can now follow the input voltage with high fidelity and at a low noise level (Fig. 5(F)).

To further demonstrate the importance of having a wideband linear stage in the EM to meet the stringent

linearity specs, the entire ET-PA using the monolithic SiGe PA is simulated with the RF/analog/digital co-simulation bench. The behavior model is used for the Op-Amp, while the realistic SPICE models are used for the PA and the other blocks of the EM. The inductor ( $L$ ) of the buck converter is first chosen around 40  $\mu\text{H}$ . The effect of the value of  $L$  on the EM design will be discussed in the next section. The simulated output error-vector-magnitude (EVM) values of the ET-PA against different 1-dB bandwidths of the Op-Amp stage are plotted in Fig. 6 for the WiMAX 64QAM 8.75 MHz signal. As shown in Fig. 6, the EVM values of the ET-PA decrease as the 1-dB bandwidth of the Op-Amp increases, and become saturated to  $\sim 1.8\%$  after the 1-dB bandwidth of the Op-Amp becomes larger than 18 MHz. Fig. 7 shows the simulated transmission output spectra of the ET-PA with different bandwidths of the Op-Amp. There is a large improvement on the Adjacent Channel Power Ratio (ACPR) when the 1-dB bandwidth of the Op-Amp increases from 8 MHz to 18 MHz, enabling the output spectrum passing the stringent WiMAX spectral mask specs for the case of 18 MHz. As indicated by Figs. 6-7, the required bandwidth of the EM for the ET-PA needs to be able to respond to the envelope frequency contents to at least 2x of the original instantaneous signal bandwidth.

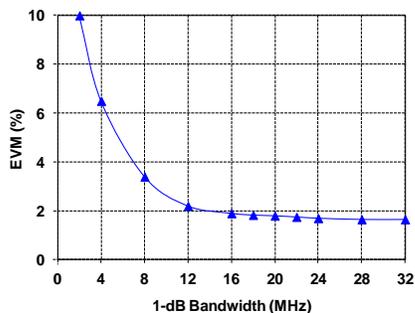


Fig. 6 Simulated output EVM of the ET-PA vs. 1-dB bandwidth of the Op-Amp for the WiMAX 64QAM 8.75 MHz signal. The behavior model is used for the Op-Amp, while realistic SPICE models are used for the SiGe PA and other blocks of the EM  $P_{in} = 6 \text{ dBm}$ ,  $P_{out} = 16 \text{ dBm}$ .

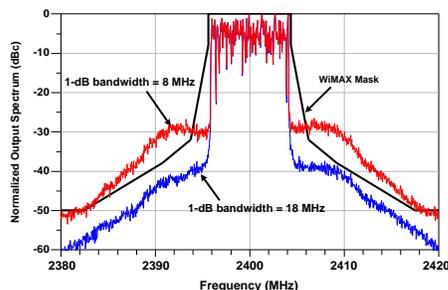


Fig. 7 Simulated output spectra of the ET-PA using different bandwidths of Op-Amp for the WiMAX 64QAM 8.75 MHz signal. The behavior model is used for the Op-Amp, while realistic SPICE models are used for the SiGe PA and other blocks of the envelope modulator.  $P_{in} = 6 \text{ dBm}$ ,  $P_{out} = 16 \text{ dBm}$ .

## 2) Switching Frequency of Envelope Modulator

The average switching frequency of the EM shown in Fig. 4 is well analyzed in [4] and can be expressed as:

$$f_{switch} = \frac{R_{sense}}{L} \cdot \frac{V_{DD}}{2h} D \left( 1 - D \cdot \frac{V_{rms}^2}{V_{dc}^2} \right) = \frac{R_{sense}}{L} \cdot \frac{V_{dc}}{2h} \left( 1 - \frac{V_{rms}^2}{V_{DD} \cdot V_{dc}} \right) \quad (3)$$

, where  $V_{dc}$  and  $V_{rms}$  are the average and root-mean-square voltages of the output envelope signal, respectively;  $h$  is the hysteresis voltage of the comparator, and  $D$  is the duty ratio that can be calculated from  $V_{dc}/V_{DD}$ . In this design, the comparator LMV7219 has a predetermined internal hysteresis  $h$  of 7-10 mV according to the data sheet and the SPICE simulations. Therefore, from (3) the average switching frequency can now be mainly controlled by the value of  $L$ . The drawback of using a small  $L$  is that it usually generates more switching ripples at high frequencies, making the design of the linear stage more challenging [11].

Fig. 8 shows the SPICE simulated waveforms and spectra of the EM designed using two different values of  $L$  with an input waveform of  $1.25 + \sin(2\pi \cdot 500\text{kHz} \cdot t) \text{ V}$ . This time, the realistic SPICE models are used for all blocks of the EM simulations. The switching current  $I_{sw}$  supplies both DC and AC components of the load current ( $I_{load}$ ) by using an  $L$  of 4.7  $\mu\text{H}$ ; a higher switching frequency and large switching ripples on the waveform of  $I_{sw}$  can be observed from Fig. 8(A). Such large switching ripples need to be suppressed or cancelled by the output current of the linear Op-Amp ( $I_{lin}$ ), which can be clearly shown by the spectra of  $I_{sw}$  and  $I_{load}$  in Figs. 8(E) and (G). On the other hand, for the case of  $L = 68 \mu\text{H}$ ,  $I_{sw}$  supplies only the DC component of  $I_{load}$ , while the AC component is taken care of by the linear Op-Amp, as shown in Fig. 8(B). Also, the spectra of  $I_{sw}$  and  $I_{load}$  for the case of  $L = 68 \mu\text{H}$  have smaller harmonics than those using  $L = 4.7 \mu\text{H}$ . These SPICE simulations indicate that the optimal value of  $L$  should be selected according to the best trade-off of maximum efficiency and linearity.

Fig. 9 shows the SPICE simulated efficiency of the EM and the EVM of the ET-PA using the monolithic SiGe PA presented in Fig. 1. The realistic SPICE models are used for the SiGe PA and all blocks of the EM. From the pure view point of efficiency, the optimal value of  $L$  for the best efficiency is 8.2  $\mu\text{H}$ . Smaller  $L$  results in higher switching frequency that can cause significant switching loss and ripples. On the other hand, too large of the  $L$  makes the buck-converter only able to supply the DC component of the load current, and in that case the lower efficiency Op-Amp has to deliver the remaining AC contents (as illustrated in Fig. 8 (B)), leading to lower efficiency for the EM and thus the overall ET-PA system. Rather large  $L$  can also cause high parasitic resistance to decrease its efficiency. In addition, as shown in Fig. 9, the output EVM of the ET-PA is increased (i.e., linearity degraded) as  $L$  decreases. For example, the efficiency of the EM is increased by 4% by reducing  $L$  from 100  $\mu\text{H}$  to 8.2  $\mu\text{H}$ , but at the slight cost of worse EVM from 2.45% to 2.8%.

Fig. 10 shows the SPICE simulated output spectra of the ET-PA using  $L$  of 8.2  $\mu\text{H}$  and  $L$  of 27  $\mu\text{H}$ , respectively. When the larger  $L$  (27  $\mu\text{H}$ ) is chosen, the ACPR is 4-6 dB better at the offset of 5-8 MHz from the center frequency. Please note the output spectrum for the case of  $L = 8.2 \mu\text{H}$  slightly violates the WiMAX spectral mask.

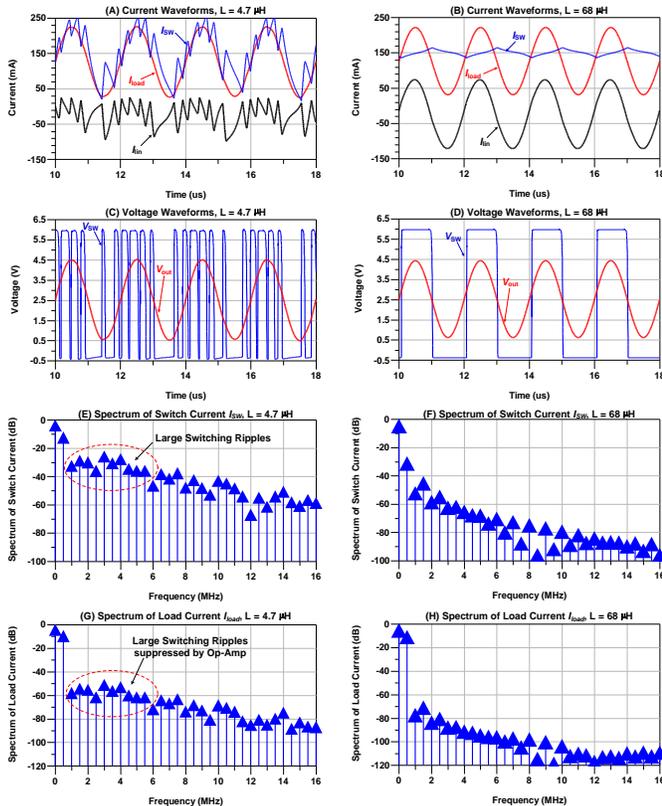


Fig. 8 SPICE simulated (A, B) current waveforms, (C, D) voltage waveforms, (E, F) spectra of the switching current  $I_{SW}$ , and (G, H) spectra of the output load current  $I_{load}$  of the EM using two different values of  $L$  ( $4.7 \mu\text{H}$  vs.  $68 \mu\text{H}$ ). The realistic SPICE models are used for the envelope modulator.  $R_{load} = 22 \Omega$ , the input voltage =  $1.25 + \sin(2\pi \cdot 500\text{kHz} \cdot t)$  V.

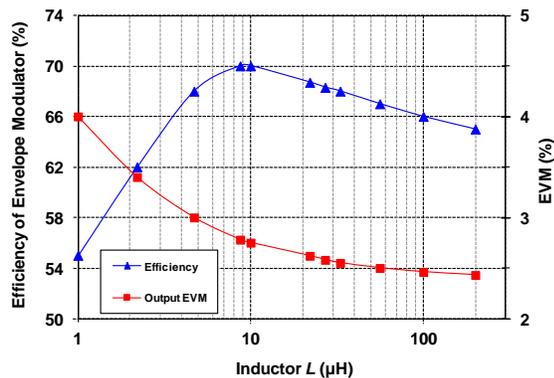


Fig. 9 SPICE simulated efficiency of the EM and output EVM of the ET-PA using different values of  $L$  for WiMAX 64QAM 8.75MHz signal. Realistic SPICE models were used for the PA and EM.  $V_{DD} = 4.2$  V,  $P_{out} = 17$  dBm.

Figs. 9-10 indicate that a small efficiency improvement may not be worthwhile if one has to sacrifice the linearity of overall ET-PA. Therefore, the  $L$  of  $27 \mu\text{H}$  is chosen in the design for our EM to achieve best trade-off of efficiency vs. linearity, which is of course dependent on the best high-Q inductor available with the given budget. Fig. 11 shows the measured efficiency of the EM with different bandwidths of the WiMAX 64QAM signals. Note the efficiency of the EM shown here is only reduced by 2.5% when the signal bandwidth increased from 1.5 MHz to 20 MHz.

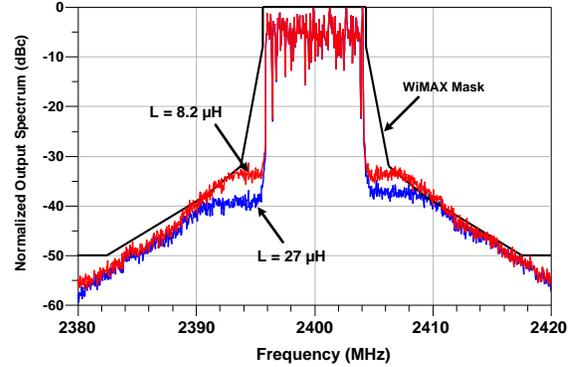


Fig. 10 Simulated output spectra of the ET-PA using different values of  $L$  for the WiMAX 64QAM 8.75 MHz signal. Realistic SPICE models were used for the SiGe PA and the EM.  $V_{DD} = 4.2$  V,  $P_{out} = 17$  dBm.

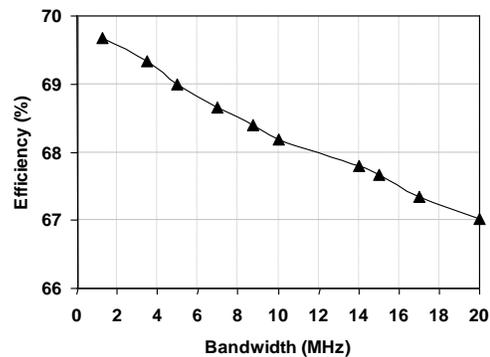


Fig. 11 Measured efficiency of the EM for different bandwidths of WiMAX 64QAM signals;  $V_{DD} = 4.2$  V,  $R_{load} = 22 \Omega$ , average output voltage = 2.3 V.

When the output current is low, the inductor may be completely discharged at the “OFF” state of the buck converter before the switcher is turned on again, which is often called as the “discontinuous mode” for DC-DC converter design [22]. Therefore, another concern in the selection of the inductor value is to ensure the buck converter does not go into the discontinuous mode operation [22]. The boundary of the discontinuous mode occurs at where the output DC current ( $I_o$ ) equals to one half of the peak-to-peak inductor ramp current  $\Delta I$  (i.e.,  $0.5\Delta I = I_o$ ). For the stand-alone buck converter controlled by the conventional pulse-width modulation (PWM) scheme, the minimal  $L$  should be determined to avoid the discontinuous mode at the minimum DC output current ( $I_{o,min}$ ) as [22]

$$L_{min} = \frac{(V_{DD} - V_{out,DC}) \cdot D}{\Delta I \cdot f_{SW}} = \frac{(V_{DD} - V_{out,DC}) \cdot V_{out,DC}}{2I_{o,min} \cdot f_{SW} \cdot V_{DD}} \quad (4)$$

, where  $V_{out,DC}$  is the output DC voltage,  $D$  is the duty cycle, and  $f_{SW}$  is the switching frequency determined by the PWM control scheme. For example, if the EM were to be implemented using the conventional PWM control scheme for this ET-PA, one could obtain the  $V_{out,DC} = 2.3$  V,  $I_{o,min} = 33$  mA (i.e., at  $R_{load} = 70 \Omega$  presented by the SiGe PA as shown in Fig. 2). Therefore, the minimal inductor value calculated based on (4) would be  $\sim 16 \mu\text{H}$  for a PWM-controlled buck-converter, assuming  $f_{SW} = 1$  MHz.

However, for the linear-assisted switching EM design presented here, the peak-to-peak inductor ramp current  $\Delta I$  is limited under  $2h/R_{\text{sense}}$ , which is not related with the inductor value [4]. This is because once the switching current  $I_{\text{sw}}$  is  $h/R_{\text{sense}}$  lower than the load current  $I_{\text{load}}$ , the hysteresis comparator will immediately sense the current difference and turn on the switcher again, assuming the switcher can response fast enough [4]. In the practical design, however, the switcher is not ideal due to its intrinsic gate capacitance and resistance, therefore it may not respond fast enough with a high switching frequency, and this frequency is directly determined by the inductor value. Also, the hysteresis window  $h$  increases with higher input slew rate [23]. The SPICE simulations show that  $h$  is  $\sim 7$  mV with the input voltage ramp below  $0.2$  V/ $\mu$ s, but increases to  $\sim 43$  mV with the input voltage ramp of  $4$  V/ $\mu$ s. According to the simulation, the minimal  $L$  is  $1.2$   $\mu$ H to avoid the discontinuous mode operation for this EM design. Fig. 12 shows the SPICE simulated current waveforms of the linear-assisted EM at the boundary of the discontinuous mode. Depending on the accuracy of the device modeling and the packaging parasitic, the inductance in the practical implementation may be lower than the simulated value.

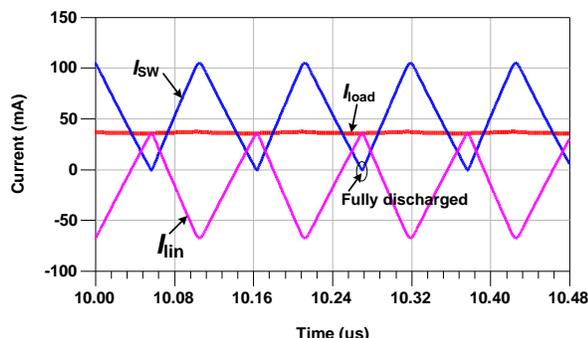


Fig. 12 SPICE simulated current waveforms of the discrete linear-assisted switching EM at the boundary of the discontinuous mode ( $L = 1.2$   $\mu$ H)

### III. EXPERIMENTAL RESULTS OF THE SiGe ET-PA

#### A. Efficiency and Linearity of the Stand-Alone PA

First, the stand-alone PA with fixed-supply voltage is tested to serve as a reference for the comparison with the ET-PA. Fig. 13 shows the measured gain, PAE and EVM vs.  $P_{\text{out}}$  for the WiMAX 64QAM 8.75 MHz signal (PAR = 10.5 dB) at 2.3 GHz. Without using any off-chip matching elements, the PAE of the SiGe PA reached 39% at the  $P_{\text{out}}$  of 17.8 dBm, but with a rather high output EVM of 11.7% (the EVM spec of WiMAX 64QAM is 5.0% or -26 dB). At  $P_{\text{out}}$  of 16 dBm, the stand-alone PA already violates the lenient EVM spec. Increasing  $V_{\text{CC}}$  could reduce the EVM as shown in Fig. 14, but at the cost of lower efficiency.

#### B. Efficiency and Linearity of the ET-PA

The discrete EM discussed earlier is used to modulate the supply voltage  $V_{\text{CC}}$  of the PA to form an ET-PA system. Please note that no predistortion is used in this work. The entire ET-PA operates at  $V_{\text{DD}}$  of 4.2 V. Fig. 15 shows the measured EVM, gain and overall PAE vs.  $P_{\text{out}}$  of the ET-PA.

Note the overall PAE (or the ET-PA composite PAE) includes the power consumption of the EM. The overall PAE is 30.5% at  $P_{\text{out}}$  of 17 dBm with an EVM of 4.4%. Judging from the PAE of  $\sim 50\%$  for the SiGe PA at  $P_{\text{out}}$  of 17 dBm (Fig. 2) and the efficiency of the EM of  $\sim 68\%$  (Fig. 11), the expected overall efficiency of the ET-PA would be 34%, which is close to the measured data observed from Fig. 15.

Fig. 16 shows the output spectra of the ET-PA and the stand-alone PA with fixed supply. At the same  $P_{\text{out}}$  of 17 dBm, the ET-PA successfully passes the stringent WiMAX 64QAM mask defined by European Telecommunications Standards Institute (ETSI), while the stand-alone PA fails the spectral mask badly. The ET-PA operates at its  $P_{2\text{dB}}$  point for  $P_{\text{out}}$  of 17 dBm, but still impressively passes the WiMAX emission mask. The better linearity of the ET-PA is probably due to the same envelope shaping function applied in the system as [24], which linearizes the AM-AM of the ET-PA across the instantaneous  $P_{\text{out}}$  range. Fig. 17 further shows that the maximum linear  $P_{\text{out}}$  of the fixed-supply PA is only  $\sim 13.5$  dBm in order to pass the WiMAX spectral mask, leading to a PAE of only  $\sim 26\%$ . Therefore, the ET-PA outperforms the fixed-supply PA by 3.5 dB on maximum linear  $P_{\text{out}}$  and 4.5% on PAE. Table I summarizes the performances of our ET-PA and its comparison with other state-of-the-art polar/ET-PAs.

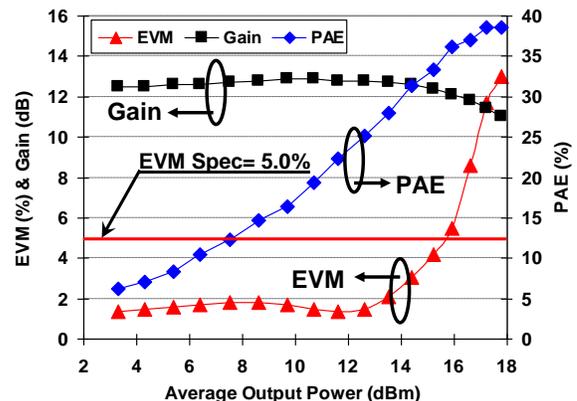


Fig. 13 Measured EVM, gain and PAE vs. average  $P_{\text{out}}$  of the stand-alone SiGe PA at  $V_{\text{CC}}$  of 3.6 V for WiMAX 64QAM 8.75MHz signal at 2.3 GHz.

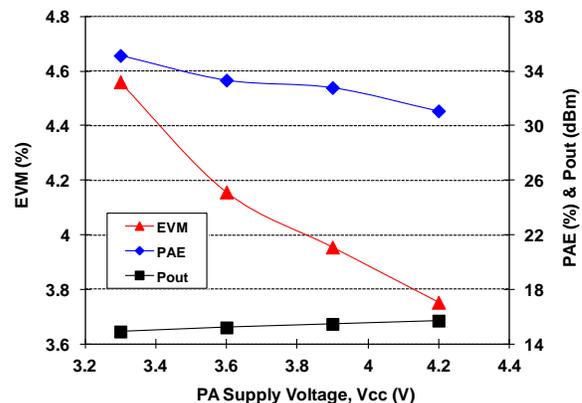


Fig. 14 Measured EVM, PAE and  $P_{\text{out}}$  vs.  $V_{\text{CC}}$  of the stand-alone PA for the WiMAX 64QAM 8.75 MHz signal at 2.3 GHz.  $P_{\text{in}} = 3$  dBm

TABLE I. PERFORMANCE SUMMARY AND COMPARISON OF OUR ET-PA WITH STATE-OF-THE-ART ET OR POLAR PAs

	Freq. (GHz)	Gain (dB)	<sup>(1)</sup> P <sub>out</sub> (dBm)	Overall PAE	EVM	EM Efficiency	Signal BW (MHz)	Modulation	<sup>(2)</sup> PD	Technology
[1]	0.9	8.41	20.41	44.4%	6.0%	65%	0.384	EDGE	No	PA: 0.18μm SiGe BiCMOS EM: discrete COTS
[4]	2.4	6.5	19	28%	2.8%	60%	20	WLAN 64QAM	Yes	Discrete COTS
[6]	1.75	---	23.8	22%	1.69%	---	0.384	EDGE	Yes	0.18μm CMOS
[10]	1.88	27.9	23.9	34.3%	2.98%	75%	5	WiMAX 64QAM	No	PA: 2μm InGaP/GaAs EM: 0.13μm CMOS
[17]	2.0	---	19.6	22.6%	2.5%	---	20	WLAN 64QAM	No	0.13μm CMOS
[18]	2.4	11	20	28%	5%	65%	20	WLAN 64QAM	Yes	0.18μm SiGe BiCMOS
[19]	1.56	---	14.7	8.9%	4.6%	---	20	WLAN 64QAM	No	0.18μm CMOS
[20]	1.92	---	15.3	22%	1.5%	---	5	WiMAX 64QAM	Yes	0.13μm SOI-CMOS
This work	2.3	10.5	17.0	30.5%	4.4%	69%	8.75	WiMAX 64QAM	No	PA: 0.18μm SiGe BiCMOS EM: discrete COTS

Note:

- (1) P<sub>out</sub>: The maximum linear P<sub>out</sub> that passes the linearity specs of the specific wireless standard.
- (2) PD: Predistortion
- (3) [1] [4] [10] [18] proposed ET-PAs, [6] [17] proposed polar PAs, and [19] [20] proposed digitally modulated polar PAs.

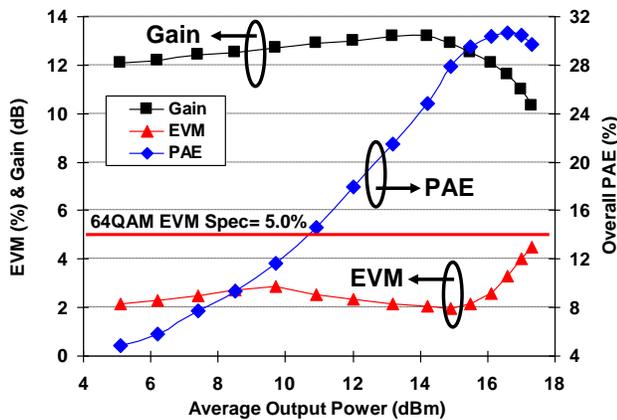


Fig. 15 Measured EVM, gain and overall PAE vs. P<sub>out</sub> of the ET-PA system for the WiMAX 64QAM 8.75 MHz signal at 2.3 GHz; V<sub>DD</sub> = 4.2 V

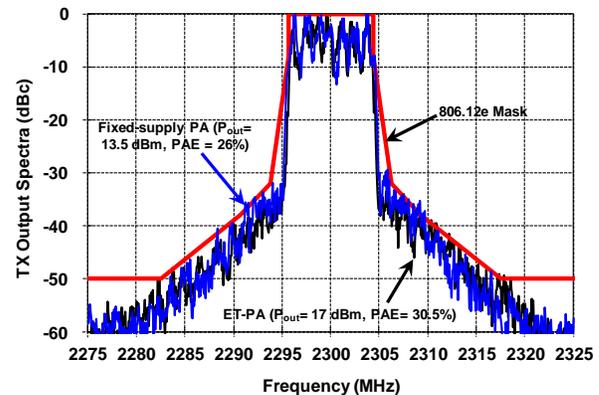


Fig. 17 Measured output spectra of the ET-PA (P<sub>out</sub>=17 dBm) and fixed-supply PA (P<sub>out</sub>=13.5 dBm) for WiMAX 64QAM 8.75 MHz at 2.3 GHz.

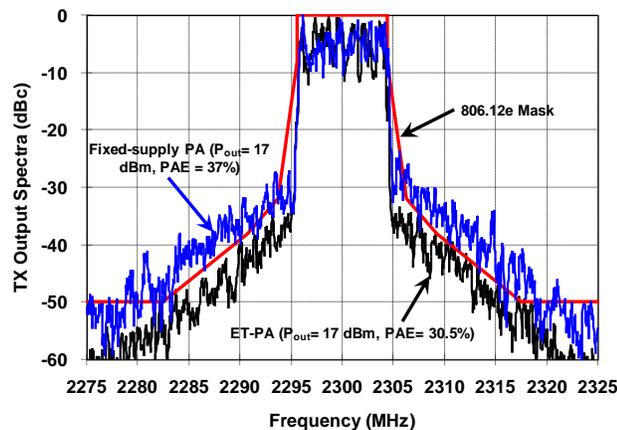


Fig. 16 Measured output spectra of the ET-PA and fixed-supply PA for WiMAX 64QAM 8.75 MHz at 2.3 GHz. P<sub>out</sub> kept at 17 dBm for both cases.

#### IV. CONCLUSION

The circuits and system design insights of a high efficiency and linear ET-PA system have been discussed using a SiGe PA and a discrete linear-assisted switching EM. The switching frequency of the switching stage and the bandwidth of the linear stage have been studied carefully to make the EM suitable for high PAR wideband applications. The WiMAX 64QAM 8.75 MHz signal was used to characterize the linearity and efficiency performances of the ET-PA system. Without needing any predistortion, the entire ET-PA system achieved an overall PAE of 30.5% at P<sub>out</sub> of 17 dBm with an EVM of 4.4%, while successfully passing the stringent WiMAX spectral mask.

#### ACKNOWLEDGMENT

The authors like to thank Dr. K.-S. Lu, CEO of Diodes Inc., Plano, TX, for generously donating the Keh-Shew Lu Regents Endowment Fund at Texas Tech University (TTU).

We also like to acknowledge of the funding support from Texas Instruments (TI) and other companies.

## REFERENCES

- [1] J. Lopez *et al.*, "Design of highly efficient wideband RF polar transmitter using the envelope-tracking technique," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2276-2294, Sept., 2009.
- [2] Y. Li *et al.*, "A Highly Efficient SiGe Envelope-Tracking Power Amplifier with an Integrated CMOS Envelope Modulator for Mobile WiMAX/3GPP LTE Transmitters," *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 10, pp. 2525-2536, Oct. 2011.
- [3] D. Y.C. Lie *et al.*, "Highly-efficient monolithic class E SiGe power amplifier design at 900 and 2400MHz," *IEEE Trans. Circuits Syst. I – Reg. Papers*, vol. 56, no. 7, pp. 1455-1466, July 2009.
- [4] F. Wang *et al.*, "An improved power-added efficiency 19-dBm hybrid envelope elimination and restoration power amplifier for 802.11g WLAN application," *IEEE Trans. Microw. Theory Tech.*, vol. 54, no. 12, pp. 4086-4099, Dec. 2006.
- [5] B. J. Minnis, P. A. Moore, P. N. Whatmough, P. G. Blanken, and M. P. van der Heijden, "System-efficiency analysis of power amplifier supply-tracking regimes in mobile transmitters," *IEEE Trans. Circuits Syst. I – Reg. Paper*, 56, 1, pp. 268-279, Jan. 2009.
- [6] P. Reynaert and M. S.J. Steyaert, "A 1.75-GHz polar modulated CMOS RF power amplifier for GSM-EDGE," *IEEE J. Solid-State Circuits*, vol. 40, no. 12, pp. 2598-2608, Dec. 2005.
- [7] D. K. Su and W. J. McFarland, "An IC for linearizing RF power amplifier using envelope elimination and restoration," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 2252-2258, Dec. 1998.
- [8] V. Pinon, F. Hasbani, A. Giry, D. Pache, and C. Garnier, "A single-chip WCDMA envelope reconstruction LDMOS PA with 130 MHz switched-mode power supply," *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers*, Feb. 2008, pp. 564-565.
- [9] J. Kitchen, W. Chu, I. Deligoz, S. Kiaei, and B. Bakkaloglu, "Combined linear and  $\Delta$ -modulated switch-mode PA supply modulator for polar transmitters," *IEEE Int. Solid-State Circuits Conf. Tech. Dig.*, Feb. 2007, pp. 82-83.
- [10] J. Choi, D. Kim, D. Kang, and B. Kim, "A polar transmitter with CMOS programmable hysteretic-controlled hybrid switching supply modulator for multistandard applications," *IEEE Trans. Microw. Theory Tech.*, vol. 57, no.7, pp. 1675-1686, July 2009.
- [11] R. Shrestha, R. v.d. Zee, A. d. Graauw, and B. Nauta, "A wideband supply modulator for 20 MHz RF bandwidth polar PAs in 65 nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 4, pp. 1272-1280, Apr. 2009.
- [12] F. H. Raab, "Split-band modulator for Kahn-technique transmitters," *IEEE MTT-S Int. Microw. Symp. Dig.*, 2001, pp. 887-890.
- [13] M. Kwak *et al.*, "Design of a wideband high-voltage high-efficiency BiCMOS envelope amplifier for micro-base-station RF power amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 60, pp. 1850-1861, June 2012.
- [14] F. Wang, A. H. Yang, D. F. Kimball, L. E. Larson, and P. M. Asbeck, "Design of wide-bandwidth Envelope-Tracking power amplifiers for OFDM applications," *IEEE Tran. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1244-1255, Apr. 2005.
- [15] M. Helaoui, S. Boumaiza, A. Chazel, and F. M. Ghannouchi, "On the RF/DSP design for efficiency of OFDM transmitters," *IEEE Trans. Microw. Theory Tech.*, 53, 7, pp. 2355-2361, 2005.
- [16] J. Jeong, D. F. Kimball, M. Kwak, C. Hsia, P. Draxler, and P. M. Asbeck, "Wideband envelope tracking power amplifiers with reduced bandwidth power supply waveforms and adaptive digital predistortion techniques," *IEEE Tran. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1244-1255, Apr. 2005.
- [17] J. S. Walling, S. S. Taylor, and D. J. Allstot, "A class-G supply modulator and class-E PA in 130nm CMOS," *IEEE J. Solid-State Circuits*, vol. 44, no. 9, pp. 2339-2347, Sept. 2009.
- [18] F. Wang, D. Kimball, D. Y.C. Lie, P. Asbeck, and L. E. Larson, "A monolithic high-efficiency 2.4-GHz 20-dBm SiGe BiCMOS envelope-tracking OFDM power amplifier," *IEEE J. Solid-State Circuits*, vol. 42, no. 6, pp. 1271-1281, June 2007.
- [19] A. Kavousian, D. K. Su, M. Hekmat, A. Shirvani, and B. A. Wooley, "A digitally modulated polar CMOS power amplifier with a 20-MHz channel bandwidth," *IEEE J. Solid-State Circuits*, vol. 43, no. 10, pp. 2251-2258, Oct. 2008.
- [20] C. D. Presti, F. Carrara, A. Scuderi, P. M. Asbeck, and G. Palmisano, "A 25 dBm digitally modulated CMOS power amplifier for WCDMA/EDGE/OFDM with adaptive digital predistortion and efficient power control," *IEEE J. Solid-State Circuits*, vol. 44, no. 7, pp. 1883-1896, July 2009.
- [21] I. Kim, Y. Woo, J. Kim, J. Moon, J. Kim, and B. Kim, "High-efficiency hybrid EER transmitter using optimized power amplifier," *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 11, pp. 2582-2593, Nov. 2008.
- [22] A. I. Pressman, *Switching Power Supply Design*, Second Edition, New York, NY: McGraw Hill, 1998.
- [23] R. Gregorian, *Introduction to CMOS Op-Amps and Comparators*, New York, NY: Wiley, 1999.
- [24] Y. Li *et al.*, "Circuits and system design of RF polar transmitters using envelope-tracking and SiGe power amplifier for mobile WiMAX," *IEEE Trans. Circuits Syst. I – Reg. Papers*, vol. 58, no. 5, pp. 893-901, May 2011.

# Fully Functional Passive RFID Tag with Integrated Sensor for Item Level Tagging Based on Collective Communications and Organic Printed Electronics

Predrag Jakimovski

Ali Hadda and Michael Zangl

Nina Oertel

Pervasive Computing Systems, TecO  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
Email: jakimov@teco.edu

Faculty of Computer Science  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
Email: {firstname.lastname}  
@student.kit.edu

SAP AG  
Karlsruhe, Germany  
Email: nina.oertel@sap.com

**Abstract**—The development of Organic Electronics (OE) is promoted by the promise of being ultra-low cost technology in the future. One key application scenario is item level tagging, e.g., smart packaging, where items are equipped with organic printed tags using Radio Frequency Identification (RFID) technology. However, current OE allows the printing of simple tags with ultra-low hardware complexity, but the printing of RFID chips with large numbers of transistors that implement RFID anti-collision protocols are far away from realization. In this paper, we present one sensor-based and fully functional passive RFID tag implementing collective communications a proprietary transmission protocol. Based on OE constraints, we describe the printable passive smart label down to the logic gate layer. Further on, we present an external clock synchronization strategy required by the passive reader system. Additionally, we provide a feasibility assessment towards printing the smart label, and conclude with simulating the proposed reader system.

**Keywords**—smart label; RFID; polymer electronics; organic printed circuits; wireless communication.

## I. INTRODUCTION

One key point in promoting polymer electronics is the promise of cost efficiency in relation to printing electronic devices in mass production. Another, the incorporated properties being light, thin and flexible are enabling novel applications in the area of pervasive systems and ubiquitous computing. The area of intelligent packaging is one key application for utilizing Organic Electronics (OE), for instance, the development of organic printed smart labels using Radio Frequency Identification (RFID) technology [1]. OE tags can contain and provide useful information, such as the item identification and sensor readings. This data is created continuously while items travel through the supply chain, i.e., from the manufacturing plant to the final consumer, and enable participants and other stakeholders to improve their business processes: quality control, certification, logistics optimization or fraud detection are but a few examples. Such dealing results in cost savings for the producer, suppliers, as well as vendors and, increases product safety for consumers. The approach we describe in this paper can be applied in those business processes. It can, for example, be used to determine product quality in bulk reading scenarios, such as the goods inbound processing in a supermarket. For each item the quality is determined by the sensor-based smart label attached to the item. In operation, the sensor readings are

acquired in a single query and interpreted by our algorithm that is specific to product type, e.g., to a quality rating on a scale from 1 to 7. The quality scale, however, is identical for all product types, i.e., a rating of 1 corresponds to a perished item for both milk packages and tomatoes. Otherwise, no compliance violation is detected, and the goods can be sold. Based on collective and simultaneous querying a single reader device can thus determine the quality ratings of multiple items belonging to different types.

Our paper is structured as follows. In Section II, we provide an overview about printed RFID technologies, and bring the state-of-the-art in relation to our smart RFID label approach. In Section III, we implement the smart label based on our earlier developed collective transmission protocol [2]. The specially developed communications for bulk reading of minimalistic tags is outlined in Subsection III-A and, its implementation into passive RFID circuitry is given in Subsection III-B. Thereafter, we describe an external clock synchronization strategy implemented within the proposed tag circuitry. The paper concludes with a validation of the proposed approach by utilizing simulations and a feasibility assessment towards printing the smart label device with current print technology.

## II. PRINTABLE SMART LABELS

RFID technology is in comparison to barcode in terms of memory capacity, readability, speed, being re-programmable, robustness and scalability vastly superior, but the high fabrication costs for Si-based RFID chips impede their wide deployment in item-level-based applications. However, the emergence of organic and printed electronics arises the opportunity to create affordable all-printed RFID tags, which can compete with the well established barcode system. Apart from the advantages of OE being thin, light-weighted, flexible and presumably ultra-low cost in the near future, the realization of this technology into printable smart labels is rather challenging, because the printing process itself is still an issue of current research. Further on, the created conductive inks indicate low electro-mobility which result into slow-switched transistors, i.e., yielding 627 Hz at 10 V electricity supply [3]. In addition, the printing of circuits on substrates, such as paper and plastic foils, consumes large areas, so that yet complex integrated circuits, such as a microchip can not be printed.

However, an all-printed 1-bit RFID tag based on carbon nano-tubes and gravure printing process has been reported from the Sunchon University [4] in 2010. More recently, the research group presented an improved version of 16-bit RFID tag at the IDTechEx RFID Europe event [5]. The RFID circuit is printed entirely on a polyester film, whereby nano-silver is used to print the antenna, because of the better interconnectivity. The all-printed rectifier of the tag provides a 10 V Direct Current (DC) from the 13.56 MHz reader signal. The all-printed ring oscillator can generate a clock signal of 102.8 Hz, which enables the transmission of a 96-bit transponder Identification number (ID) in a second. In collaboration with the company Paru their goal is to achieve printing 96-bit RFID tags conform to the ISO 14443 (International Organization for Standardization).

The company ThinFilm Electronics and its partner companies PARC developing addressable memory, and Polyera developing printed displays, presented recently a smart sensor label, which can be printed at high volumes on roll-to-roll printing process. Their smart label system consists of a writable memory, sensing capability and a printed display which indicates visually the information when, for instance, the temperature of the product has exceeded the maximum threshold. This system is based on line-of-sight reading, and thus less applicable for mass-reading scenarios, such as monitoring of pallets in a cool chain, where items are highly dense stacked and invisible for inspection. However, it offers an additional support in supply chains, such as picking out of the damaged and expired products in a super market. In this way, options for improving supply chain management are extended.

The first fully functional printed RFID tag based on silicon ink has been announced by the company Kovio. The internal circuit logic of the tag consists of about 1000 printed transistors, and has the capability to radio its stored data from a 128-bit printed Read-Only-Memory (ROM) to an interrogator. The transmission is based on High Frequency (HF) (13.56 MHz) and synchronous tag-talks-first protocol. The entire RFID circuit is printed by inkjet on a thin metal foil substrate. The application area focuses on Near Field Communication (NFC), such as the NFC Barcode [6]. An implementation of long range communication is not available. The printing process uses inorganic silicon-based technology that can reduce the cost advantage the polymer electronics is offering.

The research group Holst Center and their partners in the EU FP7 project ORICLA created a fully functional RFID tag based on polymer electronics and thin-film technology. The RFID prototype is made on plastic foil with organic thin-film semiconductors and realizes reader-talks-first communication. In contrast, the above presented approaches toward printed RFID tags are based on tag-talks-first principle: as soon as the RFID tag gets powered from the RF field of the RFID reader, it transmits its data to the reader. Here, the problem occurs, when many tags try to contact the reader at the same time, which requires an effective anti-collision mechanism. However, current generation of printed thin-film RFID tags are in fact enabled with a basic slotted ALOHA protocol, but the implementation of the anti-collision scheme is limited to about maximum 4 tags, and come at the cost of a slow reading.

Driven by lowering the costs for fabricating RFID tags, further techniques are promoted to realize item level tagging,

such as printable and chipless RFID technology. In contrast to printed RFID tags including an Integrated Circuit (IC) chip, printed chipless RFID tags are characterized mainly by encoding the product ID by using Time Domain Reflectometry-based (TDR) and Frequency Domain Signature-based (FDS) techniques, respectively. While TDR-RFID tags are based on microwave circulators and capacitors to generate RF wave transmission-lines in order to transmit the binary ID code, the FDS-RFID tags consist of resonators with different resonant frequencies which encode, i.e., the binary data. The coding capacity of TDR-based tags is currently limited by up to 8 bits, whereby FDS-based tags can contain theoretically unlimited capacity depending on frequency band and tag size. For example, Islam and Karmakar describe in their work [7] a 16-bit chipless RFID tag printed on a  $1.65 \times 1.65 \text{ cm}^2$  substrate area, which operates in the 6 – 13 GHz frequency band. The first type of fully printed chipless RFID tags suffer from limited encoding capacity, whereas the second type requires large-scale size and wide frequency bands, which can be scarcely deployed for applications operating in the Industrial, Scientific and Medical (ISM) radio bands. An overview about recent advances in printable chipless RFID technology is given in [8] and [9].

To reflect the related work the aim of all described approaches is focusing on implementing the EPCglobal Tag Data Standard (TDS) into printable RFID tags, which have the same capabilities as the conventional Si-based RFID technology is already providing. For instance, bi-directional communication between reader and tags, anti-collision protocol for avoiding interference, and large data storage capacity on tag. The implementation of these features into a polymer tag, that can be even printed on roll-to-roll process at high volumes, is currently rather challenging. The lack of printing integrated circuits with large numbers of transistors into small area and, the lack of mass production accuracy and reliability, is suspending the realization of fully functional organic and all-printed RFID tags into the far-distant future.

### III. COLLECTIVE COMMUNICATIONS FOR PRINTED SMART LABELS

In our earlier work, we developed with respect to future organic and printed smart labels robust and highly scalable communications based on collective transmission mechanism called *collective communications* [2][10]. In the following, we implement with regards to OE constraints the targeted smart label device into a fully functional polymer-based circuitry, and provide a feasibility assessment towards printing the smart RFID label in mass production.

#### A. Collective Communications

In general, collective communications is designed how to obtain information from a set of simultaneously sending nodes, in our application scenario the tags attached to goods on a shelf. We request from the set of items which proportion of tags measured (*proportion query*), respective, contain which values (*binary query*). In principle, this could be done by querying tags individually using any of the well-established protocols. However, implementing these protocols is not feasible in our scenario, since the senders need to be simple due to polymer electronics deficiency, and we consider large number

of senders. In particular, our transmission algorithm is laid out to be on the one hand simple, so that the smart label device can be implemented in polymer electronics, and on the other hand to acquire maximal performance regarding information transfer, scalability and robustness. We achieve these objectives by employing the following fundamental principles:

- 1) All queried tags transmit their stored data simultaneously and start transmission at the same time
- 2) A randomly drawn bit sequence encodes a physical entity, such as an identification, sensor value, temperature range, etc.

By applying the rules, the signals sent from tags superimpose on the RF channel. However, the transmission mechanism enables to analyze the superimposed signal statistically at the receiver side, so that the transferred information, for instance, which proportion of senders sent which value, can be extracted. Basically, our communication mechanism is based on bit sequences  $c$  of fixed length, that are shared between a sender  $S$  and a receiver  $R$ . A bit sequence  $v$  is sent from  $S$  as  $s = c \oplus v$ , where  $\oplus$  is the bitwise *exclusive or*. The receiver extracts  $v$  from  $s$  by computing  $v = s \oplus c$ . The double application of  $\oplus c$  cancels out  $c$  and  $v$  is regained. Simultaneous connections between a number of senders  $S_i$  and corresponding receivers  $R_i$  can then be achieved: simultaneous transmission yields the superimposed signal as the sum  $s = s_1 + s_2 + \dots + s_n$  of signals  $s_i$  sent, since the amplitudes of synchronized signals of the same frequency are approximately added to each other when the bit sequences  $s_i$  are sent. The resulting signal  $s$  is *similar* to each of the original signals  $s_i$ , where similarity can be based on various distance metrics on bit sequences  $v, w \in \{0, 1\}^n$ , e.g., on the Hamming distance:

$$d_H(v, w) = \sum_{i=1}^n |v_i - w_i|. \quad (1)$$

The similarity is then defined by choosing a threshold  $T_n$  suitable for the length of the vectors  $n$ . Two bit sequences  $v, w \in \{0, 1\}^n$  are called *similar* if they differ only in a small number  $T_n$  of bits:

$$v \sim w \stackrel{\text{def}}{\iff} d_H(v, w) \leq T_n. \quad (2)$$

A number of pairs of senders and receivers can thus communicate via codes  $c_i$ . If the codes  $c_i$  are chosen so as to be orthogonal ( $d_H(v, w) = 0$ ), or at least sufficiently different ( $d_H(v, w) \ll n$ ) from each other, this entails that we can obtain  $v_i$  from  $s$  by applying  $v'_i = s \oplus c_i$ . The result  $v'_i$  is similar to  $v_i$ , so that  $v_i$  can then be regenerated from  $v'_i$ , using error correcting codes. Codes  $c_i$  can be generated so as to be orthogonal, however, sufficiently long random bit sequences, are also suitable: statistical theory suggests that the probability to obtain two random bit sequences of low similarity is higher, the longer the sequences are. In order to ensure that different values transmitted can be retrieved from the superimposed signal, we directly encode the numerical values by using a single random bit vector  $z_0$  shared by all tags and the receiver. We obtain sufficiently different codes  $z_i$  for numerical values  $i$  by circularly shifting  $z_0$  by the amount of  $i$  bit, since shifting is a distancing operation. In this way, a single bit vector  $z_0 \in \{0, 1\}^n$  can be used to encode  $n$  values, and thus save memory capacity on the smart label device.

The received signal  $s = s_1 + s_2 + \dots + s_n$  is then simply a sum of encoded numbers  $z_i$ , directly encoding the multi-set of measured values. The algorithm for the reader system is outlined in Figure 1.

- 1) Tags come initialized with a register  $t$  set to default value 0, and transmit code  $z$  set to  $z_0$ .
- 2) Each tag measures continually its environment:
- 3) if the measured value is  $m > t$ , then
  - a) it sets  $t := m$ .
  - b) it shifts the code  $z$  accordingly:  
 $z := z_t$ .
- 4) Reader sends *start* signal to tags.
- 5) Tags send their respective  $z$ .
- 6) Reader receives overlaid signal  $s$ :
  - a) **Binary Query:**
    - i) Set  $S := \emptyset$ .
    - ii) For each possible value  $z_i$ :  
if  $z_i \sim s$  then  $S := S \cup \{z_i\}$ .
  - b) **Proportion Query:**  
For each value  $z \in S$ : use Least Squares Estimation to compute proportion of contribution of  $z$ :
    - i) Generate linear equation system for the found values  $z_i \in S$ .
    - ii) Estimate parameters  $a_i$  so that error is minimal.
    - iii)  $M := \{(a_i, z_i) | s = \sum_{z_i \in S} a_i * z_i\}$ .
  - c) **Output:** return  $M$ .

Figure 1. The Collective Communications algorithm.

## B. Circuit design

Based on our communication protocol described above in Subsection III-A, the implementation of the smart label device relies merely on features, such as fetching the sensor value, reading and sending out the corresponding bit sequence. Hence, the necessary building blocks of the circuit device are derived as follows: a sensor unit that consists of one Analog-Digital-Converter (ADC) and the sensor itself, a Read-Only-Memory (ROM) building block of fixed length that contains a randomly drawn binary sequence, and a circuit logic block that is enabled to generate the corresponding bit sequence to the sensed value, i.e., by employing the hardwired bit sequence stored in the ROM and carrying out an internal function operation on it. Further on, the smart label device contains a clock generator and counter to drive the transponder circuit. Exemplary, the block diagram in Figure 3 illustrates the interacting between all participating blocks of an 64-bit polymer tag, that implement, i.e., the clock generator realized by one ring oscillator, a 6-bit counter realized by 6 D-flip-flops, a ROM with 64-bit capacity, and a 3-bit ADC. Since the target is to create a reader system for passive polymer tags, the smart label circuitry includes in addition a DC rectifier concerning power supply, and a modulator transistor to convey the binary information to the reader device by deploying inductive coupling. Figure 4 shows in essence the schematic of the all-printable and passive polymer tag with its analog

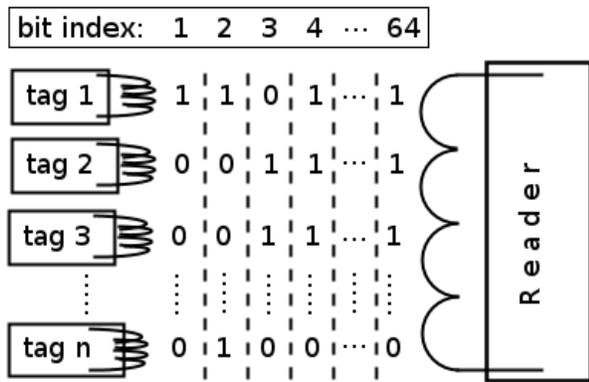


Figure 2. A passive tag reader system illustrates concurrent transmission of bit streams sent by several 64-bit smart labels.

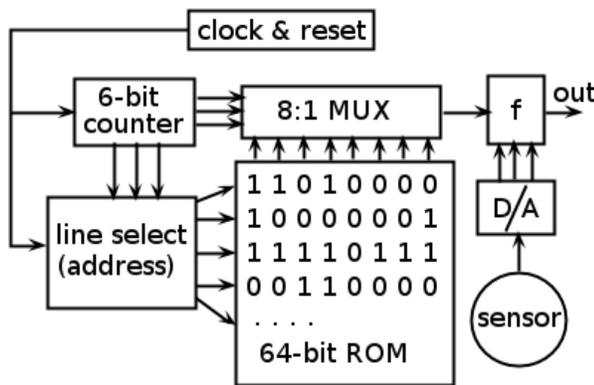


Figure 3. The circuitry implements the principle of collective communications.

and digital electronic components. The principle for collective communications is integrated in the protocol mechanism building block, where a bit sequence is generated according to the sensed value. The binary sequence is controlling the gate of the modulation transistor. When the modulator transistor is switched on, the transponder circuit draws power from the electromagnetic RF field of the reader. In the reverse, when the modulation transistor is switched off, the transponder circuit draws still power from the RF field, but considerably less than it is consuming in the on-state. In this way, the binary information is transferred by means of load modulation to the reader, i.e., by varying the power consumption.

C. Collective clock synchronization

With respect to bulk reading of smart labels in a dense area, i.e., when a set of tags are queried simultaneously, the response of the interrogated tags need to be collectively clock synchronized, but not phase synchronized as the collective transmission protocol is providing. On that issue, we developed for the reader system a collective synchronization method implemented into the transponder circuit. The circuit functionality is placed in the *clock & reset* building block (cf. Figure 4). Since conventional strategies, such as external reader-tag-synchronization applied in traditional RFID reader systems can not be employed straightforward in polymer electronics due to its hardware performance weakness, such as slow-

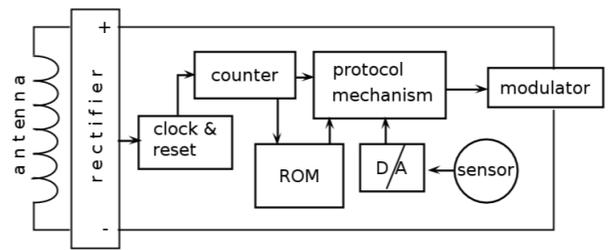


Figure 4. The circuit block diagram shows the holistic printable passive tag.

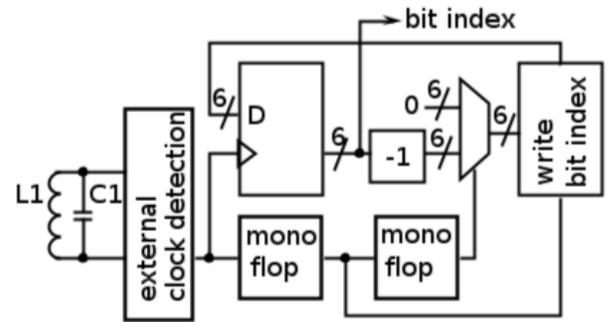


Figure 5. The schematic shows partially the circuit logic in connection with the tag rectifier for interpreting signaling coming from reader device.

switched printed transistors, appropriate solutions for reader-tag-communication needed to be elaborated. With respect to polymer electronics constraints, we developed collective synchronization procedures based on switching on/off the electromagnetic RF field of the reader device coil that empowers the polymer tags. By interrupting the wireless power supply all responsive tags can be instructed to either start with data transmission or, to send the next bit of their binary information. The transmission of data is based on inductive coupling. Figure 2 illustrates concurrent transmission of bit streams transmitted by several smart labels. In each synchronized time slot at the current bit index the bit information from all smart labels superimposes on the RF channel, and generates a specific overlaid signal. At the receiver side the superimposed signal in each time slot is captured and evaluated based on the collective communication algorithm (cf. mechanism in Figure 1).

In order to initiate every tag to start transmission at the same time the signaling can be done either by (i) turning off the reader device coil for a long time period and then start empowering the tags periodically with short intermissions, or (ii) begin with transmission by transmitting the start bit in a long time slot followed by transmitting the subsequently bit stream in regular sized time slots. Either way, the circuit logic in the clock & reset building block is tuned to interpret request orders sent from reader device based on capacitor discharge behavior and voltage comparator device. The circuit logic is connected with the tag rectifier shown in Figure 6. Parallel to the circuit logic that detects the external clock synchronization request, the internal clock counter in the clock & reset building block is resetting the bit index, and start increasing the count when the next intermission is registered. In this way, the smart label acquires the actual bit position in the bit sequence. In addition to the clock detection circuit a memory buffer realized through D-flip-flops preserves the bit index. This cache device

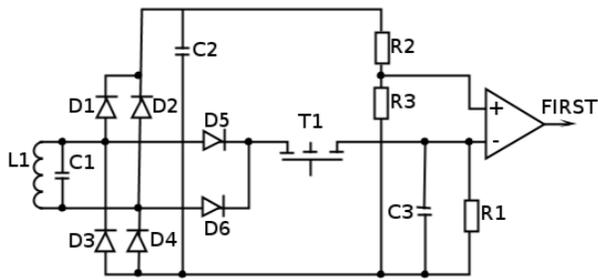


Figure 6. Schematic of the clock & reset building block.

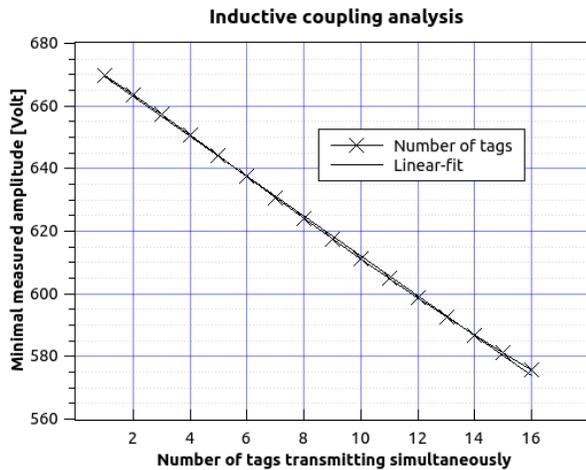


Figure 7. The overall signal analysis indicates the maximal amplitude difference between different trails with up to 16 simulated passive tags transmitting simultaneously.

keeps the storage contents even if the polymer tag is not powered for a short term, i.e., when the reader device causes an intermission to signalize the next time slot for transmission. Before increasing the bit index an additional memory buffer realized by D-latches stores temporarily the current bit index and induces the bit transmission. After increasing the bit index, it is delayed rewritten in the bit index buffer. The mono-flops in the circuit block provide the required delay for stabilized rewriting. The circuitry of the clock & reset block is depicted in Figure 5, whereby the external clock detection circuit is shown in Figure 6.

#### D. Feasibility assessment

The realization of printing electronic devices in mass production is one key issue of present-day research. With regard to organic and printed smart labels the state-of-the-art is described in Section II. All references indicate the feasibility of printing circuits with ultra-low hardware complexity, but non of them has yet enabled the printing of circuits implementing highly complex functionalities, such as an organic printed RFID chip with large number of transistors. With respect to current OE limitations the collective communication protocol was developed with the purpose to facilitate applications in the near future, though. The ability of exploiting concurrent and thus superimposed information transfer, has provided to move the hardware complexity from tag to the reader side. To the sake of completeness and, in order to indicate the hardware

complexity of the proposed passive tag, the remaining of the circuitry is shown in Figure 8. It entails a 6-bit counter that enables to read out the hardwired 64 bit sequence shown partially in Figure 3. Due to lack of space the sensor unit is not shown here. As can be seen from the circuit block diagram in Figure 4, the depicted circuits in Figure 5, in Figure 6 and in Figure 8, the entire hardware complexity of the tag is reduced to about few hundreds of transistors. Considering the tag circuitry in relation to already manufactured printed electronics by other research facilities, the fabrication in a clean room of 64-bit ROM devices has been experimentally verified [11]–[13]. The printing of holistic 4-bit RFID tags employing load modulation for communication is described in [14]. A challenge for realizing the proposed passive tag into polymer electronics is to achieve hazard-free printed circuit devices. For example, the analog components in the clock & reset building block, such as the comparator and the RC-element for timekeeping (cf. Figure 6), need to be printed with high accuracy, so that the circuit characteristic remains constant. Further on, the transmission range regarding polymer-based printed transponders indicates low modulation index as reported in [15]. This may constrain the free scalability property of collective communications, i.e., to read-out many tags at once.

#### E. Simulation

The communication protocol collective communications is designed to allocate information instantly from a large number of nodes transmitting data simultaneously to a receiver. The communication relies on ON-OFF-Keying (OOK), and works generally for active reader systems. Here, we analyze the applicability of the collective transmission mechanism for passive reader systems. For our analysis, we established in LTSpice a simulated passive reader system, where the antenna coil generates in HF an electromagnetic RF field providing up to 16 simulated passive transponder circuits with wireless power. Depending on passively switched on tags the change in the electromagnetic RF field, i.e., the amount of drawn energy from the RF field is recorded and evaluated at the reader side. With regard to parameterizing the simulation, we chose default parameters from standard RFID reader systems, such as 10 cm in diameter for the reader antenna coil, 2 cm in diameter for the tag antenna, which conforms to the inductive coupling coefficient of 0.025. The portability of the simulation with respect to polymer electronics is taken for granted, because the key parameter for inductance applies for printed antennas and circuits. In order to simulate different tag positions in relation to the reader antenna, respective, different sizes of tag antennas the coupling coefficient is varied. For the resonant circuit of the reader device generating the 13.56 MHz signal the hardware parameters are chosen according to Finkenzeller [16, Chapter 4.1]. The inductance is set by  $2\mu\text{H}$  with resistance of  $2.5\Omega$ . The capacitor is set by  $68.8\text{pF}$ . Further on, for the tag coil we chose  $100\text{nH}$  and  $0.1\Omega$ . The results of the signal analysis are shown in Figure 7, that indicate the minimal measured amplitude strength for  $n$  simulated passive tags transmitting simultaneously. Therewith, the maximal amplitude difference is acquired for distinguishing different number of responsive tags. Further on, the simulations show the more passive tags being responsive the greater the distinctive features appear among the captured superimposed signals.

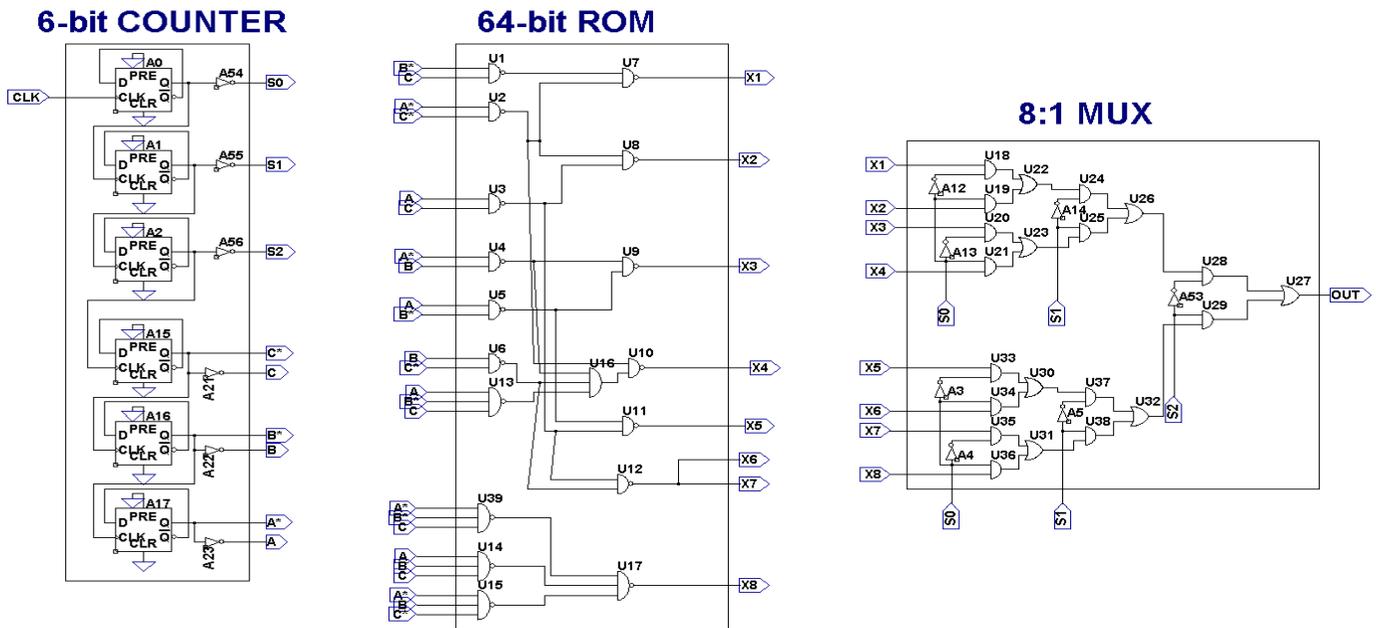


Figure 8. Schematic of the basic circuit arrangement realizing the part of circuit logic for a 64-bit polymer tag.

#### IV. CONCLUSION

With regard to smart packaging, item level tagging and organic printed smart labels, we described the implementation of one fully functional passive RFID tag, that is based on our earlier developed collective communication protocol and regards current organic electronics constraints. Apart from describing the state-of-the-art regarding printable smart labels and, putting this in relation to our smart label circuitry, we introduced with regard to reader-tag-communication an external and centralized clock synchronization strategy. Further on, we provided for the reader system a feasibility assessment, that indicates the printing of the proposed smart RFID label is in principle with present-day printing technology possible, but entails risks, such as low communication range and hazard-afflicted circuits that have been encountered by organic printed devices with similar hardware complexity. The causes for it were led back in the used conductive inks and fluctuations at the printing process. Hence, the specified hardware parameters may vary drastically, so that the printed polymer tags may exhibit unexpected and consequently unwanted behavior. However, the development and testing of novel conductive inks in material science, and improving the design rules for printing circuits on large area substrates, promises to resolve current issues of mass production in the near future.

#### ACKNOWLEDGMENT

This work was supported by the project PolytoS, in the BMBF funded cluster "Printed Organic Circuits and Chips".

#### REFERENCES

[1] Organic Electronics Association, "OE-A Roadmap for organic and printed electronics," White Paper, 2008.  
 [2] P. Jakimovski, F. Becker, S. Sigg, H. R. Schmidtke, and M. Beigl, "Collective communication for dense sensing environments," Intelligent Environments, International Conference on, vol. 0, 2011, pp. 157–164.

[3] K. Myny et al., "An inductively-coupled 64b organic rfid tag operating at 13.56mhz with a data rate of 787b/s," in Solid-State Circuits Conference, 2008. ISSCC 2008. Digest of Technical Papers. IEEE International, 2008, pp. 290–614.  
 [4] M. Jung et al., "All-printed and roll-to-roll-printable 13.56-mhz-operated 1-bit rf tag on plastic foils," Electron Devices, IEEE Transactions on, vol. 57, no. 3, 2010, pp. 571–580.  
 [5] P. Harrop. Sunchon university printed transistors. IDTechEx. [Online]. Available: <http://www.printedelectronicsworld.com/print-articles.asp?articleids=2664> [retrieved: April, 2014]  
 [6] The kovio website. [Online]. Available: <http://www.kovio.com/> [retrieved: April, 2014]  
 [7] M. Islam and N. Karmakar, "A novel compact printable dual-polarized chipless rfid system," Microwave Theory and Techniques, IEEE Transactions on, vol. 60, no. 7, 2012, pp. 2142–2151.  
 [8] B. Shao, Q. Chen, Y. Amin, R. Liu, and L. Zheng, "Chipless rfid tags fabricated by fully printing of metallic inks," annals of telecommunications - annales des telecommunications, vol. 68, no. 7-8, 2013, pp. 401–413. [Online]. Available: <http://dx.doi.org/10.1007/s12243-013-0378-3>  
 [9] S. Preradovic and N. Karmakar, "Fully printable chipless rfid tag," in Advanced Radio Frequency Identification Design and Applications, D. S. Preradovic, Ed., 2011.  
 [10] P. Jakimovski, H. R. Schmidtke, S. Sigg, L. W. F. Chaves, and M. Beigl, "Collective communication for dense sensing environments," JAISE, vol. 4, no. 2, 2012, pp. 123–134.  
 [11] E. Cantatore et al., "A 13.56-mhz rfid system based on organic transponders," Solid-State Circuits, IEEE Journal of, vol. 42, no. 1, 2007, pp. 84–92.  
 [12] A. Ullmann, M. Bhm, J. Krumm, and W. Fix, "Polymer multi-bit rfid transponder," in ICOE, June 2007, pp. 4–7.  
 [13] K. Myny et al., "Plastic circuits and tags for 13.56 mhz radio-frequency communication," Solid-State Electronics, vol. 53, 2009, pp. 1220–1226.  
 [14] J. Krumm and W. Clemens, "Printed electronics first circuits, products, and roadmap," in AACD Workshop, 2010, pp. 333–346.  
 [15] M. Böhm, A System Study for RFID. Transponders Based on Polymer Semiconductors. Cuvillier Verlag, 2007.  
 [16] K. Finkensteller and D. Müller, RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication. Wiley, 2010.

## An End-to-End QoS Performance Evaluation of VoLTE in 4G E-UTRAN-based Wireless Networks

Myasar R. Tabany

Wireless Communications Research Lab  
School of Systems Engineering, The University of Reading  
Reading, Berkshire, UK  
m.tabany@pgr.reading.ac.uk

Chris G. Guy

Wireless Communications Research Lab  
School of Systems Engineering, The University of Reading  
Reading, Berkshire, UK  
c.g.guy@reading.ac.uk

**Abstract**— Long Term Evolution (LTE) is the Fourth-Generation (4G) mobile broadband technology. Its standardization has been finalized by Third-Generation Partnership Project (3GPP) in Release 8 technical specifications (R8). As users' demand for higher data rate continues to rise, LTE and its ability to cost effectively provide fast, highly responsive mobile data services, a scalable bandwidth and a reduced latency will become ever more important. However, the Evolved Packet Core (EPC) of the Evolved Universal Terrestrial Radio Access Network (E-UTRAN) based wireless networks (LTE and LTE-Advanced) is all-IP Packet-Switched (PS) core network and lacks native support for Circuit-Switched (CS) services. This introduces the problem of how to provide voice services in these networks. The controversy around many of the proposed solutions to provide a PS voice and the effects of this step on the deployment of LTE networks is presented. This paper also deals with the Quality of Service (QoS) in a Voice over LTE (VoLTE) service. It provides a comprehensive evaluation and validation of VoLTE QoS based on the International Telecommunication Union standard Recommendations (ITU-R) and 3GPP standard technical specifications. The initial results obtained give clear evidence that the VoLTE service fulfills the ITU-R and 3GPP standard requirements in terms of end-to-end delay, jitter and packet loss rate. Furthermore, the results related to implementing different LTE bandwidths clearly reflect how these bandwidths affect the overall network performance and end-user experience.

**Keywords**- VoLTE; E-UTRAN; LTE; QoS; IMS.

### I. INTRODUCTION

The Third-Generation Partnership Project (3GPP) has developed a new technology called Long Term Evolution (LTE) in Release 8 (R8) Technical specification [1]. 3GPP LTE aims to improve the Third-Generation (3G) Universal Mobile Telecommunication System (UMTS) technology to meet the International Mobile Telecommunications Advanced (IMT-A) requirements determined by the International Telecommunication Union (ITU) [2]. Some of the agreed features of LTE are a significant increase in data rates with up to 300 Mbps downlink (DL) and 75 Mbps uplink (UL); a scalable bandwidth of 1.4, 3, 5, 10, 15 and 20 MHz and a reduced latency [3]. However, the changes in

this design were significant, with a flat all-IP Evolved Packet Core network (EPC) only supporting Packet-Switched (PS) services [1]. The EPC lacks native Circuit-Switched (CS) services support, including voice, which is considered as the main revenue for Mobile Service Providers (MSPs). This is different from most of the legacy UTRAN/GERAN wireless networks such as UMTS, which support both CS and PS services [2]. A user always expects voice as a basic service provided by the network operator and this raises the question of how to provide a voice call service to LTE users. The Evolved Universal Terrestrial Radio Access Network (E-UTRAN) is the radio wireless access for LTE and LTE-A and its architecture is simpler and flatter than Radio Access Network (RAN) in the 3G mobile networks as shown in Fig. 1. 3GPP presented the technical specifications for the E-UTRAN based networks (LTE and LTE-Advanced) in Release 8, 9 and 10. According to the 3GPP specifications, there is no guarantee that LTE has the ability to fulfil the ITU-R and 3GPP technical requirements related to QoS, especially with one way VoLTE end-to-end delay of less than 150 ms and a minimum of 98% packets successful delivery rate [4]. In this work, firstly we introduce a brief comparison between the proposed solutions for deploying voice service over LTE wireless networks. We also evaluated the VoLTE QoS performance in terms of end-to-end delay, jitter and packet loss rate. For this purpose, we designed a realistic baseline simulation for LTE wireless networks based on 3GPP R8 technical specifications, including IMS. We then simulated different LTE bandwidths and depending on the results we investigated the effects of deploying these bandwidths on the service quality and end-user experience.

The rest of this paper is organized as follows. Section II briefly introduces a description of the VoLTE proposed technologies. Section III explains the QoS architecture in LTE wireless networks. Section IV describes the designed simulation environment. Section V discusses and analyses the simulation results and finally, we conclude this work in Section VI.

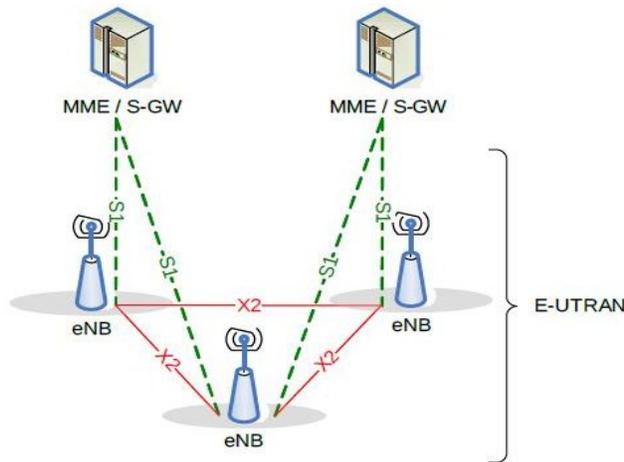


Figure 1. E-UTRAN System Architecture [5]

## II. DEPLOYMENT OF VOLTE OVER 3GPP 4G E-UTRAN -BASED NETWORKS

This paper discusses mainly Voice over LTE (VoLTE) technology based on IP Multimedia Subsystem/ Multi Media Telephony (IMS/MMTel), which is standardized by 3GPP to provide voice service to LTE wireless networks. Other proposed technologies such as Circuit Switched Fall Back (CSFB), Voice over LTE via Generic Access (VoLGA) and Over The Top (OTT) were investigated and described briefly in this paper.

### A. VoLTE via IMS/MMTel

Voice is a fundamental service to consider in any Next Generation Mobile Networks (NGMNs). In fact, IMS with MMTel are the key to make this possible and provide a required High Definition (HD) telephony system to LTE [2] [6]. In VoLTE, a software upgrade is required to the LTE network and its PS core network (EPC). VoLTE uses a QoS Class Indicator value equal to one (QCI=1) and the Conversational QoS class for either originating or terminating a voice call. For more detailed information about the procedure of UE to originate a voice call in a roaming scenario refer to [7]. According to the 3GPP technical specification in [6] IMS is an access independent based on the Session Initiation Protocol (SIP), defined by the Internet Engineering Task Force (IETF) to support voice and other multimedia services. The reference architecture of IMS is illustrated in Fig. 2. IMS provides a complete solution to handle voice over all-IP and PS wireless networks. GSM Association (GSMA) announced in 2010 it will consider IMS as a major solution in the one voice profile recommendations [8].

The first step of User Equipment (UE) to start a voice call is an IMS registration. Next the UE obtains the required bearers to complete the call followed by IP address allocation to be known by other users. Multi Media Telephony (MMTel) has originated in 3GPP Release 7. It is a service set in the IMS standard architecture that defines both Network to Network Interface (NNI) and User to Network Interface (UNI) [10]. It offers real time multimedia

services based on IMS and allows users to use voice and other services. One of the major roles of MMTel is to maintain service quality of a minimum performance voice and video which support 3GPP codecs.

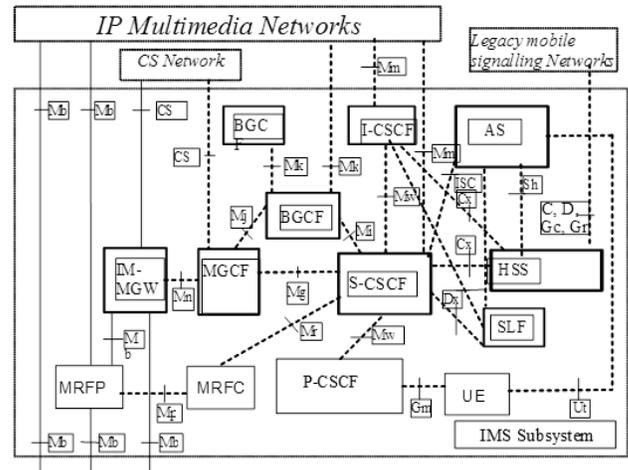


Figure 2. The IMS Reference Architecture [9]

### B. Circuit Switched Fall Back (CSFB)

CSFB is a 3GPP standard bridging technology between the LTE PS and legacy CS wireless networks to obtain CS services [11]. The NGMN alliance has recommended CSFB to enable non-IMS roaming subscribers to use both PS and CS voice services in legacy CS networks. The precondition in CSFB is the LTE coverage must overlap with UTRAN/GERAN. CSFB was specified in the 3GPP technical specification in [11]. CSFB is an interim solution which is suitable to use when the visiting LTE networks do not have IMS or IMS still not fully deployed.

### C. Voice over LTE via Generic Access (VoLGA)

VoLGA is a different mechanism which was defined by the VoLGA forum in 2009 based on the 3GPP Generic Access Network (GAN) specified in [12] and [13]. VoLGA connects the LTE PS network with MSC/VLR in UTRAN/GERAN using a special gateway called VoLGA Access Network Controller (VANC). However, VoLGA has not been accepted by the 3GPP standardization body as standard technology to provide voice to LTE users, which is the biggest disadvantage of this technology.

### D. Over The Top (OTT)

Over The Top (OTT) means providing voice services through third party providers such as Skype or Google Talk. This service is provided either free of charge or is relatively inexpensive. Mobile operators might use OTT when they do not want to invest too much on deployment a very expensive IMS. However, there is no guaranteed QoS and no service continuity using this method, especially when UE moves outside an LTE coverage area. Call drop and call failure is always possible using this method [13].

### III. QoS IN E-UTRAN-BASED WIRELESS NETWORKS

Quality of Service (QoS) is a concept of providing a particular quality for a specific type of service. QoS is one of the main and the greatest challenges for the IP-based services that lack of a dedicated connection channel. As part of the rapid growth of multimedia applications over cellular networks, the QoS needs to be maintained a guaranteed service through wireless networks [14]. It is essential for LTE to provide an efficient QoS solution that the user experience of each service running over the shared radio link is satisfied. Thus, the Evolved Packet System (EPS) system selects different QoS data flows for each service.

#### A) QoS Architecture in LTE Networks

3GPP introduced the QoS architecture of the LTE/EPS in R8 technical specifications. As can be seen from Fig. 3, this end-to-end class-based QoS architecture has been introduced to support a mix of Real Time (RT) and non-Real Time (non-RT) services.

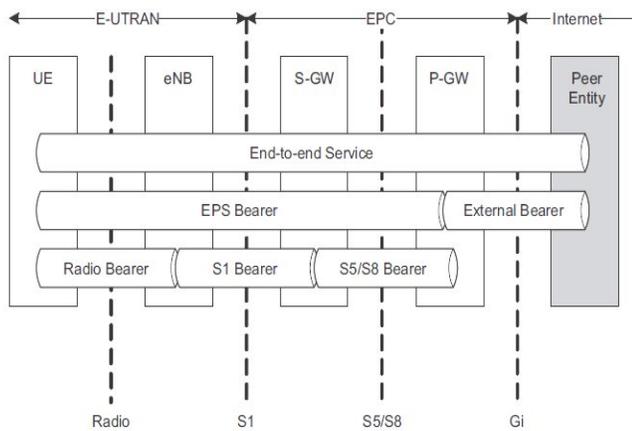


Figure 3. LTE/EPS Bearer Service Architecture [17]

The QoS in EPS is based on the data flows concept and bearers. Such flows of data are established between the UE and the Packet Data Network Gateway (PDN-GW) and mapped to bearers, with three individual bearers (Radio, S1 and S5/S8). The combinations of them provide the end-to-end QoS support to the LTE system. With the help of bearers, a scalar value, referred to as a QoS Class Identifier (QCI) with the help of bearers specifies the class to which the bearer belongs [15]. Table I illustrates the standardized QoS classes.

#### B) QoS for Voice over LTE (VoLTE)

For running VoLTE service over LTE networks, two default and one dedicated bearers are mostly required [18]. The first bearer is the default bearer, which is used for signaling messages. IMS uses this default bearer with QCI=5 for any SIP signaling related to it. The Packet Delay Budget (PDB) in this bearer is 100 ms between the PDN-

GW and the UE with up to  $10^{-6}$  Packet Loss Rate (PLR). This default bearer has the highest priority amongst all other QCI classes. The second bearer is also a default bearer, which is used for all other TCP-based traffic (e.g., Email). Up to 300 ms PDB is allowed in this bearer with a maximum of  $10^{-6}$  PLR. It has the lowest priority among all other QCI classes. The last bearer is dedicated bearer. This dedicated bearer is used for conversational voice (VoLTE). Up to 100 ms PDB is allowed with maximum  $10^{-2}$  PLR. This bearer has the second highest priority among all other QCI classes and unlike the other default bearers, it is always GBR. It is associated with the first default bearer with Linked EPS Bearer ID (L-EBI) and has also Traffic Flow Template (TFT) which determines the rules of sending and receiving IP packets.

TABLE I. STANDARDIZED QCI CHARACTERISTICS [16]

QCI	Resource type	Packet error/loss rate	Delay Budget	QCI priority	Example services
1	GBR	$10^{-2}$	100 ms	2	Conversational voice
2			150 ms	4	Real-time video
3		$10^{-3}$	50 ms	3	Real-time gaming
4			300 ms	5	Buffered video
5	Non-GBR	$10^{-6}$	100 ms	1	IMS signaling
6			300 ms	6	Buffered video, email
7		$10^{-3}$	100 ms	7	Voice, RT video
8		$10^{-6}$	300 ms	8	TCP-based services
9				9	

### IV. SIMULATION ENVIRONMENTS

The OPNET modeler from Riverbed Technologies Ltd. has been used to design the baseline for the LTE wireless network. The practical side of this study started by designing a baseline for the LTE wireless network. This baseline includes a complete implementation for VoLTE, the topic in our study. The LTE baseline network consists of seven LTE base stations (eNBs), three mobile stations (UEs) in each eNB (total of 21 UEs in the whole network), one IP Multimedia System (IMS) in addition to the LTE core network (EPC). The designed network contains Application Definition, Profile Definition, Mobility Management and OPNET LTE configuration entities. In addition, a number of wired and wireless links to connect between different nodes and the LTE EPC were used. Mobility is implemented with node velocity equal to 5 meters per second (m/s). Intra-frequency is the only handover likely to happen between cells of the same frequency. The only path loss model used in each UE is set to free space with no any obstruction for the propagated signal. The UE and the eNB transmission power are set to cell size based for both of them. This parameter is used to represent the transmission power (in wattage) to use for each UE/eNB. 20 MHz FDD bandwidth

has been chosen to simulate the physical profile in the LTE designed network. In order to give a very realistic scenario to the designed network, we assumed that all the 7 eNBs are located in London, UK, which will be our simulation work space with 1 km eNB radius each. Additionally, we adopted a hexagon overlay for the implemented eNBs and put the UEs randomly between these eNBs for the same reason. The overall network model of the baseline LTE network is shown in Fig. 4. The LTE mobile nodes (UEs) are programmed and configured to run VoLTE services. The `lte_wkstn_adv` node model is used to represent a workstation with source and destination application running over TCP/IP and UDP/IP. Table II shows the important configuration parameters of the UEs in the baseline LTE network.

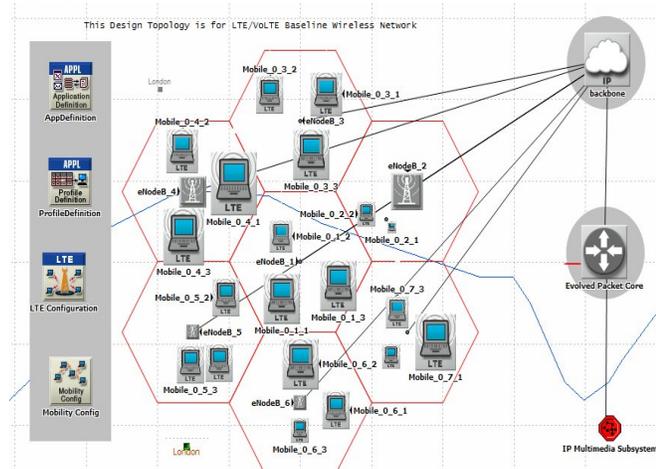


Figure 4. The Baseline LTE Designed Network (Network Model)

TABLE II. USER EQUIPMENT (UE) CONFIGURATION PARAMETERS

PARAMETER	VALUE
ANTENNA GAIN dBi	-1 dBi
MODULATION and CODING SCHEME INDEX	9
MULTIPLE CHANNEL MODEL (DL)	LTE OFDMA ITU Pedestrian B
MULTIPLE CHANNEL MODEL (UL)	LTE SC-FDMA ITU Pedestrian B
PATH LOSS MODEL	FREE SPACE
DL MIMO TRANSMISSION	Same NB Setting
NUMBER of RECEIVE ANTENNAS	2
NUMBER of TRANSMIT	1
HANDOVER TYPE	INTRA-FREQUENCY
VELOCITY	5 M/S
MEASUREMENT WINDOW SIZE	100 ms
CELL RESELECTION MEASUREMENT THRESHOLD	-112 dBm

The LTE base stations (eNBs) are programmed and configured to provide radio coverage to the UEs in the LTE network. The `lte_enodeb_3sector_4slip_adv` node model is used to represent the LTE eNBs. This model of eNBs includes 3 sectors in each eNB and can maintain up to 4 serial line interfaces at a selectable data. Any one of the 7 eNBs nodes can communicate with one or more UEs, in

addition to the EPC. The eNBs, UEs and EPC in the design network have been programmed in a way such that each one of them has a unique ID and name. Table III shows the important configuration parameters of the eNBs in the designed network. The IMS model is used to deliver High Definition (HD) voice and a set of Rich Communications Services (RCSs); it also gives a more realistic scenario to deliver VoLTE service. The IMS model consists of Proxy Call Session Control Function (P-CSCF), Serving-CSCF (S-CSCF) and Interrogating-CSCF (I-CSCF). These components are used for the signaling procedures of the VoLTE calls between different users in the network. The IMS signaling flow in the LTE network requires the highest priority as it is the first procedure which is invoked towards the establishment of the VoLTE call. Hence, all the IMS signaling packets are marked with priority equal to 1 in both radio and core networks in the QCI [17]. Note that 1 is IMS priority while 5 is the value of IMS QCI.

TABLE III. EVOLVED NODEB (eNB) CONFIGURATION PARAMETERS

PARAMETER	VALUE
ANTENNA GAIN dBi	15 dBi
DUPLEXING SCHEME	FDD
PATH LOSS MODEL	Free Space
BANDWIDTH	20 MHz
NUMBER OF RECEIVE/ TRANSMIT ANTENNAS	2
HANDOVER TYPE	INTRA-FREQUENCY
DL MIMO TRANSMISSION TECHNIQUE	Spatial Multiplexing Codewords 2 Layers
MEASUREMENT THRESHOLD	-44 dBm
eNB SELECTION THRESHOLD	-110 dBm

The EPC is one entity which includes all the main required core network parts; the Mobility Management Entity (MME), the Serving Gateway (S-GW), and the Packet Data Network Gateway (PDN-GW). The `lte_access_gw_atm8_ethernet8_slip8_adv` model is used to represent the LTE core network. The voice model used to generate VoLTE in the designed LTE network is a G.711 Pulse Code Modulation (PCM) voice codec. A summary of the G.711 parameters which have configured in the baseline simulation is illustrated in Table IV.

TABLE IV. VOICE CONFIGURATION PARAMETERS

PARAMETER	VALUE
VOICE FRAMES PER PACKET	1
TYPES OF SERVICE	Interactive Voice
SILENCE LENGTH (SECONDS)	0.65 Second
TALK SPURT LENGTH	0.352 Second
COMPRESSION DELAY	0.02 Second
DECOMPRESSION DELAY	0.02

The VoLTE application in OPNET simulation enables two UEs to establish a virtual channel and they can communicate using digitally encoded voice signals. Fig. 5 describes the data traffic flow in OPNET LTE simulation

between EPC and the LTE eNB through GPRS Tunneling Protocol (GTP). GTP tunneling is located at both nodes (EPC, eNB) and it is dynamically established between them to carry the EPS required bearers shown before in Fig. 3. The User Datagram Protocol (UDP) is the default transport protocol used for this application. The voice data arrive in spurts called talk spurts that are followed by silent periods. A talk spurt is an uninterrupted burst or a period of time in which the listener does not detect a pause. During a silent period, packets are transmitted quite rarely. Internally, the voice packets are sent over Real-Time Protocol (RTP) streams. Traffic is generated in the network model only when the application is active, therefore the traffic duration equals the application duration. The voice application in our simulation starts at 100 seconds from the simulation start time. The period of time before the traffic is generated is called warm-up time, which is very important for any simulation scenario [4]. The reason why this time is important is because any simulation running is started with empty systems. During this time all buffers are configured before starting the traffic generation. The node and the PHY process models of the LTE eNB and LTE UE of our simulation are illustrated in Fig. 6 and Fig. 7, respectively.

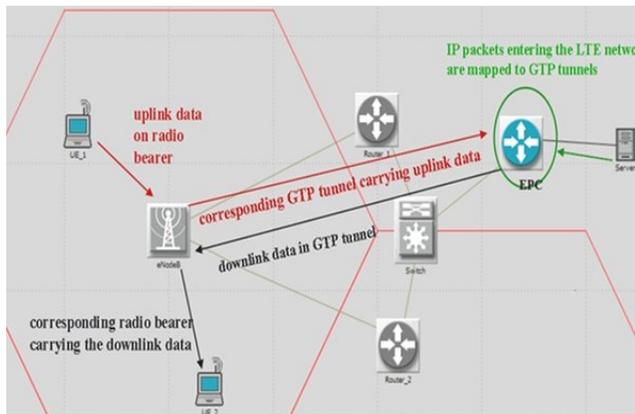


Figure 5. GTP Tunneling Between LTE EPC and LTE eNB and Data Traffic Flow

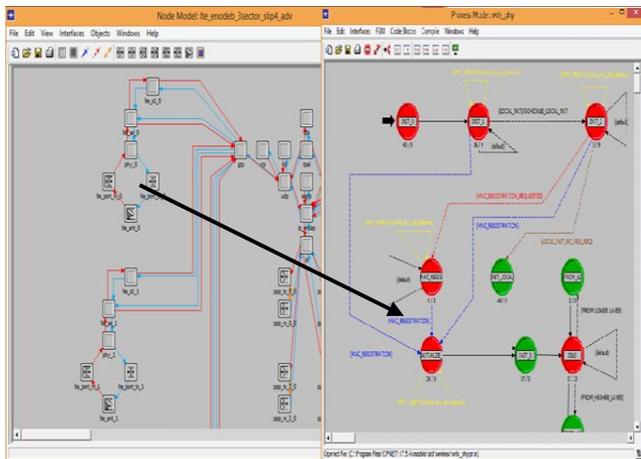


Figure 6. Node and PHY Process Models for LTE eNB

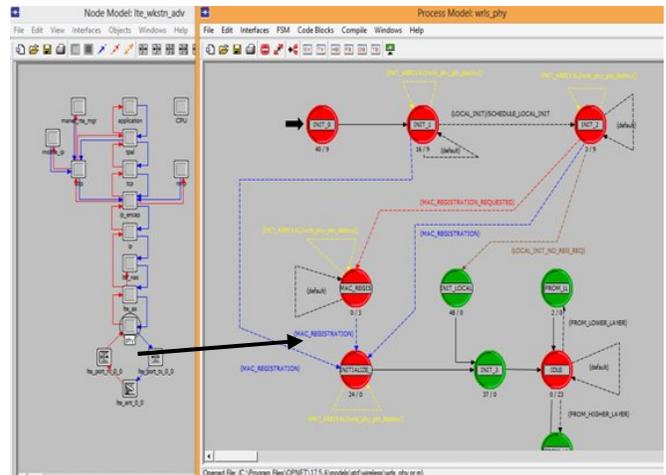


Figure 7. Node and PHY Process Models for LTE UE

## V. SIMULATION RESULTS

The performance evaluation has been conducted in terms of different QoS factors such as end-to-end delay, packet loss rate and jitter in two main scenarios.

### A. VoLTE QoS Baseline Scenario

The VoLTE end-to-end delay (mouth-to-ear delay) is one of the most important factors to consider when we measure the VoLTE QoS. It should be strictly maintained under reasonable limits and must be carefully monitored. End-to-End delay is measured from the ingress of the UE at the sender side to the egress of the UE at the receiver side. The equation used to calculate this QoS factor based on our simulation design is:

$$\text{VoLTE end\_to\_end\_delay} = \text{Network delay} + \text{Encoding delay} + \text{Decoding delay} + \text{Compression delay} + \text{Decompression delay} + \text{Dejitter\_buffer delay} \quad (1)$$

Fig. 8 shows end-to-end delay for VoLTE service in the baseline LTE network during 600 seconds simulation time. The X-axis represents the simulation time in seconds, while the Y-axis represents the end-to-end delay in seconds. The VoLTE traffic starts at 100 seconds as we mentioned earlier (see Section IV for more details). At 100 seconds, the end-to-end delay was 165 ms. There was a slight increase in this value to become 184 ms and then a gradual decrease until reaching its stable level equal to 119 ms. It then continues with that value until the end of the simulation. The average end-to-end delay for the VoLTE is found 126 ms. This value fulfills the ITU-R and the 3GPP standard requirements with up to 150 ms for one way VoLTE end-to-end delay to experience high quality [19] [16]. In fact, an end-to-end delay up to 250 ms is still quite satisfactory for the majority of users if we considered about 100 ms required delay for packet processing and propagation delay in the core network [20] [21].

The VoLTE packet loss rate is another important QoS factor to examine. Packet loss rate generally refers to the percentage of packets that are lost during the transition from the sender to the receiver in the network. Ideally, in VoLTE

steady networks, there should be no packet loss [22]. However, VoLTE users are still satisfied if this percentage is a maximum 2% based on the 3GPP requirements [23] [16]. This means, at least 98% of the total transmitting packets have to arrive successfully to the final destination. The equation to calculate the packet loss rate is as follows:

$$\text{Packet Loss Rate} = \frac{\text{Packets sent} - \text{Packets received}}{\text{Packets sent}} * 100 \% \quad (2)$$

Fig. 9 demonstrates a comparison between VoLTE traffic sent and traffic received. The X-axis represents the simulation time in seconds, while the Y-axis represents the VoLTE traffic sent/received in packets per seconds. It is clear from Fig. 9 that the VoLTE traffic sent/received increased sharply after 100 seconds. At 104 seconds, the rate of the packets sent was 69.64 packets/second, while on the other hand, the number of the packets received was 67.41 packets/second at the same time. The traffic reaches its peak values at 138 seconds and then it stays on this steadily level with 1800 packets/second until the end of the simulation. Overall, the amount of traffic generated and received was almost identical. This is due to the stable LTE network that does not involve any congestion by other applications. The total number of packets sent/received in the baseline scenario were 290300/290286 packets respectively. The packet loss rate for VoLTE was found to be 0.0048%, which is an excellent rate. This result meets the ITU-R and 3GPP standard requirements which were clarified before in Section I.

The difference in response time between different packets received in the destination side is called jitter. For any stable system with steady packet stream, the value of this QoS factor should be always 0 as there is no variation in the delay of the received packets. However, if the jitter is so large then it can cause an out of order situation to the receiving packets. This can lead to confusion if the working application is a voice service, which results in poor service quality. The ITU-R has recommended 25 ms jitter as an acceptable value for the delay variation [21]. From Fig. 10, it can be seen that the jitter value was - 0.00000136 at time 120 seconds and then after 6 seconds (time=126 seconds) becomes 0 and stayed at this value until the end of the simulation. Negative jitter indicates that the time difference between the packets at the destination node was less than that at the source node. This result indicates clearly that the overall jitter value is ideal and reflects that the designed system is very stable.

#### B. LTE Bandwidth Implementation Scenario

One of the interesting features of LTE is its ability to support scalable bandwidths from 1.4 MHz up to 20 MHz. It is necessary to examine the effects of using different LTE bandwidths on the service quality and end-user experience. In the second scenario, we implemented other bandwidths (1.4 MHz and 5 MHz). Fig. 11 compares between end-to-

end delay in the three different LTE bandwidths (1.4, 5 and 20 MHz). The higher the bandwidth (20 MHz) is the higher the data rate supported, as a result, the lowest end-to-end delay. 1.4 MHz is the highest end-to-end delay in the same figure. However, all the results for all the implemented bandwidths are still within the acceptable threshold of 150 ms. Fig. 12 shows a comparison between the same groups of LTE bandwidths, but in this case in terms of the VoLTE jitter. For jitter calculation, we used the following formula:

$$\text{Jitter} = (T4 - T3) - (T2 - T1) \quad (3)$$

Where:

T1 and T2: The time of leaving two consecutive packets from the source node.

T3 and T4: The time of arrival same packets to the destination node.

Contrary to expectations, Fig. 12 shows that LTE jitter with 20 MHz has better jitter values than LTE jitter with 5 MHz, which in turn also, has a better jitter values than 1.4 MHz as it includes many negative values. These results reflect that the higher the data rate gives better jitter performance. However, in fact, theoretically, there is no direct relation between LTE bandwidth and its jitter. The higher bandwidth can affect positively on the end-to-end delay, but not necessarily lead to reduced jitter value.

The downlink (DL) delay in LTE networks is the time started from when the traffic arrives at the LTE layer of the eNBs until it is delivered to the higher layer of the corresponding UEs. On the other hand, the uplink (UL) delay in LTE networks is the time started from when the traffic arrives at the LTE layer of the UEs until it is delivered to the higher layer of the corresponding eNBs. It is straightforward to show that the Packet Delay in the DL direction is less than the UL direction, although the resulting delay in both sides follows the same standard requirements of 50 ms one way delay [20]. This is due to the higher power in the DL side from the eNBs which is around 43 dBm, compared to the UE's power which is about 23 dBm in the UL side [17]. The UL/DL delay in the designed LTE network for different LTE bandwidths is illustrated in Fig. 13 and Fig. 14 respectively. As can be seen from these figures, the relation between the UL/DL delay and the LTE bandwidth is directly proportional. The results show that there is up to 0.03 seconds DL delay and up to 0.017 seconds UL delay. These results meet the requirements mentioned before, for the one way voice radio UL/DL delay.

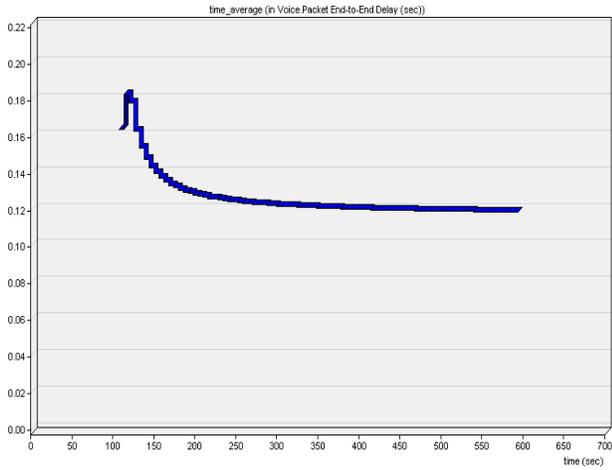


Figure 8. VoLTE End-to-End Delay

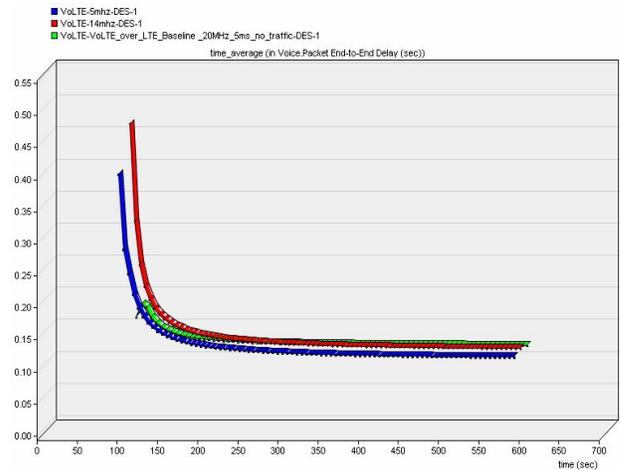


Figure 11. End-to-End Delay for Different LTE Bandwidths

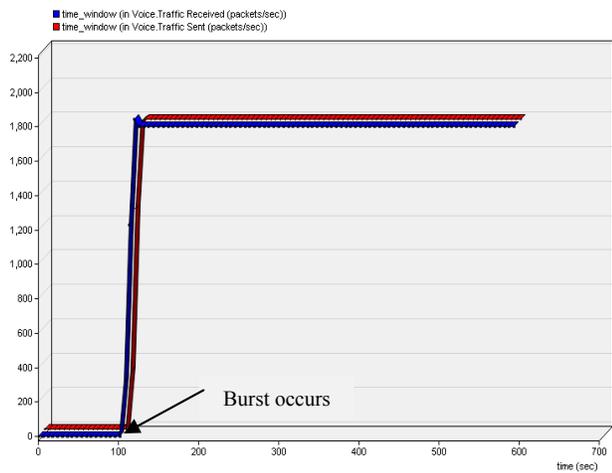


Figure 9. VoLTE Traffic Sent/Received

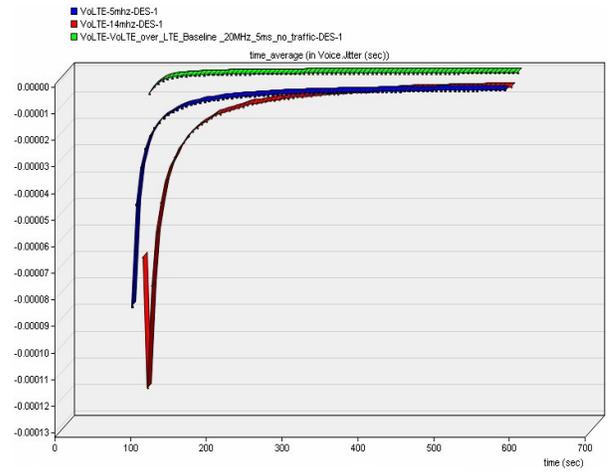


Figure 12. Jitter for Different LTE Bandwidths

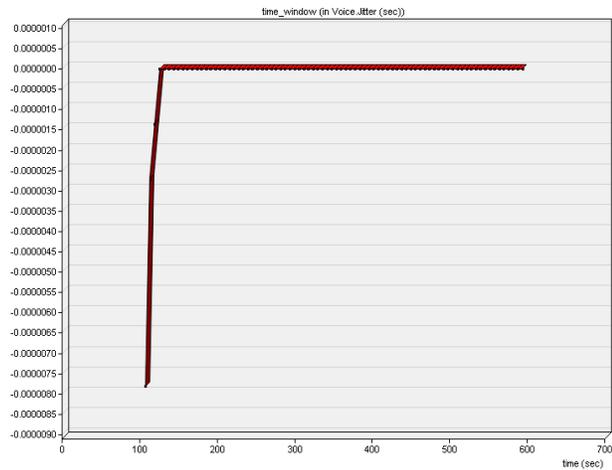


Figure 10. VoLTE Jitter

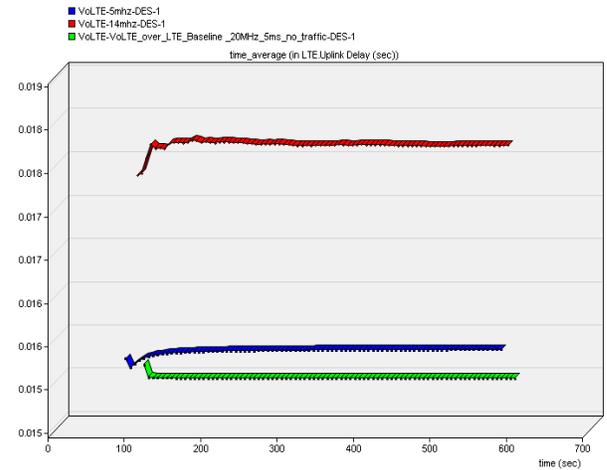


Figure 13. LTE Uplink Delay for Different Bandwidths

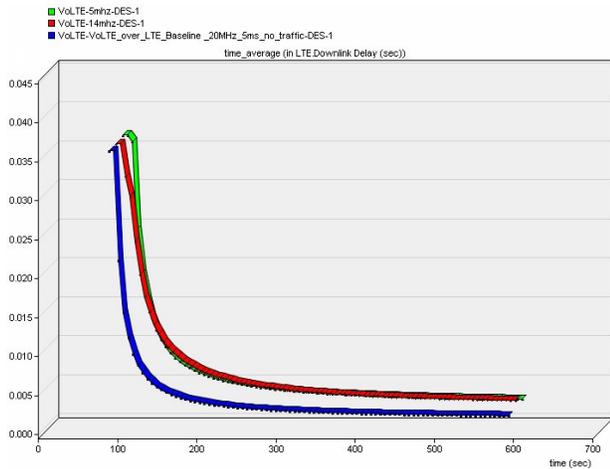


Figure 14. LTE Downlink Delay for Different Bandwidths

## VI. CONCLUSION

In this paper, a realistic Voice over LTE (VoLTE), including IMS over the baseline LTE wireless network was simulated and its performance in terms of Quality of Service (QoS) was evaluated and validated using OPNET modeler wireless suite 17.5. VoLTE is a standard technology that is required to support packet voice calls over a purely Packet-Switched (PS) LTE wireless networks. It provides better QoS, which results in better end-user experience over CSFB, VoLGA and OTT. In conclusion, this work has demonstrated that the simulation results have matched the ITU-R and 3GPP standard requirements related to the VoLTE over 4G LTE. The simulation results are significant in three different QoS respects; end-to-end delay, jitter and packet loss rate. It has been found that the overall VoLTE end-to-end delay was about 0.12 ms, and its packet loss rate was about 0.005%, while jitter was almost 0. Furthermore, another different simulation scenario was designed to investigate the effects of different LTE bandwidths on the VoLTE service quality. Three different LTE bandwidths (1.4, 5 and 20 MHz) were implemented and their effects on the VoLTE end-to-end QoS and LTE DL and UL delays were also studied. The results show that LTE can achieve better performance with a 20 MHz bandwidth.

## ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to Riverbed Technologies Ltd., for providing the OPNET modeler wireless suite 17.5 licenses required for this work.

## REFERENCES

[1] ETSI TS 136.300, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access (E-UTRAN); Overall description; Stage 2," v8.4.0, April 2008.

[2] M. Sauter, Beyond 3G-Bringing networks, terminals and the web together: LTE, WiMAX, IMS, 4G Devices and the Mobile Web 2.0: Wiley, com, 2011.

[3] D. Astély, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," Communications Magazine, IEEE, vol. 47, pp. 44-51, 2009.

[4] ITU-T Recommendation G.114, "One way transmission time", February 2003.

[5] C. C. Lin, K. Sandrasegaran, H. Ramli, and M. Xue, "Requirement of handover modeling in the downlink 3GPP long term evolution system," in Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on, 2010, pp. 305-310.

[6] ETSI TS 124.237, "LTE; IP Multimedia (IM) Core Network, IMS service continuity; Stage 3," v10.3.0, June 2011.

[7] M. R. Tabany and C. G. Guy, "Performance Analysis and Deployment of VoLTE Mechanisms over 3GPP LTE-based Networks," International Journal of Computer Science and Telecommunications, Vol. 4, Issue. 10, Oct. 2013, pp.1-8.

[8] R. Ferrus, O. Sallent, and R. Agusti, "Interworking in heterogeneous wireless networks: comprehensive framework and future trends," Wireless Communications, IEEE, vol. 17, pp. 22-31, 2010.

[9] 3GPP TS 23.228, "Technical specification Group Services and System Aspects; IP Multimedia Subsystem (IMS); Stage 2, (Release 9) v9.3.0, March 2010.

[10] ETSI TS 122.173, "UMTS; LTE; IP Multimedia Core Network Subsystem (IMS) Multimedia Telephony Service and supplementary services; Stage 1," v10.3.0, April 2011.

[11] ETSI TS 123.272, "Technical Specification Group Services and System Aspects; Circuit Switched (CS) fallback in EPS; Stage 2," (Release 10), v9.13.0, January 2013.

[12] VoLGA TS stage 3, "Voice over LTE via Generic Access; Stage 3 Specification; Phase1," v1.5.0, June 2010.

[13] 3GPP TR 43.318, "Technical Specification Group GSM/EDGE Radio Access Network; Generic Access Network (GAN); Stage 2 (Release 8)," v8.4.0, February 2009.

[14] L. Li and S. Shen, "End-to-End QoS performance management across LTE networks," in Network Operations and Management Symposium (APNOMS), 2011 13th Asia-Pacific, 2011, pp. 1-4.

[15] H. Ekstrom, "QoS control in the 3GPP evolved packet system," Communications Magazine, IEEE, vol. 47, pp. 76-83, 2009.

[16] 3GPP TS 23.203, "Technical Specification Group Services and System Aspects; Policy and charging control architecture (Release 10)," v10.6.0, March 2012.

[17] S. Sesia, I. Toufik, and M. Baker, LTE: the UMTS long term evolution: Wiley Online Library, 2009.

[18] GSMA PRD IR.92: "IMS Profile for Voice and SMS," March 2013.

[19] Y. Fan, M. Kuusela, P. Lunden, and M. Valkama, "Downlink VoIP support for evolved UTRA," in Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE, 2008, pp. 1933-1938.

[20] J. Puttonen, T. Henttonen, N. Kolehmainen, K. Aschan, M. Moisio, and P. Kela, "Voice-over-IP performance in UTRA Long Term Evolution downlink," in Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, 2008, pp. 2502-2506.

[21] S. A. Ahson and M. Ilyas, VoIP handbook: applications, technologies, reliability, and security: CRC Press, 2008.

[22] Cisco, "Quality of Service for VoIP," Cisco systems, 2011.

[23] T. Henttonen, et al., "Performance of voip with mobility in ultra long term evolution," in Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE, 2008, pp. 2492-2496.

# Enhancing the Vector-Based Forwarding Routing Protocol for Underwater Wireless Sensor Networks: A Clustering Approach

Dina M. Ibrahim, Tarek E. Eltobely, Mahmoud M. Fahmy, and Elsayed A. Sallam

Computers and Control Engineering Department

Faculty of Engineering, Tanta University

Tanta, Egypt

emails: {dina.mahmoud@f-eng.tanta.edu.eg, tarekt@f-eng.tanta.edu.eg, mfn\_288@hotmail.com, and sallam@f-eng.tanta.edu.eg}

**Abstract**— Underwater Wireless Sensor Networks (UWSNs) have an important role in different applications, such as offshore exploration and ocean monitoring. The networks consist of a considerably large number of sensor nodes deployed at different depths. Many routing protocols have been proposed in order to discover an efficient route between the sources and the sink. In this paper, we propose an algorithm to improve the performance of the Vector-Based Forwarding (VBF) protocol which we call a Clustering Vector-Based Forwarding algorithm (CVBF). In the proposed algorithm, the space volume of the network is divided into a number of clusters where one virtual sink is assigned to each cluster. Then, the nodes inside each cluster are allowed to communicate with themselves just to reach its virtual sink node, which in turn sends the packets to the main sink in the network. Simulation results demonstrate that the proposed algorithm reduces the energy consumption especially in dense networks, increases the packet delivery ratio especially in sparse networks, and decreases the average end-to-end delay in both sparse and dense networks. These advantages are emphasized when the algorithm is compared with four other powerful routing algorithms: VBF, Hop-by-Hop VBF (HH-VBF), Vector-Based Void Avoidance (VBVA), and Energy-Saving VBF (ES-VBF) routing protocols.

**Keywords**—wireless networks; underwater sensor networks; multiple clusters; routing protocols.

## I. INTRODUCTION

At the end of the twentieth century, wireless sensor networks became a hot research area. At the beginning, these networks covered only terrestrial applications. However, the earth is known to be a water planet, with 70 % of its surface being covered with water (principally oceans). With the increasing role of oceans in human life, discovering all of the ocean parts became of prime importance. On one side, traditional approaches formerly used for underwater monitoring missions have several drawbacks [1] and on the other side, these harsh environments are not feasible for human presence as unpredictable underwater activities, high water pressure, predatory fish and vast areas are major reasons for un-manned exploration. Due to these reasons, Underwater Wireless Sensor Networks (UWSNs) attract the interest of many researchers lately, especially those working on terrestrial sensor networks [2]. Over the last three decades, significant contribution has been made in the area

of scientific, commercial, and military applications [3]. In particular, highly precise real-time continuous-monitoring systems are essential for vital operations such as off-shore oil field monitoring, pollution detection, disaster prevention, assisted navigation, mine reconnaissance, and oceanographic data collection. All these significant applications call for building UWSNs. The work done by Akyildiz et al. [4] is considered as the pioneering effort towards the deployment of sensor nodes for underwater environments.

Though there exist many network protocols for terrestrial wireless sensor networks, the underwater acoustic communication channel has its unique characteristics, such as limited bandwidth capacity and high delays, which require new efficient and reliable data communication protocols [5]. Major challenges in the design of underwater wireless sensor networks are: i) the limited bandwidth; ii) the underwater channel is severely impaired, especially due to multipath and fading problems; iii) high propagation delay in underwater which is five orders of magnitude higher than in Radio Frequency (RF) terrestrial channels; iv) high energy consumption due to longer distances; v) battery power is limited and usually batteries cannot be recharged, also because solar energy cannot be exploited underwater; vi) underwater sensor nodes are prone to failures due to fouling and corrosion. All the factors mentioned above, especially limited energy, would make designing a routing protocol for UWSN an enormous challenge.

Routing is a fundamental issue for any network, and routing protocols are considered to be in charge of discovering and maintaining the routes [2]. Most of the research works concerning UWSNs have been on the issues related to the physical layer, while issues related to the network layer such as routing techniques are a relatively new area. Thus, an efficient routing algorithm is to be provided. Although underwater acoustic networks have been studied for decades, underwater networking and routing protocols are still at the infant stage of research.

A review of underwater network protocols till the year 2000 can be found in [1]. Several routing protocols have been proposed for underwater sensor networks. A good survey until year 2012 about underwater wireless sensor routing techniques is presented in [2]. Here, Ayaz et al. introduced an overview of the state of the art of routing protocols in UWSNs and thoroughly highlighted the advantages, functionalities, weaknesses and performance

issues for each technique. Based on network architecture, UWSNs routing protocols are classified into: location-based, flat, and hierarchical routing protocols. Vector-Based Forwarding (VBF) protocol has been suggested in order to solve the problem of high error probability in dense networks [6]. It is a location-based routing protocol. Here an idea of a virtual routing pipe from the source to the destination is proposed, and all the flooding data packets are carried out through this pipe. An enhanced version of VBF called Hop-by-Hop VBF (HH-VBF) has been proposed [7]. They use the same concept of virtual routing pipe as used by VBF, but instead of using a single pipe from source to destination, HH-VBF defines per hop virtual pipe for each forwarder [8]. Another extension of VBF protocol is introduced in [9] called Vector-Based Void Avoidance (VBVA) routing protocol which extends the VBF routing protocol. It addresses the routing void problem in underwater sensor networks. VBVA assumes two mechanisms, vector-shift and back-pressure, to handle voids. In [10], an energy-aware routing algorithm, called Energy-Saving Vector Based Protocol (ES-VBF), is proposed. In this protocol, Bo et al. put forward an energy-aware routing algorithm to save network energy. It takes both residual energy and location information into consideration, which shows a promising performance in balancing network energy consumption and packet reception ratio.

Other UWSNs routing protocols, such as Dynamic Source Routing (DSR), Time division Multiple Access (TDMA), Focused beam Routing (FBR), Directional Flooding-Based (DFR), and Depth-Based Routing (DBR) are found in [2][8][11][12].

The remainder of this paper is organized as follows. In Section II, the functionality and performance issues of VBF, HH-VBF, VBVA, and ES-VBF location-based routing protocols which will be used in a comparison with our algorithm are discussed. Section III presents the details of the proposed algorithm. In Section IV, we show the performance results of the proposed algorithm. Finally, we draw the main conclusions in Section V.

## II. REVIEW OF LOCATION-BASED ROUTING PROTOCOLS

In this section, we discuss in brief four location-based routing protocols which we will choose to compare our algorithm with. These protocols are:

### A. Vector-Based Forwarding (VBF) Routing Protocol

VBF is a location-based routing approach for UWSNs proposed by Xie et al. [6]. In this protocol, state information of the sensor nodes is not required since only a small number of nodes are involved during packet forwarding. Data packets are forwarded along redundant and interleaved paths from the source to the sink, which helps handling the problem of packet losses and node failures. It is assumed that every node previously knows its location, and each packet carries the location of all the nodes involved including the source, forwarding nodes, and final destination. The forwarding path is specified by the routing vector from the sender to the target. As soon as a packet is received, the node computes its relative position with respect to the forwarder.

Recursively, all the nodes receiving the packet compute their positions. If a node determines that it is close enough to the routing vector, it puts its own computed position in the packet and continues forwarding the packet; else, it simply discards the packet. In this way, all the packet forwarders in the sensor network form a “routing pipe”, the sensor nodes in this pipe are eligible for packet forwarding, and those which are not close to the routing vector do not forward. Fig. 1 illustrates the basic idea of VBF. In this figure, node  $S_1$  is the source, and node  $S_0$  is the sink. The routing vector is specified by  $S_1S_0$ . Data packets are forwarded from  $S_1$  to  $S_0$ . Forwarders along the routing vector form a routing pipe with a pre-controlled radius,  $W$ .

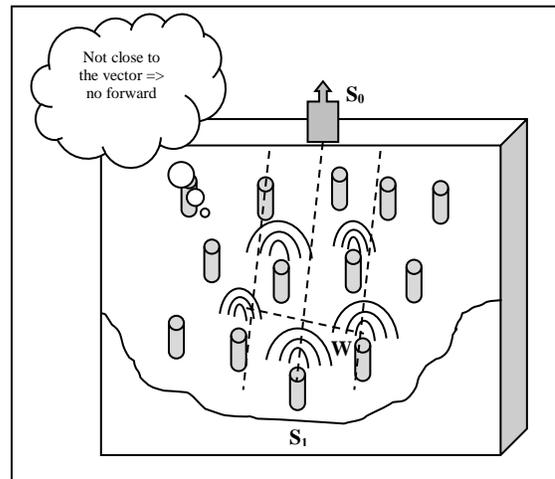


Figure 1. VBF routing protocol for UWSNs.

Additionally, a localized and distributed self-adaptation algorithm is developed to enhance the performance of VBF [6]. The self-adaptation algorithm allows each node to estimate the density in its neighborhood and forward packets adaptively. This algorithm is based on the definition of a desirableness factor,  $\alpha$  [6]. This factor measures the suitability of a node to forward packets. Given a routing vector  $S_1S_0$  and forwarder  $F$ , the desirableness factor of a node  $A$  is:

$$\alpha = \frac{P}{W} + \frac{R - d \times \cos \theta}{R} \quad (1)$$

where  $P$  is the projection length of  $A$  onto the routing vector  $S_1S_0$ ,  $d$  is the distance between node  $A$  and node  $F$ ,  $\theta$  is the angle between vector  $FS_0$  and vector  $FA$ ,  $R$  is the transmission range, and  $W$  is the radius of the routing pipe.

Fig. 2 represents the different parameters used in the definition of the desirableness factor [6]. From the definition, we see that for any node close enough to the routing vector, i.e., inside the pipe ( $0 \leq P \leq W$ ), the desirableness factor of this node is in the range of  $[0, 3]$  depending on position of node  $A$ .

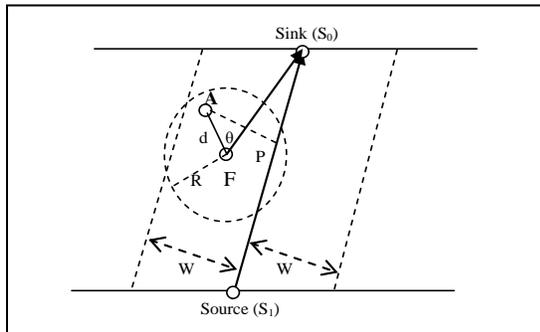


Figure 2. Desirability factor in self-adaptation algorithm.

In this algorithm, when a node receives a packet, it first determines if it is eligible for packet forwarding (i.e., close enough to the routing vector) [6][7]. If yes, the node then holds the packet for a time period,  $T_{adaptation}$ , related to its desirability factor and other network parameters (2). In other words, each qualified node delays forwarding the packet by a time interval calculated as follows:

$$T_{adaptation} = \sqrt{\alpha} \times T_{delay} + \frac{R-d}{v_0} \quad (2)$$

where  $T_{delay}$  is a pre-defined maximum delay,  $v_0$  is the propagation speed of acoustic signals in water, i.e., 1500m/s, and  $d$  is the distance between this node and the forwarder [7]. Principally, this self-adaptation algorithm gives higher priority to the desirable node to continue forwarding the packet. The theoretical analysis can be found in [6].

VBF has many essential drawbacks. First, using a virtual routing pipe from source to destination can affect the routing efficiency of the network with different node densities. In some spaces, if node deployment is sparser or become sparse due to some node movement, then it is possible that very few or even no node will lie within that virtual pipe, which is responsible for the data forwarding; even it is possible that some paths may exist outside the pipe. Eventually, this will result in small data deliveries in sparse spaces. Second, VBF is very sensitive about the routing pipe radius threshold, and this threshold can affect the routing performance significantly; such feature may not be desirable in the real protocol developments. Furthermore, some nodes along the routing pipe are used again and again in order to forward the data packets from sources to the sink, which can exhaust their battery power.

### B. HH-VBF Routing Protocol

The need to overcome two problems encountered by the VBF, i.e., small data delivery ratio in sparse networks, and sensitivity to the routing pipe's radius, the HH-VBF (hop-by-hop VBF) is proposed by Nicolaou et al. [7]. HH-VBF forms the routing pipe in a hop-by-hop method, enhancing the packet delivery ratio significantly. Although it is based on the same concept of routing vector as VBF, instead of using a single virtual pipe from the source to the sink, it defines a different virtual pipe around the per-hop vector from each forwarder to the sink. In this protocol, each node can

adaptively make packet forwarding decisions based on its current location. This design can directly bring the following two benefits: First, since each node has its own routing pipe, the maximum pipe radius is the transmission range. Second, in sparse networks, HH-VBF can find a data delivery path even so the number of eligible nodes may be small, as long as there exists one in the network.

In HH-VBF, the routing virtual pipe is redefined to be a per-hop virtual pipe, instead of a unique pipe from the source to the sink [7]. When some areas of the network are not occupied with nodes, for example there exist "voids" in the network, even a self-adaptation algorithm may not be able to route the packets. In such a case, a forwarder is unable to reach any node other than the previous hop. Although simulation results show that HH-VBF considerably produces better results for packet delivery ratio, but still it has an inherent problem of routing pipe radius threshold, which can affect its performance. Moreover, due to its hop-by-hop nature, HH-VBF is not able to add a feedback mechanism to detect and avoid voids in the network and energy efficiency is still low compared to VBF [7].

### C. VBVA Routing Protocol

Xie et al. [9] introduce a Vector-Based Void Avoidance (VBVA) routing protocol, which extends the VBF routing protocol to handle the routing void problem in UWSNs. VBVA assumes two mechanisms, vector-shift and back-pressure. The vector-shift mechanism is used to route data packets along the boundary of a void. The back-pressure mechanism routes data packets backward to bypass a concave void. VBVA handles the routing void problem on demand and thus does not need to know network topology and void information in advance. Hence, it is very robust to cope with mobile voids in mobile networks. Simulation results in [9] show that VBVA can handle both concave and convex voids effectively and efficiently in mobile underwater sensor networks only when these voids are inside the forwarding pipe, while the voids outside the forwarding pipe is not solved by VBVA.

### D. ES-VBF Routing Protocol

To solve the energy problem in UWSN, Bo et al. [10] put forward an energy-aware routing algorithm, called Energy-Saving Vector-Based Protocol (ES-VBF). The main purpose of this routing protocol is saving energy. ES-VBF takes both residual energy and localization-based information into consideration while calculating the desirability factor as in (3), which allows nodes to weigh the benefit for forwarding packets. The ES-VBF algorithm modifies the calculation of the desirability factor of (1) for VBF protocol to be calculated if the node residual energy is smaller than 60% of initial energy as:

$$\alpha = 0.5 \times \left( 1 - \frac{energy}{initialenergy} \right) + \left( \frac{P}{W} \right) + \left( \frac{R - d \times \cos \theta}{R} \right) \quad (3)$$

where *energy* is the residual energy of nodes and *initialenergy* is the initial energy of nodes. By simulation results in [10], it is shown that the performance is promising in balancing network energy consumption and packet

reception ratio. This means that the ES-VBF protocol saves energy in an efficient manner. At the same time, there is a small falling in packet reception ratio, which needs further research aiming at finding a better solution not only reducing energy consumption but also achieving high packet reception ratio.

### III. CLUSTERING VBF ROUTING ALGORITHM: THE PROPOSED ALGORITHM

In this paper, we propose an algorithm for UWSNs which we call a Clustering Vector-Based Forwarding algorithm (CVBF). The objective of the proposed routing algorithm is to reduce energy consumption, increase the packet delivery ratio, and decrease the average end-to-end delay. This is emphasized through comparison with VBF, HH-VBF, VBVA, and ES-VBF routing protocols.

According to our approach, the whole network is divided into a predefined number of clusters. All sensor nodes are assigned to the clusters on the basis of their geographic location, and then one node at the top of each cluster is selected as a *virtual sink* for that cluster. The rest of nodes in each cluster transmit the data packets to their respective cluster virtual sink. The routing inside each cluster follows the VBF routing protocol discussed in Section II. This implies that the concept of using one virtual routing pipe for all network nodes in VBF is replaced by defining one virtual routing pipe for each cluster to forward the packets from any node in the cluster to its virtual sink in that cluster. We assume that the routing pipe radius is equal to the transmission range of a node. Each intermediate node in any cluster selects the next hop to a node inside its cluster. In this way, the network will have many virtual routing pipes, one pipe per cluster, which guarantees forwarding the packets in the upward direction instead of forwarding the packets widely across the network nodes in the VBF algorithm. It is well expected that this will decrease the average end-to-end delay node and reduce the number of hops to reach the virtual sink node which will enhance the network performance. In addition, CVBF avoid voids in the network because each node belongs to a specific cluster.

Also, if a small number of nodes are available in the neighborhood, CVBF can still find a data delivery path. After receiving the data packets from cluster sensor nodes, cluster virtual sinks perform an aggregation function on the received data, and transmit them towards the main sink node using single-hop routing. Cluster virtual sink nodes are responsible for coordinating their cluster members and communicating with the main sink node.

The proposed algorithm is stated in the following steps:

#### Step1: Clustering the Nodes

This step involves dividing the network into groups of nodes according to their geographic location producing non-overlapping clusters excluding the main network sink which is allocated on the water surface. The following values are given: the network space  $X \times Y \times Z$ , node transmission range, routing pipe width, and the node speed. We divide the network space into equal space volumes; in the form of

cuboids (each cuboid has four rectangular sides and two square ones). The division is based on the values of  $X$  and  $Y$  coordinates, and the cluster width,  $cw$ , as shown in Fig. 3 (a).

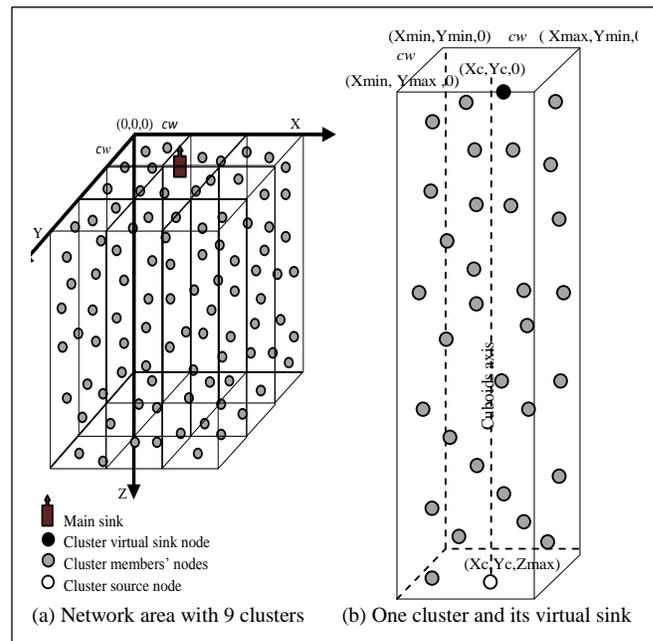


Figure 3. A CVBF network area: (a) Network area with 9 clusters, (b) One cluster and its virtual sink

Choosing the best number of clusters is proposed as:

$$N = \frac{X \times Y}{(cw)^2} \quad (4)$$

where  $X \times Y$  is the total surface area of the network and  $(cw)^2$  is the area of the cluster surface. The cluster width is thus calculated as:

$$cw = \sqrt{\frac{X \times Y}{N}} \quad (5)$$

It is given that the surface area is square; therefore, we choose  $N$  as a number raised to the power of two:  $2^2$ ,  $3^2$ ,  $4^2$ , or  $5^2$ . Here, we choose  $N$  that gives the value of  $cw$  as near as possible to  $\sqrt{2}R$  in order to make sure that the virtual pipe of the cluster includes all the nodes inside that cluster, as shown in Fig. 4.

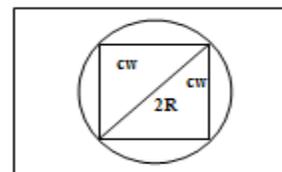


Figure 4. A horizontal section of cluster virtual pipe.

As an extreme case, if we choose one cluster ( $N=1$ ) only, then our algorithm reduces to the VBF protocol. In other words, our algorithm is a good generalization to VBF protocol.

*Step2: Selecting the Cluster Virtual Sink*

For each cluster (cuboid) which has a space volume  $cw \times cw \times Z$ , we choose the nearest node to the main sink to be a cluster virtual sink. As shown in Fig. 3 (b), the surface corner coordinates of the cluster are:  $(X_{min}, Y_{min}, 0)$ ,  $(X_{max}, Y_{min}, 0)$ ,  $(X_{min}, Y_{max}, 0)$ , and  $(X_{max}, Y_{max}, 0)$ .

All other nodes can send data to their corresponding virtual sink following the mechanism of VBF and depending on the value of its desirableness factor  $\alpha$ . If more than one node have the same depth position, we choose the nearest node to the cuboid axis, in which its surface point coordinates is the point  $(X_c, Y_c, 0)$ . The source node of the cluster is fixed at the position  $(X_c, Y_c, Z_{max})$ .

*Step3: Calculating the Cluster's Maintenance Time*

This step takes into consideration the node mobility that affects network topology and performance, thus necessitating a cluster maintenance algorithm. For a correct network operation, the maintenance algorithm should be executed simultaneously in all clusters. In this step, we propose a suitable periodical time which we call *maintenance time*,  $T_m$ . This time is enough to move a node from its cluster to another cluster according to speed and maximum distance of the node. Each node in the cluster checks its belonging to that cluster after the periodical time  $T_m$ . If a node belonging to a cluster moves away from that cluster, it naturally has two choices. The first choice is to enter another neighboring cluster, and so we transfer this node from the old cluster to the new cluster. The second choice is that it exits from all the network space, and so we leave this node in the old cluster. To calculate  $T_m$ , we divide the known maximum distance of a node movement,  $d_{max}$ , by the current speed of the node,  $S$ :

$$T_m = \frac{d_{max}}{S} \quad (6)$$

In other words, all the nodes with positions near the cluster boundaries are prone to exit from their own cluster and enter to other clusters. To avoid exiting a node from the network space, we suggest to carefully choose the node positions to be far from the network space boundaries.

The proposed algorithm is summarized in the Pseudocode of Fig.5:

Pseudocode
<p>Step 1: Clustering the nodes</p> <ol style="list-style-type: none"> <li>1. Given network space <math>X \times Y \times Z</math> and node transmission range <math>R</math>.</li> <li>2. Calculate <math>cw = \sqrt{(X \times Y)/N}</math> where <math>N=2^2, 3^2, 4^2, \text{ or } 5^2</math>, and cluster space = <math>cw \times cw \times Z</math></li> <li>3. For each cluster, <math>i=1</math> to <math>N</math> with step 1           <ol style="list-style-type: none"> <li>3.1. Given <math>X_i</math> takes values from <math>X_{min}</math> to <math>X_{max}</math>, and <math>Y_i</math> takes values from <math>Y_{min}</math> to <math>Y_{max}</math></li> <li>3.2. <math>K_i</math> is the number of nodes in the space <math>X_i \times Y_i \times Z</math></li> </ol> </li> </ol> <p>Step 2: Selecting the cluster virtual sink</p> <ol style="list-style-type: none"> <li>1. Sort the nodes, <math>K_i</math>, according to the values of their <math>Z</math> coordinate to determine the minimum value of <math>Z</math> and call it <math>Z_{min}</math>.</li> </ol>

2. Count the number of nodes,  $K_{iz}$ , in which their  $Z$  coordinate equal  $Z_{min}$ .
3. If  $(K_{iz}=1)$   
Then  
This node is the *virtual sink* of cluster  $i$   
Else  
  - (a) Calculate  $X_c = [(X_{max} - X_{min})/2]$ , and  $Y_c = [(Y_{max} - Y_{min})/2]$  to get the points of the cluster axis,  $(X_c, Y_c, Z_{min})$  and  $(X_c, Y_c, Z)$
  - (b) Calculate the nearest node to point  $(X_c, Y_c, Z_{min})$  from the given  $K_{iz}$  to become *virtual sink* of cluster  $i$

## Step 3: Calculating the cluster's maintenance time

1. If a node is near to the cluster axis and its mobility does not cause exit this node from that cluster or If a node moves outside the whole network space  
Then  
This node is still belongs to its original cluster  
Else  
If a nodes exits from the cluster  
Then  
  - (a) Given the node speed,  $S$ , and the maximum distance of any node,  $d_{max}$ ,
  - (b) Calculate  $T_m = d_{max}/S$
  - (c) For  $J=0$  to Simulation time with step  $T_m$   
For each node in the cluster  $i$  and has coordinates  $(X_i, Y_i, Z)$   
If  $(X_{min} \geq X_i \geq X_{max}$  and  $Y_{min} \geq Y_i \geq Y_{max})$   
Then  
This node is still in the cluster  
Else  
Remove this node from cluster  $i$  and enter it to the suitable neighboring cluster
2. All the nodes in cluster  $i$  forward the packets to its virtual sink following the mechanism of VBF routing algorithm
3. All the virtual sinks forward the packets to the main sink

Figure 5. Pseudocode of the proposed routing protocol CVBF.

## IV. PERFORMANCE EVALUATION

Performance is quantified through measures of energy consumption, packet delivery ratio, and average end-to-end delay [10]. The success rate is the ratio of the number of packets successfully received by the sink to the number of packets generated by the source. The energy consumption is the total energy consumed by the sensor network nodes. The average delay is the average end-to-end delay for each packet received by the sink.

Simulation is performed by the underwater package Aqua-Sim of ns-2 [13][14]. In all our simulations, we set the parameters similar to UWM1000 LinkQuest Underwater Acoustic Modem [15]. The bit rate is 10 kbps, and the transmission range  $R$  is 100 m. The energy consumption on

the sending mode, receiving mode, and idle mode are 0.6J, 0.3J, and 0.01J, respectively. The data packet size is 76 bytes and control packet is 32 bytes. The pipe radius in each cluster is 100 m. In all simulation experiments, sensor nodes are randomly distributed in a space volume of 600 m × 600 m × 600 m. They can move in a two-dimensional space, i.e., in the X-Y plane (the most common mobility pattern in underwater applications) with the medium node speed  $S$  in the range (2m/s-5m/s). The maximum distance of a node movement  $d_{max}$  is 5m. The number of clusters  $N$  used is 9 clusters (it is found in our experiments that the number  $N=9$  give better performance than that for  $N=1, 4, 16, 25$ ). The cluster width  $c_w$  is  $\sqrt{(600 \times 600)/9} = 200$  m. We have one data source, one main sink, and 9 virtual sinks. For each setting, the results are averaged over 30 runs with a randomly generated topology. The total simulation time for each run is 1000 s. The simulation results are plotted in Figures 6, 7, and 8.

Fig. 6 depicts the total energy consumption as the number of sensor nodes varies. The energy consumption increases with the number of nodes since more nodes are involved in packet forwarding. On the other hand, this figure shows that the energy consumption for the proposed algorithm is less than that in VBF and HH-VBF routing protocol only on dense networks, when the number of nodes is greater than 300 nodes, indicating that the CVBF algorithm can save more energy with high node density, as shown in Table I, extracted for Fig. 6.

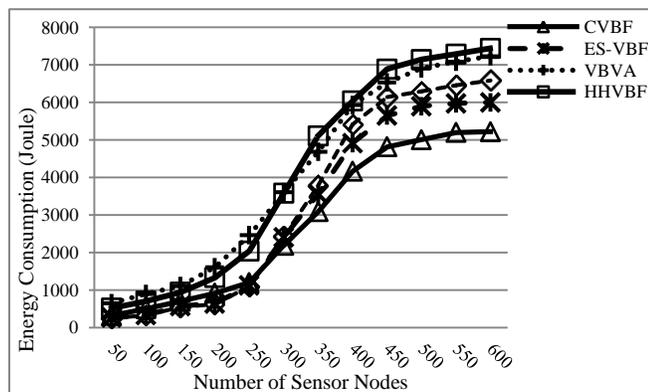


Figure 6. CVBF energy consumption vs. number of sensor nodes.

TABLE I. REDUCTION IN TOTAL ENERGY CONSUMPTION

No. of Nodes	Total Energy Consumption(Joule)					Reduction percentage
	VBF	HH-VBF	VBVA	ES-VBF	Proposed CVBF	
50	250.12	515.37	646.71	260.72	<b>320.65</b>	-28%
150	550.53	939.98	1103.65	570.23	<b>698.97</b>	-26%
300	2424.16	3578.32	3604	2400.91	<b>2190.86</b>	8.7%
450	6142.81	6890.6	6532.1	5646.4	<b>4816.87</b>	14%
600	6589.39	7440.6	7231.9	6000.03	<b>5220.02</b>	13%

Fig. 7 shows the packet delivery ratio with the number of sensor nodes. It is seen that the packet delivery ratio increases with the increase of the number of nodes. When

more than 200 nodes are deployed in the space, the packet delivery ratio remains above 90% for both ES-VBF routing protocol and CVBF algorithm. Table II, extracted from Fig. 7, shows that our algorithm gives better results in packet delivery ratio than VBF, HHVBF, VBVA, and ES-VBF protocols.

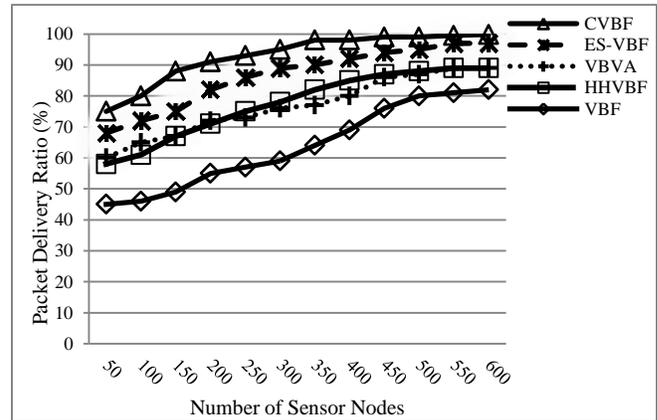


Figure 7. CVBF packet delivery ratio vs. number of sensor nodes.

TABLE II. INCREASE IN PACKET DELIVERY RATIO

No. of Nodes	Packet Delivery Ratio (%)					Increasing percentage
	VBF	HH-VBF	VBVA	ES-VBF	Proposed CVBF	
50	45	58	60	68	<b>75</b>	10%
150	49	67	67	75	<b>88</b>	17%
300	59	78	76	89	<b>95</b>	6.7%
450	76	87	86	94	<b>99</b>	5.3%
600	82	89	89	99	<b>99.8</b>	0.08%

Fig. 8 describes the average end-to-end delay with the number of sensor nodes. It is seen that the average end-to-end delay decreases with the increase of node density in the network. When the number of sensor nodes increases, the paths from the source to the sink are closer to the optimal path ( $\alpha=0$ ); therefore, the average end-to-end delay decreases, as shown in Table III, extracted from Fig. 8.

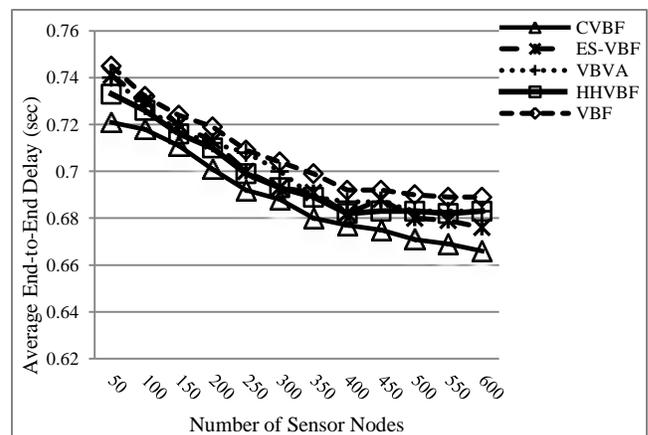


Figure 8. CVBF average end-to-end delay vs. number of sensor nodes.

TABLE III. REDUCTION IN AVERAGE END-TO-END DELAY

No. of Nodes	Average End-to-End Delay(sec)					
	VBF	HH-VBF	VBVA	ES-VBF	Proposed CVBF	Reduction percentage
50	0.745	0.733	0.74	0.741	<b>0.721</b>	1.6%
150	0.724	0.716	0.717	0.72	<b>0.711</b>	0.7%
300	0.704	0.693	0.7	0.693	<b>0.688</b>	0.7%
450	0.692	0.683	0.687	0.688	<b>0.675</b>	1.2%
600	0.689	0.683	0.683	0.676	<b>0.666</b>	1.5%

We evaluate the performance of CVBF under various network scenarios. The simulation results show that CVBF significantly exhibits a better performance than VBF, HH-VBF, VBVA, and ES-VBF protocols since it has: lower energy consumption, higher packet delivery ratio, and lower average end-to-end delay.

Calculating the cluster width  $c_w$  depends on two parameters: the surface area of the network  $X \times Y$  and choosing the number of clusters  $N$ . We choose a value of  $N$  for which the cluster width is nearest to the value of  $\sqrt{2}R$ . We conclude this after examining different values of  $N$ . This is because each node can transmit the data packets only to the neighbors allocated in its transmission range.

## V. CONCLUSIONS

In this paper, we propose a clustering vector-based forwarding algorithm to improve the performance of the location-based routing protocol in underwater wireless sensor networks. In the proposed approach, the space area of the network is divided into clusters where one virtual sink is assigned to each cluster. Choosing the number of clusters depends on the value of the network surface area and the transmission range of the sensor node. The nodes inside each cluster are allowed to communicate with themselves following the concept of VBF protocol only to reach its virtual sink node, which sends the packets to the main sink node in the network. Due to node mobility, some nodes may move outside their cluster and enter another cluster. Therefore, we check the node position periodically as a maintenance step to allocate each node to its suitable cluster.

Simulation results demonstrate that the proposed algorithm efficiently reduces energy consumption especially in dense networks, increases the packet delivery ratio especially in sparse networks, and decreases the average end-to-end delay in both sparse and dense networks, in comparison with the four routing algorithms VBF, HH-VBF, VBVA, and ES-VBF. It is interesting to note that our multiple-cluster algorithm is a good generalization to the VBF protocol. The VBF results from our algorithm by adopting the special case of single-cluster manipulation.

## ACKNOWLEDGMENT

We would like to thank the (anonymous) reviewers for their helpful comments.

## REFERENCES

- [1] E. M. Sozer, M. Stojanovic, and J. G. Proakis, "Underwater Acoustic Networks," IEEE Journal of ocean engineering, vol. 25, Jan. 2000, pp. 72-83.
- [2] M. Ayaz, I. Baig, A. Abdullah, and I. Faye, "A Survey on Routing Techniques in Underwater Wireless Sensor Networks," Journal of Network and Computer Applications, vol. 3, no. 4, 2011, pp. 1908-1927.
- [3] M. Ayaz and A. Abdullah, "Underwater Wireless Sensor Networks: Routing Issues and Future Challenges," Proc. of the 7th international conference on advances in mobile computing and multimedia. ACM, Malaysia, 2009, pp. 370-375.
- [4] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater Acoustic Sensor Networks: Research Challenges," Ad Hoc Networks (Elsevier), vol. 3, no. 3, March. 2005, pp. 257-279.
- [5] D. Pompili, T. Melodia, and I. F. Akyildiz, "Distributed Routing Algorithms for Underwater Acoustic Sensor Networks," IEEE Trans. On wireless communications, vol. 9, no. 9, Sept. 2010, pp. 2934-3944.
- [6] P. Xie, J.-H. Cui, and L. Lao, "VBF: Vector-based Forwarding Protocol for Underwater Sensor Networks," International conference on networking (IFIP networking), 2006, pp. 1-20.
- [7] N. Nicolaou, A. See, P. Xie, J.-H. Cui, and D. Maggiorini, "Improving the Robustness of Location-based Routing for Underwater Sensor Networks," Proc. Of the OCEANS'07, Europe, June 2007, pp. 1-6.
- [8] A. Sharma and H. A. Gaafar, "A Survey on Routing Protocols for Underwater Sensor Networks," International Journal of Computer Science & Communication Networks, vol. 2, no. 1, 2012, pp. 74-82.
- [9] P. Xie, Z. Zhou, Z. Peng, J.-H. Cui, and Z. Shi, "Void Avoidance in Three-dimensional Mobile Underwater Sensor Networks," Proc. of the 4th international conference of wireless algorithms, system, and applications (WASA 2009), USA, August 2009, pp. 305-314.
- [10] W. Bo, L. Yong-mei, and J. Zhigang, "ES-VBF: An Energy Saving Routing Protocol," Proc. of the 2012 International Conference on Information Technology and Software Engineering, 2012, pp. 87-97.
- [11] E. A. Carlson, P.P. Beaujean and E. An., "Location-aware Routing Protocol for Underwater Acoustic Networks," Proc. of the OCEANS, 2006, pp. 1-6.
- [12] R. Thumpi, R.B. Manjula, and S.M. Sunilkumar, "A Survey on Routing Protocols for Underwater Acoustic Sensor Networks," International journal of recent technology and engineering (IJRTE), vol. 2, May 2013, pp. 170-175.
- [13] NS-2: Network Simulator, <http://www.isi.edu/nsnam/ns>, last access: 30<sup>th</sup> April 2014.
- [14] P. Xie, et al., "Aqua-sim: A Ns-2 based Simulator for Underwater Sensor Networks," Proc. of IEEE/MTS OCEANS, 2009, pp. 1-7.
- [15] LinkQuest, <http://www.link-quest.com>, last access: 30<sup>th</sup> April 2014

# On-Demand Data Collection in Sparse Underwater Acoustic Sensor Networks Using Mobile Elements

Jalaja M.J. and Lillykutty Jacob

Department of Electronics and Communication Engineering

National Institute of Technology, Calicut, India

Emails: {jalaja@rit.ac.in, lilly@nitc.ac.in}

**Abstract**—Underwater Wireless Sensor Networks (UWSNs) is a group of sensors and underwater vehicles, networked via acoustic links, that perform collaborative tasks and enable a wide range of aquatic applications. Due to hostile environment, resource constraints and peculiarities of the underlying physical layer technology, providing energy-efficient data collection in a sparse UWSN is a challenging problem. We consider mobility-assisted routing technique for enabling connectivity and improving the energy efficiency of sparse UWSN, considering it as a Delay/Disruption Tolerant Network (DTN) or Intermittently Connected Network (ICN). We use analytical models to investigate the performance of the data collection scheme. Based on the result that the DTN scheme improves energy efficiency and Packet Delivery Ratio (PDR) at the cost of increased message latency, we investigate techniques to improve the delay performance. The effects of using multiple mobile elements for data collection and priority-polling based on traffic class and data generation rate are investigated. The analytical results are validated through extensive simulations. The results show that our model for data collection in sparse UWSNs can effectively capture the underwater acoustic network conditions. Also, the improved DTN framework shows superior performance in terms of energy efficiency and network connectivity over ad hoc multihop network, and in terms of message latency and fairness over simple polling-based DTN framework.

**Keywords**—Underwater Sensor Networks; Delay Tolerant Network; Mobile Sink; Priority Polling; Energy Efficiency; Fairness.

## I. INTRODUCTION

Underwater Wireless Sensor Networks (UWSNs) have emerged as powerful systems for providing autonomous support for several activities like oceanographic data collection, marine surveillance, disaster prediction, assisted navigation etc. Acoustic communication, with its associated pros and cons, is the underlying physical layer technology used in UWSNs. Features like high latency, low bandwidth, high error probability and 3-dimensional deployment make the UWSNs significantly different from terrestrial WSNs [1]. The energy saving/efficiency is a critical issue for UWSN because of the high cost of deploying and/or re-deploying underwater equipment. Underwater sensors are expensive, partially because of their more complex transceivers and the ocean area that needs to be sensed is quite large. Hence, UWSN deployment can be much sparser compared with terrestrial WSNs. Due to sparse deployment, harsh environment, node mobility and resource limitations, the network can be easily partitioned and a contemporaneous path may not exist between any

two nodes. This results in sparse UWSNs that need to be treated as Intermittently Connected Networks (ICN) or Delay / Disruption Tolerant Networks (DTN) [2]. DTNs are characterised by frequent partitions and potentially long message delivery delays. Such networks may never have an end-to-end contemporaneous path and traditional routing protocols are not practical since packets will be dropped when no routes are available.

The primary objective of DTN routing is to provide eventual delivery of data, rather than optimizing some routing metric, say message latency. In energy-constrained underwater sensors, for certain delay-tolerant applications like environmental sensing or continuous monitoring, enhanced network lifetime will be more important than message delay. Enabling reliable and energy-efficient data collection in resource-constrained sparse UWSNs is a challenging problem that requires specialized routing approaches and QoS metrics. Conventional DTN approaches like multipath routing are resource-hungry and hence not suitable for resource-constrained underwater applications.

The three main approaches used for data collection in wireless sensor networks, in general, are [3]: (i) Base Station (BS) approach which uses direct communication between the source and the sink; (ii) Ad hoc network which uses a multi-hop path from the source to the sink; and (iii) Mobility assisted routing which makes use of a mobile sink or mobile relays for data collection. The first approach provides fast delivery, but suffers from reduced life time of sensors due to the increased requirement of communication energy. The ad hoc network provides medium delay and medium power requirement, but suffers from the ‘hot spot’ problem and the necessity for an end-to-end contemporaneous path. Mobility assisted routing approach supports the DTN concept, reduces transmit power consumption, and eliminates the relaying overhead. However, due to the limited travel speed of the mobile elements, data collection latency will be large, but such large latency may be acceptable in certain environmental sensing applications which are not time-critical. Typical example of such an application is the continuous monitoring and recording of the behaviour of underwater plates in tectonics, for later scientific analysis. Providing support for delay-sensitive applications like pollution monitoring and earthquake prediction, and ensuring fairness among different traffic classes using energy-efficient mobility-assisted routing in sparse UWSNs is the focus of this paper.

We start with a basic DTN framework for energy efficient data collection in sparse underwater sensor networks using a mobile sink; and then augment it with techniques to improve its data collection performance by introducing priority and employing multiple data collectors. Analytical results for energy efficiency, packet delivery ratio, message latency, and sensor buffer occupancy are presented. The analytical results are validated using our own simulation model developed in Aqua-Sim [4], an NS-2 [5] based network simulator, developed by the University of Connecticut. A brief review of the related work is given in Section II. The system model is presented in Section III. The expressions used for analytical results are developed in Section IV. Section V discusses the analytical and simulation results. The paper is concluded in Section VI.

## II. RELATED WORK

Several routing protocols have been developed for underwater sensor networks, most of them suitable only for connected networks. Vector Based Forwarding (VBF) is a typical geographical routing protocol and Hop-by-hop Vector-based forwarding (HH-VBF) [6] is its more energy-efficient version, better suited for sparse networks. Both VBF and HH-VBF do not support mobility-assisted data collection and they require the network to be connected. Recently, considerable effort has been devoted to developing architectures and routing algorithms for DTNs and routing in DTNs is investigated by Jain et al. [7]. Guo et al. have proposed an adaptive routing protocol for UWSNs, considering it as a DTN [8]. Shah et al. [3] have presented a three-tier architecture based on mobility to address the problem of energy efficient data collection in a terrestrial sensor network. The same architecture with an enhanced analytical model has been presented by Jain et al. [9]. Energy analysis of routing protocols for UWSNs is presented by Domingo [10] and by Zorzi et al. [11]. An M/G/1 queueing model is used by He et al. [12] for mobility-assisted routing, proposed for reducing and balancing the energy consumption of sensor nodes. The use of controlled mobility for low energy embedded networks has been discussed by Arun et al. [13]. AUV-aided routing for UWSNs is discussed by Yoon et al. [14] and Hollinger et al. [15]. Polling-based scheduling in body sensor networks has been discussed by Motoyama [16] and the usage of message ferries in ad hoc networks is considered by Kavitha et al. [17].

The development of routing protocols for dense UWSNs and the adaptation of DTN approaches for terrestrial sensor networks have already been addressed, but the energy-efficient data collection in resource-constrained sparse/disconnected UWSNs has not been adequately investigated. Also, an analytical framework and the simulation environment for evaluating the performance metrics of data collection in UWSNs will be useful for designing application-oriented networks. In this paper, we propose a mobility-assisted DTN scheme for data collection in sparse UWSNs and propose techniques for providing support for delay-sensitive applications, by employing multiple data collectors and introducing priority.

## III. SYSTEM MODEL

We consider large and sparse underwater sensor networks with possibly disconnected components and with mobile elements used for data collection. The static sensors monitor the underwater surroundings, generate data and store it in the sensor buffer. They have limited non-rechargeable battery power and they can communicate using acoustic links. Sensors' bulk data communications are limited to transferring data to a nearby mobile collector (MC), so as to reduce energy consumption. Mobile Collectors are mobile entities with large processing and storage capacity, renewable power, and the ability to communicate with static sensors, BS and other MCs (if any). As an MC moves in close proximity to (i.e., within transmission range of) a static sensor, the sensor's data is transferred to the MC and buffered there for further processing. The mobility of the MC can be either random or controlled.

The static sensors can request the service of the MC by sending service request messages to the base station (BS) using direct or ad hoc multi-hop communication. The service request packet is assumed to be very short compared to data packets and the former will contain location information of the node, priority of application, and any other relevant information like data rate or the delay-sensitivity of request. The BS will collect the requests and based on the system load and the delay requirements, it can decide the number of MCs needed and the sequence of visiting the nodes by each MC. Accordingly, BS will create one or more visit tables specifying the order of visiting the nodes and schedule the required number of MCs with a unique visit table assigned to each one of it. Each MC will visit the sensor, collect the data generated and buffered so far, and proceed towards the next node in the table and this process is repeated. After one cycle is completed, it may visit the BS and collect the updated visit table if it has been modified by the BS during that cycle. The data is assumed to have been successfully delivered once it has been collected by the MC.

*Underwater Channel:* If a tone of frequency  $f$  and power  $P$  is transmitted over a distance  $l$ , the received signal power will be  $P/A(l, f)$ , where the attenuation factor  $A(l, f)$  is the sum of absorption loss and spreading loss. At shorter ranges, spreading loss plays a proportionally larger role compared with absorption loss. Spreading loss is frequency-independent, but depends on the geometry. The SNR of an emitted underwater signal at the receiver is expressed by the passive sonar equation [18] and the transmission loss or the attenuation factor  $A(l, f)$  of an underwater acoustic channel for a distance  $l$  and frequency  $f$  is given by Eqn. 1 as [18]:

$$10 \log A(l, f) = k \cdot 10 \log l + l \cdot 10 \log a(f) \quad (1)$$

where the first term is the spreading loss and the second term is the absorption loss. The spreading coefficient  $k = 1$  for cylindrical spreading (shallow water scenario) and  $k = 2$  for spherical case (deep water scenario). The absorption coefficient can be expressed empirically, using the Thorps formula which gives  $a(f)$  in dB/km for  $f$  in kHz as Thorp's formula [18] is

used to express the absorption coefficient as :

$$10 \log a(f) = \frac{0.11f^2}{1+f^2} + \frac{44f^2}{4100+f^2} + \frac{2.75f^2}{10^4} + 0.003 \quad (2)$$

The absorption coefficient increases rapidly with frequency (typical values being 50 dB/km at 200 kHz and 320 dB/km at 1 MHz, thus imposing a limit on the maximal usable frequency for an acoustic link of a given distance  $l$  (which may typically vary from a few metres to a few kilometres).

#### IV. ANALYTICAL STUDY

In this section, we develop the necessary analytical expressions, the numerical results of which are compared with the simulation results in Section V.

##### A. Energy Efficiency

One important motivation for employing a mobile sink is that it increases the lifetime of the network by balancing the energy consumption of the sensor nodes. The energy consumption of the static nodes alone is considered, since the mobile node is assumed to be rechargeable or having much higher initial energy compared to the static sensors. The energy consumed by the static sensor nodes for sensing and processing are negligible compared with that for underwater acoustic data transmission, and hence we consider the energy consumption for data transmission only. For a given target signal-to-noise ratio  $SNR_{tgt}$  at receiver, available bandwidth  $B(l)$ , and noise power spectral density  $N(f)$ , the required transmit power  $P_t(l)$  can be expressed as a function of the transmitter-receiver distance  $l$  [11]. If  $P_r$  is the receive power,  $L$  is the packet size in bits,  $M$  is the number of packets transferred from the source node to the destination and  $\alpha$  is the bandwidth efficiency of modulation, the energy consumption for the single hop data transfer becomes

$$E_{hop}(l) = \frac{M(P_r + P_t^{el}(l))L}{\alpha B(l)} \quad (3)$$

where  $P_t^{el}(l)$  is the electrical power (in watts) corresponding to  $P_t(l)$  in dB re  $\mu\text{Pa}$ . Compared to  $P_r$ ,  $P_t^{el}$  is very large and hence its contribution to the energy consumption of sensor nodes is significant.

In order to assess the energy efficiency of the MC-based DTN model, let us compare the energy overhead associated with transferring one packet from the sensor to the BS using the ad hoc multi-hop approach and the *store-carry-and-forward* DTN approach.

Assuming  $N$  static sensor nodes randomly and uniformly deployed over a circular area  $A$  of radius  $R$  as in [13], we can calculate the minimum energy requirement of each node for transferring one packet generated by each node to the sink at the centre of the circular area, in the ad hoc multi-hop network.

If every static node with a transmission range  $r$  and located in the  $k$ th annulus of the circular area generates one packet, then the minimum number of transmissions due to packets originated from the  $k$ th annulus is  $MinTx(k) = N \frac{A(k)}{A} k$ ,

where  $A(k)$  is the area of the  $k^{th}$  annulus and  $k = 1$  for the innermost annulus. In the mobility-assisted data collection, irrespective of the position of the nodes, each static node transmits only the packets generated by it. Instead, in the case of multi-hop architecture, if every node generates 1 packet each, for a large value of  $N$ , on an average, the number of receptions and transmissions to be undertaken by a node in annulus  $k$  will be, respectively,  $NodeRx(k) = \frac{A(k+1)}{A(k)} NodeTx(k+1)$  and  $NodeTx(k) = 1 + \frac{A(k+1)}{A(k)} NodeTx(k+1)$ , except for the outermost annulus ( $k = \lceil \frac{R}{r} \rceil$ ) where the corresponding values are 0 and 1.

The above analysis shows the increased relaying overhead of a sensor node with its proximity to the sink. If we define the *Energy Overhead Factor* (EOF) of a node as the ratio of the total number of transmissions from the node to the number of transmissions corresponding to the packets originated at that node, it is seen that all the sensor nodes have the same EOF (equal to 1 with an error-free channel) in MC-based scheme, while it is approximately equal to  $NodeTx(k)$  in multihop network. High *Energy Overhead Factor* implies low energy efficiency.

##### B. Data Collection Latency

A polling model is used to investigate the delay performance of MC-based data collection. In the basic polling model, a single server visits (or polls) the queues in a cyclic order and after completing a visit to queue  $i$ , the server incurs a switch over period or *walk time* [19]. The period during which the server continuously serves queue  $i$  is called a *service period* of queue  $i$  and the preceding period is called the *switch over period* of queue  $i$ . Different service policies can be employed, out of which the *Exhaustive* service scheme is the optimal. Mobile Collector and the static sensor buffers in our model correspond to the single server and queues of the polling model, respectively. Travel time of the MC to move from one location to the next is modelled as the *walk time* and the time spent at each location to transfer data from the near by sensor's buffer to the MC is modelled as the *service time*.

Assuming Poisson arrival of packets at rate  $\lambda$  at each sensor buffer, the offered load is given by  $\rho = N\lambda\bar{X}$ , where  $\bar{X}$  is the mean message service time. For system stability,  $\rho$  should be less than 1. If the mean of the total walk time is denoted by  $R$ , the mean cycle time of the MC is given by

$$E[C] = \frac{R}{1-\rho} \quad (4)$$

Let  $\overline{X^2}$  denote the second moment of the packet transfer time and the MC travel time between two consecutive locations be a random variable with mean and variance  $\overline{W}$  and  $\overline{W^2}$ , respectively. Under the assumption of symmetric queues and *exhaustive* service, the mean waiting time of the packet in the sensor buffer before the MC approaches it for data transfer can be obtained as:

$$W_q = \frac{\overline{W}^2}{2\overline{W}} + \frac{N\lambda\overline{X}^2 + \overline{W}(N - \rho)}{2(1 - \rho)} \quad (5)$$

Assuming that the static nodes are uniformly distributed in the network, their locations can be treated as random points in the square sensing field. The probability density function of the distance between two arbitrary points in a unit square is given by [12]  $f_D(d) =$

$$\begin{cases} 2d(\pi - 4d + d^2) & 0 \leq d \leq 1 \\ 2d[2\sin^{-1}(\frac{1}{d}) - 2\sin^{-1}\sqrt{1 - \frac{1}{d^2}} \\ + 4\sqrt{d^2 - 1} - d^2 - 2] & 1 \leq d \leq \sqrt{2} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

From this, if the MC moves at a constant velocity  $V$ , the mean and the variance of the MC travel time between two arbitrary points in a unit square area can be obtained as  $0.4555/V$  and  $3.95/V^2$ , respectively.

The expected response time of a message, buffer size, and number of messages in the system (in queue and in service) are  $\overline{X} + W_q$ ,  $W_q\lambda$ , and  $(\overline{X} + W_q)\lambda$ , respectively. Using the parameters of data generation, data transfer, and MC mobility, the delay performance of our system model is evaluated. The controlled mobility of the MC gives better performance compared to random mobility and hence the former is recommended if the deployment permits.

The delay performance of the MC-based DTN scheme with a single mobile element is not at all comparable with that of ad hoc multihop network (of the order of several minutes for the former, while a few seconds for the latter). Correspondingly, the buffer requirement of static sensors is negligible in an ad hoc network, while it is considerably high in the MC-based scheme.

### C. Packet Delivery Ratio (PDR)

In the *exhaustive* service policy of polling scheme, all the data generated at one sensor in one *cycle time* is transferred in one visit of the MC. The mean number of packets generated in 1 cycle time =  $\lambda E[C]$ . Assuming sufficiently large buffer space to avoid buffer overflow, ideal channel, and no MC failures, the PDR will be 1. But practically, there exists a probability that a node is not detected (or a *contact* does not occur) within a reasonable time period. In such situations, the significance of the data may be lost if the application is delay-sensitive, or the data itself may be lost due to buffer overflow.

### D. Performance Enhancement

To improve the delay and delivery performance of the basic DTN scheme with a single MC, two techniques can be employed: i) use of multiple mobile sinks or mobile collectors, and ii) priority polling. In the first technique, more than one mobile sink or mobile collectors are used, thus increasing the effective service rate, thereby reducing the message waiting time. In the second one, different priority is assigned to different nodes (based on data generation rate, traffic class

etc) and the order and/or frequency of polling or visiting the static sensor nodes is modified to account for the service requirement. In both cases, the scheduling of MC(s) should take into account the service demand, in terms of varying network load, meeting deadline, or ensuring fairness.

1) *Multiple Mobile Collectors*: In our basic polling model, there is only a single server, servicing a number of queues in a cyclic manner, with a non-zero switch-over time. When the number of mobile data collectors is increased, the model is converted to a Multi Server Multi Queue (MSMQ) system or *multi server polling model*, the exact analysis of which is not available. Assuming independent mobile collectors, symmetric Poisson-distributed data arrivals, independent and identically distributed *service times* and *walk times* and no server clustering, an approximate expression for the mean waiting time can be derived following the approach used in [20]. If  $S$  is the number of MCs, to get the mean message waiting time in the multiple MC case, the expression for mean waiting time in single MC case as given by Eqn. 5 can be modified by substituting  $\overline{X}/S$ ,  $\overline{X}^2/S^2$ ,  $\overline{W}/[S - (S - 1)\rho]$ , and  $\overline{W}^2/[S - (S - 1)\rho]^2$  in place of, respectively,  $\overline{X}$ ,  $\overline{X}^2$ ,  $\overline{W}$ , and  $\overline{W}^2$ . Thus the mean waiting time in the multiple MC situation becomes

$$W_q = \frac{\overline{W}^2}{2\overline{W}[S - (S - 1)\rho]} + \frac{N \left[ \frac{\lambda\overline{X}^2}{S} + \frac{\overline{W}(S - \lambda\overline{X})}{S - (S - 1)\rho} \right]}{2(S - N\lambda\overline{X})} \quad (7)$$

Compared to the basic single MC network, here the expected waiting time and the sensor buffer occupancy decrease with the number of servers  $S$ . Thus, the delay and delivery performance is improved by the use of multiple data collectors, while energy consumption and network lifetime are not affected, since the number of transmissions and the range of transmission are not changed by the use of more number of MCs.

2) *Priority Polling*: In practical situations, all the nodes may not be generating data at the same rate and hence the earlier assumption of symmetric queues may not be valid. When the data generation rates among the static sensor nodes vary considerably, it will be better to visit the nodes with higher arrival rates more frequently, rather than following the cyclic order. In cyclic polling, the server polls the queues in the order  $Q_1, Q_2, \dots, Q_N, Q_1, Q_2, \dots, Q_N, \dots$ . In *Periodic* polling, the server visits the queues in a fixed order specified by a *polling table* in which each queue occurs at least once [21].

Consider the single server polling model with the difference that the arrival rates at the queues are not equal, instead the packet arrival intensity at sensor  $i$  is  $\lambda_i$ ,  $i = 1, \dots, N$ . The offered load at sensor  $i$  is  $\rho_i = \lambda_i \overline{X}_i$ , where  $\overline{X}_i$  is the mean service time at sensor  $i$ . The total offered load in the network  $\rho = \sum_{i=1}^N \rho_i$ . The MC visits the sensors according to a periodic - not necessarily cyclic - polling scheme. The approach followed in [21] can be used to minimize the workload in the system and to ensure *fairness* among the sensors by using optimum visit frequencies. For *exhaustive* service, assuming  $W_i$  to be the switch-over time from queue  $i - 1$  to queue  $i$ , the visit

frequency at node  $i$  becomes

$$f_i^{exh} = \frac{\sqrt{\rho_i(1-\rho_i)/W_i}}{\sum_{j=1}^N \sqrt{\rho_j(1-\rho_j)/W_j}} \quad (8)$$

Now, all the nodes are not visited equally in a cycle, instead the nodes having more buffered data waiting for transmission (due to higher arrival rate) will be visited more often than those with less buffered data. Assume that sensor  $i$  is visited  $n_i$  times in a cycle of the MC and these visits are spread as evenly as possible. Considering the interval between two successive MC visits to a node  $i$  as a sub cycle, the mean residual time of a sub cycle of  $i$  will be

$$ERSC_i \propto \frac{E[C]}{n_i} \quad (9)$$

where  $E[C]$  is the mean time for one complete visit cycle of the MC according to the polling table. Now the mean waiting time at node  $i$  will be [21]:

$$(W_q)_i \propto (1-\rho_i) \frac{E[C]}{n_i} \quad (10)$$

which shows that the sensor nodes with high data generation rates (having high values of  $\rho_i$  and  $n_i$ ) get better treatment and majority of the generated packets get good treatment, in terms of waiting time and buffer requirement.

## V. ANALYTICAL AND SIMULATION RESULTS

Extensive simulations have been done to validate our analytical results using the NS-2 based network simulator for underwater applications, Aqua-Sim. It is an event-driven, object-oriented simulator written in C++ with an OTCL (Object-oriented Tool Command Language) interpreter as the front-end. We have incorporated in it, the DTN concepts of beaconing, *contact* discovery and *store-carry-and-forward* and the polling based (*exhaustive* service) data collection.

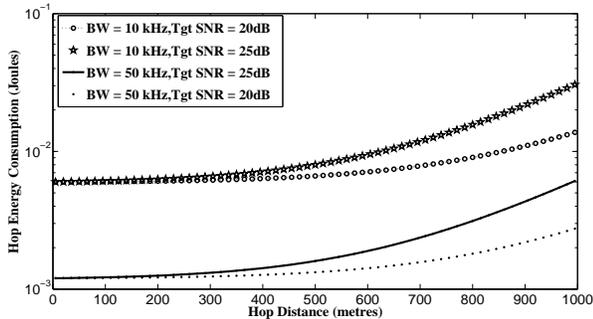


Figure 1. Hop Energy Consumption for varying hop length and bandwidth

Assuming tunable transmit power  $P_t$ , receive power  $P_r$  fixed at 0.075 W, and packet length  $L$  fixed to 400 bits [4], the effect of hop length, target SNR, and channel bandwidth on per-hop energy consumption as expressed by Eqn. 3 is plotted in Fig. 1 for shallow water environment. Decreasing the source to sink distance reduces the transmission loss and increasing

the bandwidth reduces the time required for transmission. Both situations lead to reduced transmit energy consumption, thus validating the suitability of short range communication in energy-constrained environments.

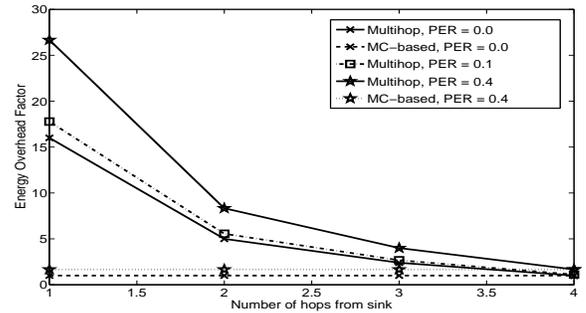


Figure 2. Transmit Energy Overhead of static sensor nodes with multi-hop and MC-based schemes for different PERs

Assuming static sensor nodes having transmission range 250m uniformly distributed in the area of radius 1000m, the variation of the *Energy Overhead Factor* (defined in Section IV A) with proximity to the sink, in multi-hop routing is illustrated in Fig. 2. Due to the increased relaying overhead, the nodes nearer to the sink will deplete their battery power soon. The impact of packet error rate (PER) due to non ideal channel is also shown in this figure. If we define the lifetime of a network as the timespan till the first node dies due to energy depletion, it is evident that the use of mobile elements for data collection leads to enhanced lifetime of the network due to reduced and balanced energy consumption among the sensor nodes.

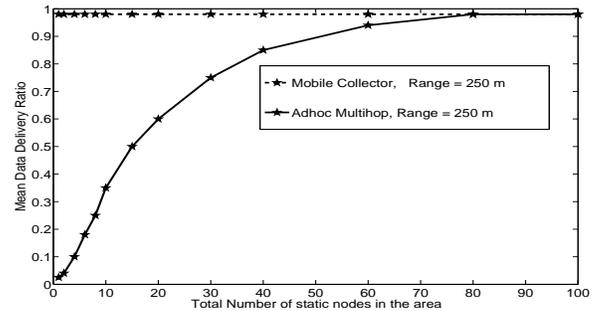


Figure 3. PDR with multi-hop and MC-based data collection

The variation of packet delivery ratio with node density is shown in Fig. 3. Assuming infinite buffer size and no communication errors, ideally the packet delivery ratio should be 1 for the DTN data collection scheme irrespective of the number of nodes in the network. For ad hoc multi-hop network, delivery ratio is very small for low node density due to end-to-end connectivity issues. As the node density is increased, PDR increases initially and finally reaches a maximum value. It then remains almost constant if only one node is transmitting, but starts reducing due to packet collisions

if multiple nodes are transmitting. For the DTN scheme, delivery ratio is independent of node density. Hence, it is the ideal one for sparse networks and heavy traffic environments, provided the network lifetime and successful data delivery are of prime concern and the application is not time-critical. If the sensors are not equipped with sufficient buffer space to avoid buffer overflow at high loads, packets are dropped and PDR is reduced.

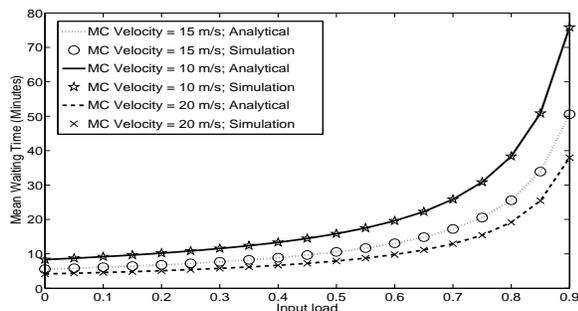


Figure 4. Variation of Mean Waiting Time

The mean waiting time for different values of data generation rate and different speeds of the single MC is plotted in Fig. 4, considering the controlled motion of the mobile sink in a square area of size  $1000m \times 1000m$  with 10 nodes randomly and uniformly distributed in this area. The sensors are equipped with sufficient buffer space so that packets are not lost due to buffer overflow. The mean waiting time increases with the packet arrival rate and decreases with the speed of the MC. Analytical and simulation results show close agreement, validating the suitability of our model. Fig. 5 shows the variation of the mean buffer occupancy with varying load and MS speeds for the same scenario. The buffer space requirement also increases with the input load and decreases with MC speed.

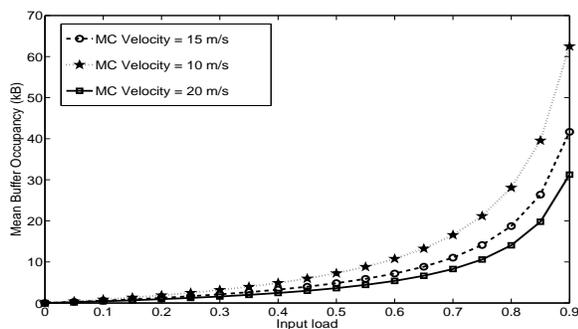


Figure 5. Variation of Mean Buffer Occupancy

Fixing the packet size to be 50 Bytes, and data rate 10 Kbps, the impact of the speed and number of MCs on the delay performance is also studied and plotted in Fig. 6. Since it not practical to have MC speeds above 20 m/s, use of multiple

MCs is to be adopted for heavy traffic environments, delay-sensitive applications, and very limited sensor buffer situations. Also, the performance gain obtained by using 3 MCs over 2 MCs is much less compared to that obtained by using 2 MCs over a single one.

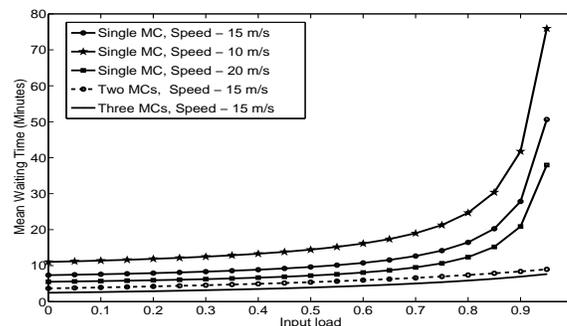


Figure 6. Mean Waiting Time with Multiple MCs

With priority polling, assuming 10 sensor nodes randomly and uniformly distributed in an area of size  $1000m \times 1000m$ , generating packets (of size 50 bytes) at four different rates, a single MC moving at 15 m/s, and having a data rate 10 kbps, Table I gives the visit frequency and the mean waiting time for different packet arrival rates. Based on the simulation for a fixed finite amount of time, the percentage of packets missed due to the MC not arriving in time is also noted. As

TABLE I. MEAN MESSAGE WAITING TIME AT DIFFERENT NODES

Arrival Rate (Pkts/min)	Visit Freq. (Percentage)	Waiting Time (Minutes)	Miss Ratio (Percentage)
0.01	1.53	22.37	71.25
0.1	5.10	17	57.42
1.0	16.58	12.43	13.10
2.0	23.7	11.47	1.01

the packet generation rate  $\lambda_i$  at node  $i$  increases, the input load  $\rho_i$ , the number of sub-cycles  $n_i$ , and the visit frequency  $f_i$  increase, while the mean waiting time  $(W_q)_i$  decreases. Due to the unequal visit frequency at different nodes, the percentage of packets collected by the MC within a finite simulation time is also not equal (more at high data rate nodes and less at low data rate nodes). Thus by reducing the unnecessary travels to the low data rate nodes, the overall system utilization is improved and majority of packets will be serviced within a reasonable waiting time.

## VI. CONCLUSION

The suitability of a mobility-assisted framework for energy-efficient data collection in sparse underwater acoustic sensor networks has been investigated in this paper. The mobility-assisted data collection improves energy efficiency and delivery ratio at the cost of increased latency and hence it is more suited for sparse or disconnected networks and in situations

where network lifetime is more important than message delay. For applications which are delay sensitive but not critical, techniques like multiple mobile collectors and priority polling have been found to improve the delay performance. The basic DTN framework having a single mobile sink and cyclic polling and the enhanced one having multiple mobile collectors and priority polling have been implemented in the NS-2 based network simulator, thus enhancing the scope for further research in this area. The enhanced model has been found to support delay-sensitive applications and optimize the delay and delivery performance.

## REFERENCES

- [1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater acoustic sensor networks: research challenges," *Ad hoc networks*, vol. 3, no. 3, 2005, pp. 257–279.
- [2] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*. ACM, 2003, pp. 27–34.
- [3] R. C. Shah, S. Roy, S. Jain, and W. Brunette, "Data mules: Modeling and analysis of a three-tier architecture for sparse sensor networks," *Ad Hoc Networks*, vol. 1, no. 2, 2003, pp. 215–233.
- [4] P. Xie, Z. Zhou, Z. Peng, H. Yan, T. Hu, J.-H. Cui, Z. Shi, Y. Fei, and S. Zhou, "Aqua-sim: an ns-2 based simulator for underwater sensor networks," in *OCEANS 2009, MTS/IEEE Biloxi-Marine Technology for Our Future: Global and Local Challenges*. IEEE, 2009, pp. 1–7.
- [5] ns-2 network simulator. [Online]. Available: <http://www.isi.edu/nsnam/ns>, [last accessed : April 2014]
- [6] X. Peng, Z. Zhong, N. Nicolas, C. Jun-Hong, and S. Zhijie, "Efficient vector-based forwarding for underwater sensor networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, 2010, pp. 1–13.
- [7] S. Jain, K. Fall, and R. Patra, *Routing in a delay tolerant network*. ACM, 2004, vol. 34, no. 4.
- [8] Z. Guo, G. Colombo, B. Wang, J.-H. Cui, D. Maggiorini, and G. P. Rossi, "Adaptive routing in underwater delay/disruption tolerant sensor networks," in *Wireless on Demand Network Systems and Services, 2008. WONS 2008. Fifth Annual Conference on*. IEEE, 2008, pp. 31–39.
- [9] S. Jain, R. C. Shah, W. Brunette, G. Borriello, and S. Roy, "Exploiting mobility for energy efficient data collection in wireless sensor networks," *Mobile Networks and Applications*, vol. 11, no. 3, 2006, pp. 327–339.
- [10] M. C. Domingo and R. Prior, "Energy analysis of routing protocols for underwater wireless sensor networks," *Computer Communications*, vol. 31, no. 6, 2008, pp. 1227–1238.
- [11] M. Zorzi, P. Casari, N. Baldo, and A. F. Harris, "Energy-efficient routing schemes for underwater acoustic networks," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 9, 2008, pp. 1754–1766.
- [12] L. He, Y. Zhuang, J. Pan, and J. Xu, "Evaluating on-demand data collection with mobile elements in wireless sensor networks," in *Vehicular technology conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*. IEEE, 2010, pp. 1–5.
- [13] A. A. Somasundara, A. Kansal, D. D. Jea, D. Estrin, and M. B. Srivastava, "Controllably mobile infrastructure for low energy embedded networks," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 8, 2006, pp. 958–973.
- [14] S. Yoon, A. K. Azad, H. Oh, and S. Kim, "Aurp: An auv-aided underwater routing protocol for underwater acoustic sensor networks," *Sensors*, vol. 12, no. 2, 2012, pp. 1827–1845.
- [15] G. A. Hollinger et al., "Underwater data collection using robotic sensor networks," *Selected Areas in Communications, IEEE Journal on*, vol. 30, no. 5, 2012, pp. 899–911.
- [16] S. Motoyama, "Flexible polling-based scheduling with qos capability for wireless body sensor network," in *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*. IEEE, 2012, pp. 745–752.
- [17] V. Kavitha and E. Altman, "Queuing in space: Design of message ferry routes in static ad hoc networks," in *Teletraffic Congress, 2009. ITC 21 2009*. 21st International. IEEE, 2009, pp. 1–8.
- [18] R. J. Urick, "Principles of underwater sound," McGraw-Hill, New York, London, 1983.
- [19] H. Takagi, *Stochastic analysis of computer and communication systems*. Elsevier Science Inc., 1990.
- [20] M. Ajmone Marsan, L. De Moraes, S. Donatelli, and F. Neri, "Cycles and waiting times in symmetric exhaustive and gated multiserver multiqueue systems," in *INFOCOM'92. Eleventh Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE*. IEEE, 1992, pp. 2315–2324.
- [21] O. J. Boxma, *Analysis and optimization of polling systems*. Centre for Mathematics and Computer Science, 1991.

# On Type II Hybrid-ARQ with Decode and Forward Relay using Non-Binary Rate-Compatible Punctured LDPC Code on MIMO SC-FDMA up-link

Tomotaka Hamada

Dept. of Computer Science and Engineering  
Nagoya Institute of Technology  
Nagoya, Japan  
E-mail: 24417582@stn.nitech.ac.jp

Yasunori Iwanami

Dept. of Computer Science and Engineering  
Nagoya Institute of Technology  
Nagoya, Japan  
E-mail: iwanami@nitech.ac.jp

**Abstract**— In this paper, Non-Binary Rate-Compatible Punctured Low Density Parity Check (NB RCP LDPC) code is designed over the extended Galois Field. The designed NB RCP LDPC code is applied to the type II Hybrid Automatic Repeat reQuest (HARQ) with Decode and Forward (DF) relay on Multiple Input Multiple Output (MIMO) Single Carrier-Frequency Division Multiple Access (SC-FDMA) up-links. The designed code enables us to decrease the coding rate with incremental redundancy for each retransmission in HARQ. The retransmission is done from the DF relay after the successful decoding in the relay. We have verified through computer simulations that the proposed type II HARQ scheme with DF relay greatly improves the throughput and average retransmission characteristics compared with the scheme without DF relay. Multiple relay cases are also considered.

**Keywords**-NB RCP LDPC code; Hybrid-ARQ; Decode and Forward Relay; MIMO SC-FDMA; Symbol-LLR.

## I. INTRODUCTION

An LDPC code which suits the flexible coding rate design and has the high error correcting capability through iterative decoding can be constructed on arbitrary extended Galois field. The Non-Binary (NB) LDPC code constructed on extended Galois field generally exhibits the better BER performance than the binary LDPC codes [1][2]. There also exist Rate-Compatible Punctured (RCP) LDPC codes with variable coding rate obtained by properly puncturing the mother LDPC code [3]. The RCP LDPC codes enable us to use the same decoder as the mother code and suit the ARQ error correcting schemes [4] with the incremental redundancy. By combining the NB LDPC codes with the RCP codes, the NB RCP LDPC codes were designed and the designed NB RCP LDPC codes were applied to the type II HARQ [5]. On the other hand, the Decode and Forward (DF) relay schemes [6] are useful for HARQ schemes. By using the DF relay, the source node can be replaced by the relay, once the relay correctly decodes the LDPC encoded packet from the source. This replacement from the source to the relay effectively reduces the number of retransmissions and improves the throughput. The NB RCP LDPC coded type II HARQ with DF relay is applied to the MIMO-OFDM modulation in [7]. The incremental redundancy in HARQ with DF relay is especially suited to the up-link transmission like in Long Term Evolution (LTE) or 4G. Due to the necessities of low Peak to Average Power

Ratio (PAPR) and the high power efficiency in the amplification, MIMO SC-FDMA [8] is usually adopted to the up-links in cellular networks. Among SC-FDMA, interleaved SC-FDMA is especially useful because of its very low PAPR nature and excellent frequency diversity effect [8]. In [9], NB LDPC coding with NB repetition codes is applied to multiple relay case for flat fading channel. However, the application of NB RCP LDPC codes to MIMO interleaved SC-FDMA with multiple DF relays has not been reported yet. In this paper, we have investigated the NB RCP LDPC coded type II HARQ with DF relays on MIMO interleaved SC-FDMA up-links. We have verified through computer simulations that the proposed up-link scheme greatly improves the throughput and the average number of retransmission characteristics compared with the case of no DF relay. Moreover, we considered multiple relay cases, i.e., serial or parallel arrangement of two relays.

The paper is organized as follows. In Section II, RCP LDPC code is introduced. In Section III, NB LDPC coded Type II HARQ scheme is described. In Section IV, we propose the DF relaying scheme. In Section V, we present the symbol LLR generation in interleaved SC-FDMA demodulation. In Section VI, computer simulation results are shown. The paper concludes with Section VII.

## II. RCP LDPC CODE

The encoding and decoding procedure of RCP LDPC code is as follows. We call the code before puncture and the code after puncture as the mother code and the punctured code, respectively. In RCP LDPC code, the encoder and decoder of mother code can also be applied to the punctured code. When the parity check matrix of mother code is given by  $\mathbf{H}_M (n_M \times n_N)$  and the generator matrix by  $\mathbf{G}_M (n_N \times n_K)$  with  $n_K = (n_N - n_M)$ , the coding rate of mother code becomes  $R_M = (1 - n_M / n_N) = n_K / n_N$ . The coding rate after the puncture of  $n_p$  symbols from the mother code is given by  $R_E = n_K / (n_N - n_p)$ . We denote the message vector as  $\mathbf{m} = (m_1, m_2, \dots, m_{n_K})$ , the code word of mother code as  $\mathbf{C}_M = (c_{M1}, c_{M2}, \dots, c_{Mn_N})$ , the index of position to be punctured as  $\mathbf{P} = (p_1, p_2, \dots, p_{n_p})$  and the code word of punctured code as  $\mathbf{C}_P = (c_{P1}, c_{P2}, \dots, c_{Pn_p})$ . The encoding procedure is first to generate the mother code by  $\mathbf{C}_M = \mathbf{m}\mathbf{G}_M$  which is systematic, and next, to puncture the position using  $\mathbf{P}$  to obtain  $\mathbf{C}_P$ . The decoding procedure is to produce the

symbol LLR from the receive signal and it is fed to the mother code decoder as the initial value for the sum-product algorithm. The symbol LLR for the position  $P$  is initially set to 0, because there is no available symbol LLR corresponding to the position  $P$ .

### III. NB RCP LDPC CODED TYPE II HARQ SCHEME

In Fig. 1, we show the transmitter and receiver block diagram of NB RCP LDPC coded Type II HARQ using  $2 \times 2$  MIMO SC-FDMA. At the transmitter, the data bits are firstly encoded by the Cyclic Redundancy Check (CRC)-16 error detecting code and secondly encoded by the NB LDPC code on GF(4) or GF(16). The encoded LDPC code word is divided into the transmission packets and they are modulated by QPSK or 16QAM depending on GF(4) or GF(16), respectively. Matching GF( $Q$ ) to the modulation level  $Q$  is preferable in calculating the symbol LLR and reduces the complexity compared with the use of bit LLR calculation. The modulated QAM symbols are then  $N$ -point FFT transformed at each antenna stream and the subcarrier mapping is done to make the interleaved SC-FDMA spectrum as shown in Fig. 2. The interleaved spectrum is then  $M$ -point IFFT transformed, where  $M = U \times N$  and  $U$  is the number of users. Cyclic Prefix (CP) is added to the time domain complex samples of an IFFT block. After the interpolation filtering and the up-conversion to carrier frequency, the RF signal is transmitted from each antenna. At the base station, after the down-conversion to baseband and the sampling, CP is removed and the  $M$ -point FFT is done to obtain the frequency domain signal. The frequency domain signal is then subcarrier-de-mapped to aggregate the interleaved spectrum of each user back to the  $N$  sample spectrum again. The Frequency Domain Equalization (FDE) is made to compensate the channel frequency response and separate the multiple spatial streams of each user. The FDE weight  $G_u(i)$  in MMSE criterion is expressed as

$$G_u(i) = \mathbf{H}_u(i)^H \left\{ \mathbf{H}_u(i) \mathbf{H}_u(i)^H + n_r \sigma^2 \mathbf{I}_{n_r} \right\}^{-1} \quad (1)$$

where  $\mathbf{H}_u(i)$  is  $n_r \times n_t$  channel matrix at subcarrier  $i$  of user  $u$ ,  $n_t$  is the number of transmit antennas,  $\sigma^2$  is the noise variance of each subcarrier and  $\mathbf{I}_{n_r}$  is the identity matrix of size  $n_r$ . After the FDE,  $N$ -point IFFT is made to obtain the time domain signal of each stream in each user. The symbol LLR defined in section V is then calculated and fed to the NB LDPC decoder. Using Sum-Product Algorithm (SPA), the LDPC code word is decoded and the hard decision is made to obtain the data bits. The data bits are then CRC-checked. NACK or ACK is returned to the transmitter of each user corresponding to the error or no error detection.

In type II HARQ, like in Fig. 3, at the first transmission, only uncoded information symbols are transmitted, and at the second transmission and after, the parity check symbols are retransmitted at the incremental redundancy policy. Accordingly, when the channel condition is good, the first uncoded transmission is successful enough and it achieves high throughput. On the other hand, when the channel is bad, by decreasing the coding rate at each retransmission, the error correction capability increases gradually. The

generation of RCP LDPC code is done only once at the transmitter and there is no need of regeneration of code word when the coding rate is decreased. Therefore there is no increase of complexity of RCP LDPC code compared with the fixed rate LDPC code. Also at the receiver side, the complexity of RCP LDPC decoder does not increase compared with the fixed rate LDPC decoder, because the same and only one LDPC decoder can be used for various coding rates of RCP LDPC code.

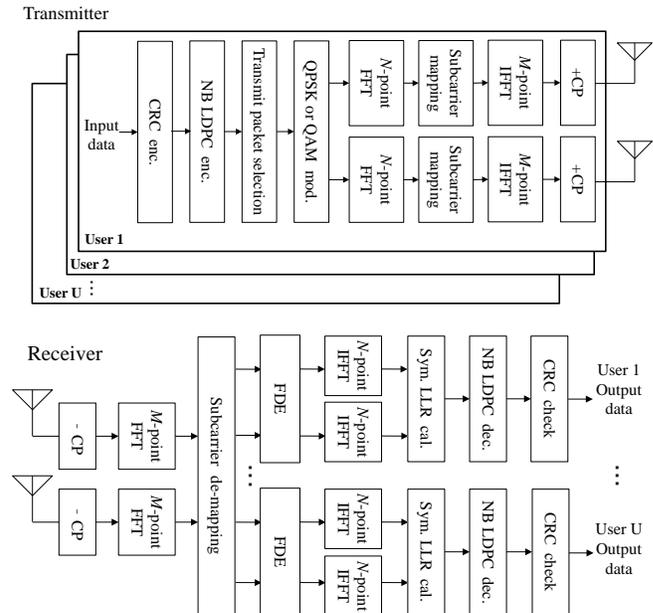


Fig. 1 Transmitter and receiver structure of NB RCP LDPC coded type II HARQ scheme using  $2 \times 2$  MIMO interleaved SC-FDMA

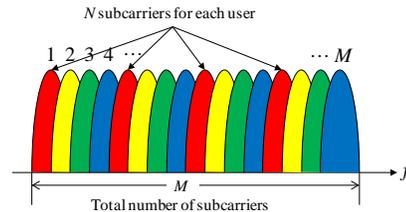


Fig. 2 Sub-carrier mapping in interleaved SC-FDMA (Red, yellow, green and blue spectrum show the subcarriers of user 1, user 2, user 3 and user 4, respectively.)

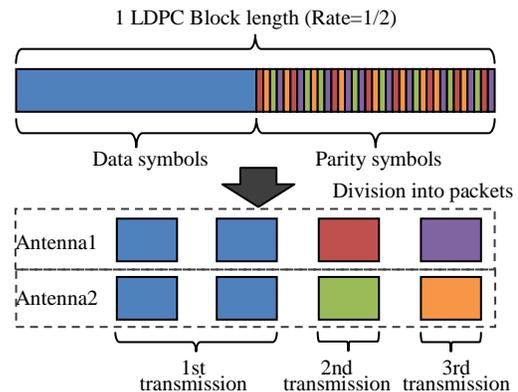


Fig. 3 Division of an LDPC code word into transmission packets

#### IV. DECODE AND FORWARD RELAYING SCHEME

The Decode and Forward relay model is shown in Fig. 4. We consider the relay arrangement where the relay locates at the middle point between the source (transmitter at each user) and the destination (receiver at base station).

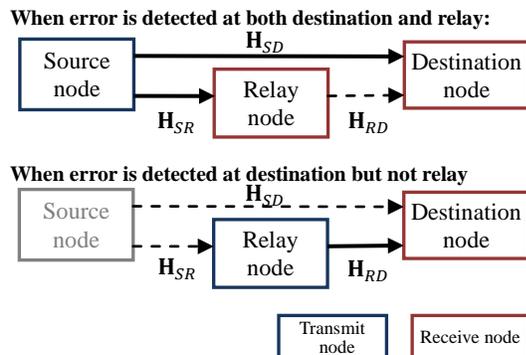


Fig. 4 DF (Decode and Forward) relaying model

At the first transmission, the source broadcasts the uncoded information packet to the relay and destination simultaneously. The relay and destination independently detect the transmission errors using CRC-16 code. The relay and destination independently return (broadcast) ACK or NACK to the source. This ACK or NACK is shared among source, relay and destination. If the destination returns ACK, the transmission finishes at the first transmission and this condition is equivalent to no relay. Otherwise, retransmission is made. The source sends parity check packets with incremental redundancy. The relay and destination receive the parity check packet and combine it with already received packet. The LDPC decoding and CRC error detection is done both at relay and destination. ACK or NACK is returned and shared among source, relay and destination. At this point, if destination returns NACK but relay does ACK, then the relay sends the parity check packet hereafter instead of source, i.e., the source is replaced by the relay which locates closer to destination. The transmission from relay to destination is more successful than source to destination due to the near distance between relay and destination. Also, as the source and relay do not simultaneously retransmit the parity check packet, the total transmission power is the same between with and without relay. This saves the total transmit energy in the case where the same power as source is allocated to the relay.

Next, we consider the two relay cases where two relays are allocated in a serial or parallel manner as shown in Fig. 5(b) or (c), respectively. Fig. 5 (a) is the arrangement of single relay already discussed. In Fig. 5 (b), two relays are allocated in the middle point between source and destination in parallel. In Fig. 5(c), relay 1 and relay 2 are allocated serially with equal distance interval between source and destination. When the power attenuation exponent is given by  $\alpha$  and the distance between source and destination is defined as 1, the relay at the middle point between source and destination in Fig. 5 (a) and (b) receives  $2^\alpha$  times more power than the direct link between source and destination. Similarly, the relay 1 and relay 2 in Fig. 5 (c) receives  $3^\alpha$  times more power than the direct link.

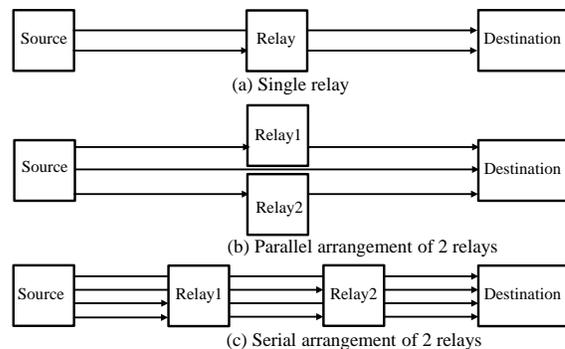


Fig. 5 DF relay arrangement in case of multiple relays

#### V. SYMBOL LLR GENERATION IN MIMO INTERLEAVED SC-FDMA DEMODULATION

The received signal in time domain after  $N$ -point IFFT in Fig. 1 is expressed as

$$y_k = h_k x_k + n_k, \quad k = 1, \dots, N, \quad \xi^2 = (1/2) \cdot E\{|n_k|^2\} \quad (2)$$

where  $h_k$  is the complex gain for the symbol  $x_k$  after the equalization and spatial de-multiplexing, and  $\xi^2$  is the variance of receive noise. When the modulation level of QAM is given by  $Q$  and the probability of occurrence of  $Q$  symbols are all equal, i.e.,  $p(s_1) = \dots = p(s_Q) = 1/Q$ , the symbol LLR is defined and calculated as

$$\begin{aligned} \lambda_a &= \log \left\{ \frac{p(s_a | y_k)}{p(s_1 | y_k)} \right\} = \log \left\{ \frac{(p(y_k | s_a) p(s_a)) / p(y_k)}{(p(y_k | s_1) p(s_1)) / p(y_k)} \right\} \\ &= \log \left\{ \frac{p(y_k | s_a)}{p(y_k | s_1)} \right\} = \log \left\{ \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|y_k - h_k s_a|^2}{2\xi^2}\right)}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|y_k - h_k s_1|^2}{2\xi^2}\right)} \right\} \\ &= \frac{-|y_k - h_k s_a|^2 + |y_k - h_k s_1|^2}{2\xi^2}, \quad a = 1, \dots, Q \end{aligned} \quad (3)$$

#### VI. COMPUTER SIMULATION RESULTS

The BER characteristics of NB RCP LDPC code for each coding rate on GF(4) or GF(16) are examined. The simulation conditions are listed in the shaded area in Table I and the simulation results are shown in Fig. 6 and Fig. 7. C++ language is utilized for programming. We can see that each punctured code having different coding rate is obtained from the mother code with rate 1/2. Each punctured code shows the BER characteristics corresponding to its coding rate.

Next, from Fig. 8~Fig. 15, we show the throughput and the average number of retransmission characteristics of NB RCP LDPC coded type II HARQ on  $2 \times 2$  MIMO interleaved SC-FDMA. We have shown the simulation results for the single relay case of Fig. 5 (a) in Figs. 16 and 17 and the serial two relay case of Fig. 5 (c) in Figs. 9, 11, 13, 15, 16 and 17. The simulation conditions are also listed in Table I.

When QPSK modulation is employed, one GF(4) LDPC mother code word is divided into 16 packets. As the coding

rate of mother LDPC code is  $1/2$ , the former 8 packets contain only information symbols and the latter 8 packets consist of parity check symbols. For the 1st transmission, 8 information packets are transmitted from two antennas. For the 2nd retransmission and after, two parity check packets are retransmitted at each retransmission resulting in lowering the coding rate at receiver from  $4/5$  to  $4/6$ ,  $4/7$ ,  $4/8$ . After all the parity check packets are retransmitted and the coding rate at destination reaches  $1/2$ , if error is still detected at destination, the same RCP LDPC code transmission is repeated two times and each time the symbol LLR's are summed up at destination by symbol LLR addition. Thus, the total 3 transmissions of RCP LDPC code word are done before the final discard of RCP LDPC code in case of failure of error correction at destination.

As a comparative scheme, we also considered the LDPC coded type I HARQ scheme where the coding rate is fixed for each retransmission. The maximum number of retransmissions is limited to 15 and the symbol LLR addition is employed at the destination. For QPSK and GF(4), the throughput characteristics and the average number of retransmission characteristics are shown in Fig. 8 and Fig. 9 and in Fig. 12 and Fig. 13 respectively.

When 16QAM modulation is employed, one GF(16) LDPC mother code word is divided into 8 packets. The former 4 packets contain only information symbols and the latter 4 packets consist of parity check symbols. The coding rate decreases from  $2/3$  to  $2/4$  at each retransmission. After all the packets are retransmitted and the coding rate at destination reaches  $1/2$ , if error is still detected at destination, the same RCP LDPC transmission is repeated 3 times in total, which is the same as the case of QPSK modulation as mentioned in the above. For 16QAM and GF(16), the throughput characteristics and the average number of retransmission characteristics are shown in Fig. 10 and Fig. 11 and in Fig. 14 and Fig. 15, respectively.

For the two serial relay case, we also show the throughput and the average number of retransmission characteristics in Fig. 16 and Fig. 17, respectively.

Regarding the simulation results, we first compare the type I HARQ with the type II HARQ. The throughput characteristic of type I HARQ saturates in the high average  $E_b/N_0$  region, because the coding rate is fixed. As the coding rate of type I HARQ increases, the throughput also increases in the high average  $E_b/N_0$  region. On the other hand, the throughput of type II HARQ approaches to 4 (bps/Hz) and 8 (bps/Hz) in case of QPSK and 16QAM, respectively, in the high  $E_b/N_0$  region. This is because type II HARQ can change the coding rate adaptively and it can use the coding rate of 1 for high SNR region. The slight decrease of throughput in type II HARQ is due to the use of CRC-16 error detection code. We also observe that for entire  $E_b/N_0$  region, the throughput of type II HARQ is optimized and is superior to type I HARQ. However, the average number of retransmission of type II HARQ is worse than type I HARQ. This is because parity check packets are sent sequentially with several time slots in type II HARQ, while the parity check packet is sent at a time in type I HARQ.

Next, we compare the case with relay and without relay. When the average  $E_b/N_0$  is high, the throughput with relay is the same as without relay. This is because, both in the case with relay and without relay, the average number of retransmission is almost 0 for the high average  $E_b/N_0$  region, and there makes no difference between the two. On the other hand, when the average  $E_b/N_0$  is low, we see that the throughput and the average number of

TABLE I. SIMULATION CONDITIONS OF NB RCP LDPC CODED TYPE II HARQ WITH DF RELAY ON  $2 \times 2$  MIMO INTERLEAVED SC-FDMA

NB LDPC mother code	Size of Galois field	GF(4)	GF(16)
	Size of parity check matrix	(512,1024)	(256,512)
	Average weight	(2.66,5.32)	(2.41,4.82)
	Coding rate	4/8	2/4
Punctured code	Information bit length	1024	
	Coding rate	4/4,4/5,4/6,4/7,4/8	2/2,2/3,2/4
Max SPA iteration		20	
Number of users U		4	
Transmit and receive antennas		$2 \times 2$	
modulation		QPSK	16QAM
Number of subcarriers / user		$N = 64$	
Number of total subcarriers		$M = 256$	
CP length ( $T_s$ :QAM symbol length)		$16 \times (T_s / 4) = 4T_s$	
Channel model between each transmit and receive antenna		Quasi-static Rayleigh fading with 16 delay paths having equal average power	
Interval of delay paths		$T_s / 4$	
Channel State Information		Known at receiver	
Error detecting code		CRC-16	
Power attenuation exponent		$\alpha = 3$	

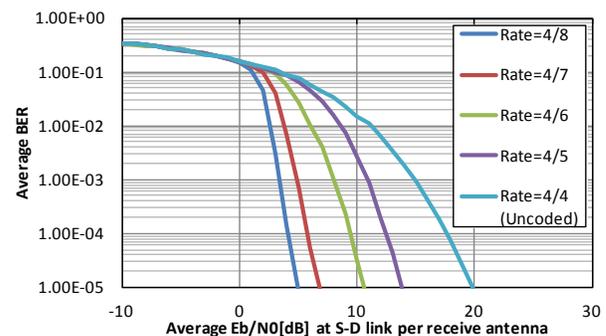


Fig. 6 BER characteristics of NB RCP LDPC GF(4) code using  $2 \times 2$  MIMO interleaved SC-FDMA (QPSK)

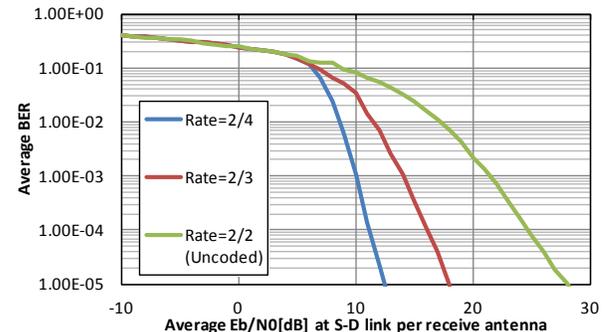


Fig. 7 BER characteristics of NB RCP LDPC GF(16) code using  $2 \times 2$  MIMO interleaved SC-FDMA (16QAM)

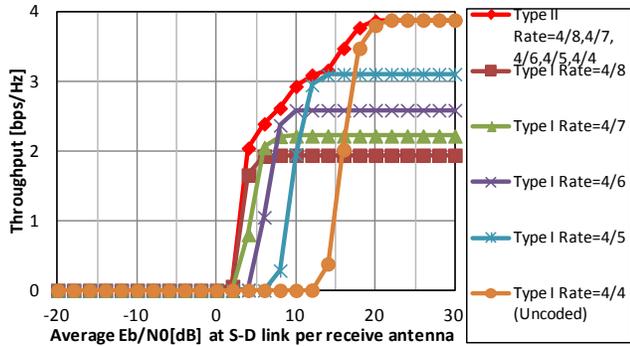


Fig. 8 Throughput characteristics of NB GF(4) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (without relay,  $2 \times 2$ , QPSK)

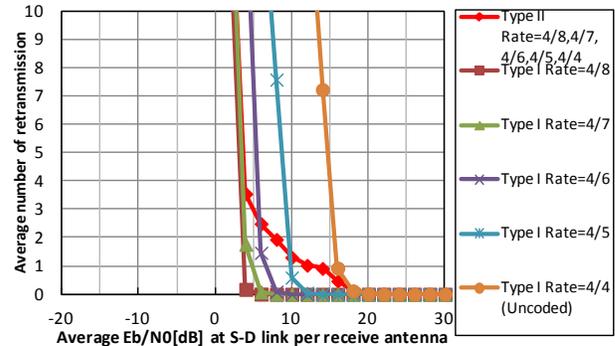


Fig. 12 Average number of retransmission of NB GF(4) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (without relay,  $2 \times 2$ , QPSK)

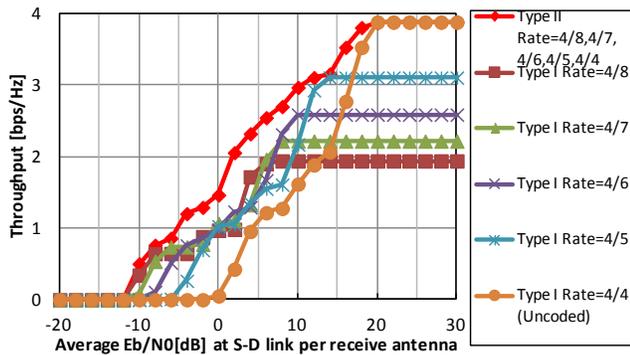


Fig. 9 Throughput characteristics of NB GF(4) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (with 2 serial relays,  $2 \times 2$ , QPSK)

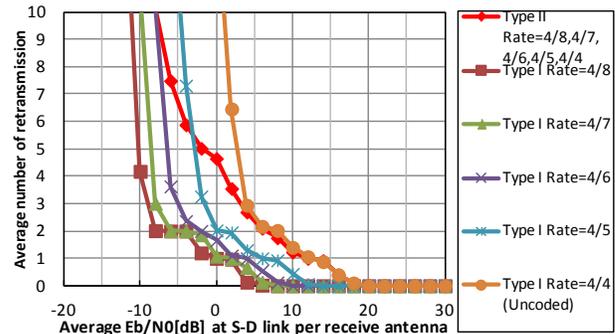


Fig. 13 Average number of retransmission of NB GF(4) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (with 2 serial relays,  $2 \times 2$ , QPSK)

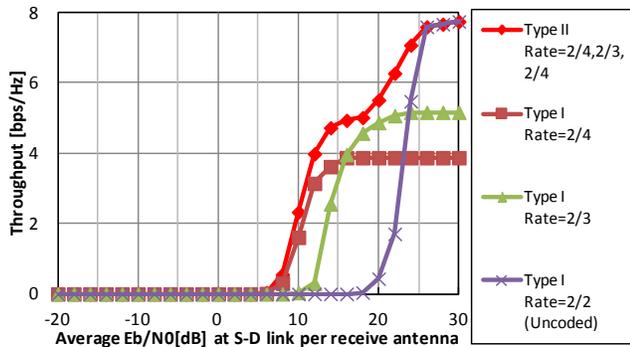


Fig. 10 Throughput characteristics of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (without relay,  $2 \times 2$ , 16QAM)

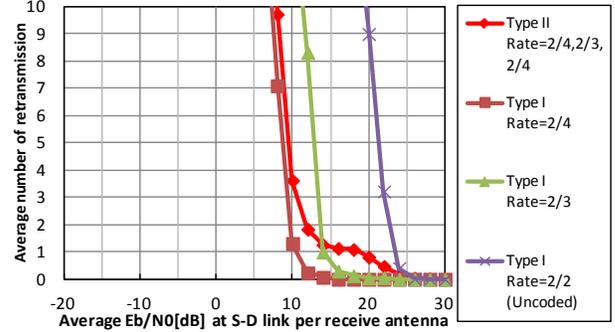


Fig. 14 Average number of retransmission of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (without relay,  $2 \times 2$ , 16QAM)

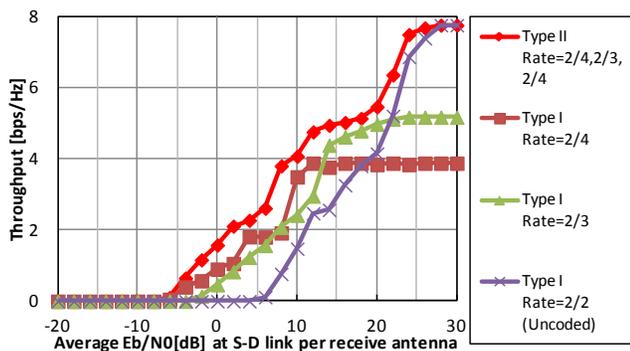


Fig. 11 Throughput characteristics of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (with 2 serial relays,  $2 \times 2$ , 16QAM)

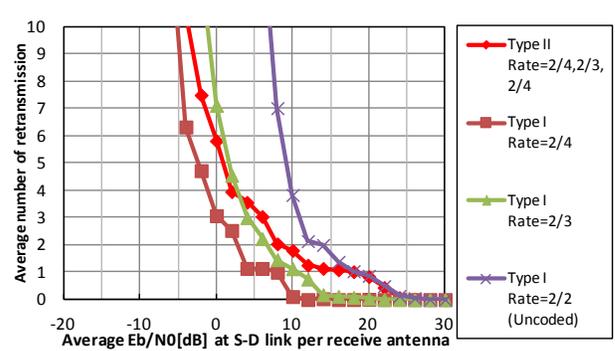


Fig. 15 Average number of retransmission of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy and type I HARQ with fixed coding rate (with 2 serial relays,  $2 \times 2$ , 16QAM)

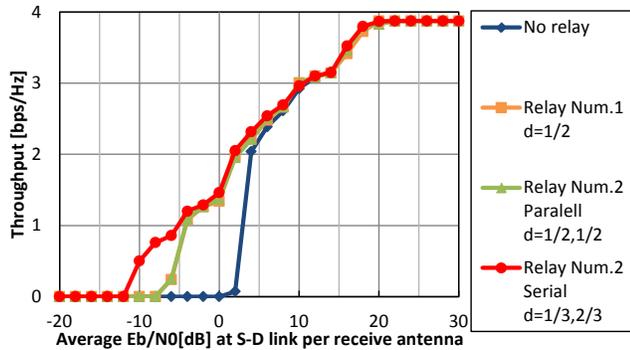


Fig. 16 Comparison of throughput characteristics of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy for three relay arrangements ( $2 \times 2$ , 16QAM)

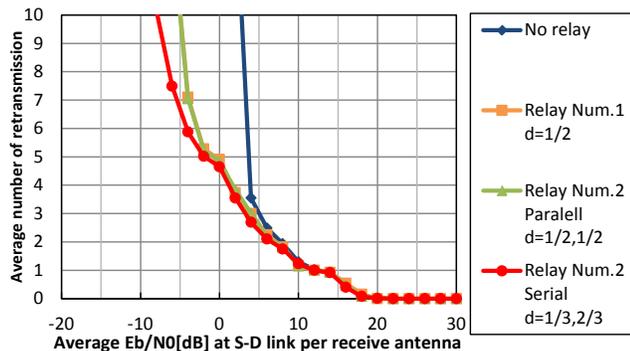


Fig. 17 Comparison of average number of retransmission of NB GF(16) RCP LDPC coded type II HARQ scheme with incremental redundancy for three relay arrangements ( $2 \times 2$ , 16QAM)

retransmission characteristics with relay are much better than the ones without relay. This is because for low average  $E_b/N_0$  region, although the destination frequently fails to decode the code word correctly, the relay succeeds in decoding with high probability. Accordingly, as the retransmission is executed from the relay to the destination instead of the source to the destination, the probability of successful decoding at destination is increased. We also observe that when the number of average retransmission is greater than 0, i.e., the retransmission is done and the total number of transmission is more than 2, the throughput with relay is largely improved compared with the one without relay. This improvement happens below  $E_b/N_0 \approx 20$  (dB) for GF(4) and QPSK in Figs. 8,9,12 and 13, and below  $E_b/N_0 \approx 25$  (dB) for GF(16) and 16QAM in Figs. 10,11,14 and 15.

As for the relay arrangements in Fig. 5, from Fig. 16 and Fig. 17, we see that the throughput and the average number of retransmission characteristics for the serial arrangement in Fig. 5 (c) show the best. The parallel arrangement in Fig. 5 (b) exhibits almost the same performance as the one relay case in Fig. 5 (a). This observation comes from the fact that the receive power at relay or destination in the serial arrangement becomes larger than the parallel arrangement.

## VII. CONCLUSIONS AND FUTURE WORKS

In this paper, assuming the up-link transmission in cellular

wireless networks, we have investigated the throughput and the average number of retransmission characteristics of the proposed NB RCP LDPC coded type II HARQ with DF relays using MIMO interleaved SC-FDMA. We have verified the effectiveness of the proposed scheme through computer simulation. In the proposed scheme, for the first transmission, only uncoded information packet is broadcasted to both for relay and destination. If error is detected at destination, parity check packets are retransmitted for the 2nd and the subsequent retransmission. The error correction decoding is done both at relay and destination. When the destination fails in decoding, but the relay succeeds, the relay replaces the source hereafter. The relay retransmits the remaining packets instead of source. The destination receives the parity check packets with incremental redundancy till the coding rate reaches 1/2. We made clear that by using DF relay the throughput and the average number of retransmission characteristics are improved for low receive SNR region. We also clarified that two DF relays serially arranged between source and destination improve the characteristics further. MIMO interleaved SC-FDMA seems to be a promising candidate for up-link transmission in 4G and after 4G, because of its low PAPR and frequency diversity effect. Considerations on other channel models among nodes, the improvement of BER characteristic of MMSE nulling receiver in SC-FDMA, etc., will be future studies.

## ACKNOWLEDGEMENT

This study is partially supported by the Grants-in-Aid for Scientific Research 24560454 of Japan and the Sharp Corporation.

## REFERENCES

- [1] D. Declercq and M. Fossorier, "Decoding algorithm for nonbinary LDPC codes over GF(q)," *IEEE transactions on Communications*, vol. 55, pp. 633-643, April 2007.
- [2] D. Kimura, F. Guilloud, and R. Pyndiah, "Application of non-binary LDPC codes for small packet transmission in vehicle communications," *The 5th International Conference on ITS Telecommunications*, pp. 109-112, Brest France, June 2005.
- [3] J. Ha, J. Kim, D. Klinc, and S. W. McLaughlin, "Rate-compatible punctured low-density parity-check codes with short block lengths," *IEEE transactions on Information Theory*, vol. 52, no. 2, pp. 728-738, Feb. 2006.
- [4] Y. Tsuruta, Y. Iwanami, and E. Okamoto, "A Study on LDPC Coded Hybrid-ARQ Using Spatially Multiplexed MIMO-OFDM," S36-1, 6 pages, WPMC2009, CD-ROM 5 pages, Sept. 2009.
- [5] T. Kozawa, Y. Iwanami, and E. Okamoto, R. Yamada, and N. Okamoto, "An evaluation on throughput performance for Type II Hybrid-ARQ using non-binary Rate-Compatible-Punctured LDPC codes," *IEICE transactions on fundamentals*, vol. E93, no.11, pp. 2089-2091, November 2010.
- [6] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behaviour," *IEEE transactions on Information Theory*, vol. 50, no. 12, pp. 3062-3080, Dec. 2004.
- [7] H. Tanaka and Y. Iwanami, "On Throughput Characteristics of Type II Hybrid-ARQ with Decode and Forward Relay using Non-Binary Rate-Compatible Punctured LDPC Codes," *ICWMC 2012*, pp. 272-277, June 2012.
- [8] H. G. Myung, J. Lim, and D. J. Goodman, "Single Carrier FDMA for Uplink Wireless Transmission," *IEEE Vehicular Technology Magazine*, pp. 30-38, Sept. 2006.
- [9] D. Declercq, V. Savin, and S. Pfletschinger, "Multi-Relay Cooperative NB-LDPC Coding with Non-Binary Repetition Codes," *ICN 2012*, pp. 205-214, March 2012.

# Interference Management for D2D Communications Underlying Cellular Networks at Cell Edge

Bin Guo, Shaohui Sun

School of Electrical and Information Engineering  
Beihang University  
Beijing, China  
guobin.buaa@gmail.com

Shaohui Sun, Qiubin Gao

State Key Laboratory of Wireless Mobile  
Communications, China Academy of  
Telecommunications Technology (CATT)  
Beijing, China  
sunshaohui@catt.cn

**Abstract**—Device-to-Device (D2D) communications underlying cellular networks enhance the network capacity and spectrum efficiency, but make interference situation more complicated. In this paper, we concentrate on managing the interference between D2D communications and cellular networks at cell edge when sharing channel resources. First of all, a scheme based on interference-suppression-area (ISA), which contains downlink part and uplink part, is proposed to classify the strength of interference between D2D user equipments (UEs) and cellular user equipments (CUEs). Secondly, power control and resource allocation are processed to reduce mutual interference inside ISA. Finally, the range of ISA is discussed, which influences the system performance heavily. The simulation results show that this interference management scheme significantly improves system performance, and the optimal system performance can be obtained by adjusting the range of ISA.

**Keywords**—Device-to-Device (D2D); interference management; cellular networks

## I. INTRODUCTION

The development of mobile communications puts forward higher requirements on transmission rate, spectrum efficiency and network capacity. While radio frequency resources are quite limited, it has become a research hotspot to find efficient ways to fully utilize the channel resources. Because of the remarkable spectrum efficiency, D2D communications are considered as a promising solution to solve these problems.

Under the control of base station (BS), D2D communications underlying cellular networks can obtain the required frequency resources and transmit power, share resources with CUEs and improve spectrum efficiency. In addition, D2D communications reduce the burden of cellular networks, increase the system throughput, reduce the power consumption of the mobile terminal and increase the bit rate, etc. [1][2].

However, resources sharing between D2D communications and cellular networks will cause additional interferences [3]. It is likely to assign orthogonal or nonorthogonal channel resources to D2D links and cellular links. Orthogonal resources are safe but wasted. Nonorthogonal resources cause interferences but improve

spectrum efficiency. Due to limited frequency resources, the nonorthogonal channel resources are more practical to be considered.

In nonorthogonal resources sharing mode, BS can adopt various resource allocation strategies with different gains and complexities. Random resources allocation is a simple way and the interferences between D2D communications and cellular networks are also random. Another simple way is that BS allocates the resources used by the CUEs which are far away from the D2D pairs. This method can make the interference between D2D communications and cellular networks as small as possible.

At cell edge, the interference between BS and D2D UEs is weak, because of pathloss. Hence, there are two main interferences existing in the system. In downlink, the interference from the D2D transmitting user equipment (TUE) to the CUEs around it, and in uplink, the interference from surrounding CUEs to the D2D receiving user equipment (RUE).

Some efforts have been taken to overcome the interferences when D2D communication underlying cellular networks [4-8]. In [4], the problem of interference management for D2D communications where multiple D2D users coexist with one cellular user was discussed. To optimize the transmit power levels of D2D users to maximize the cell throughput while preserving the SINR performance for the cellular user, the authors investigated the availability of the instantaneous or average channel state information (CSI) at the base station and studied the trade-off between the signaling overhead and the overall system performance. In [5], an interference avoiding scheme for D2D communications was proposed when frequency is persistently allocated, which did not allocate subchannels that nearby cellular users currently use via overhearing signal power of uplink cellular users and calculating the interference in the frequency domain. It is noticed that some researches utilized the interference limited area to manage the interference. In [6], the mutual interferences between D2D communications and cellular networks were restricted under the constraints by adopting the interference limited area control method in downlink. Simulation results showed that the proposed scheme can significantly improve the total capacity of cellular and D2D communication, in addition to suppressing the mutual interferences. In [7], the authors

proposed a  $\delta_D$ -interference limited area control scheme to manage the interference from cellular networks to D2D communication while exploiting the same uplink resources. The results indicated that the system capacity can be improved only at a small cost of cellular communication performance. In [8], D2D communication underlying a 3GPP LTE-Advanced cellular network was considered and an interference limited local area scenario was used in system simulation. The results showed that D2D communication in this scenario increased the total throughput. However, these works didn't clarify the relationship between system performance and the range of the area. Actually, the area range has a great influence on system performance.

In this paper, an interference management scheme based on ISA is proposed to manage the interference between D2D communications and cellular networks. Firstly, the whole area can be divided into two parts by interference strength. The area with strong interference in downlink and uplink is defined as ISA. Then power control and resources allocation are applied to decrease the strong interference. Furthermore, the range of ISA is discussed. It is analyzed that the optimal system performance can be obtained by adjusting the range of ISA.

The rest of the paper is organized as follow. In Section II, the D2D communications underlying cellular networks is described and the interference problem is formulated. In Section III, the interference management scheme based on ISA is discussed in details. The simulation results are presented and analyzed in Section IV. The main conclusions are drawn in Section V.

## II. SYSTEM MODEL

### A. System Model

The system model considered in this paper is a cellular network with D2D communication underlying it and sharing resources with cellular links as depicted in Fig.1. There are  $M$  CUEs uniformly distributed in the cell and one D2D pair located at the cell edge. In downlink, BS transmits signals to CUEs, and in uplink, CUEs transmit signals to BS. The TUE transmits signals to the RUE and the maximum distance between them is  $D$ .

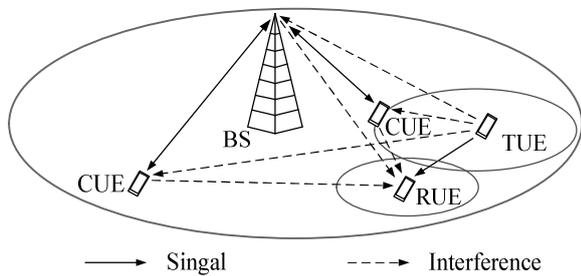


Figure 1. System model of D2D communication underlying cellular systems.

BS allocates channel resources and controls transmit power to cellular links and the D2D link. It is assumed that channels are orthogonal, and the interference only exists in intra-channel due to the channel sharing of the D2D

communication and the cellular communications. Given that all channel resources, both uplink and downlink channel resources, have already equally allocated to CUEs and the number of channel resources equals the number of CUEs  $M$ , i.e.,  $M$  uplink channel resources and  $M$  downlink channel resources. The D2D pair is allowed to share multiple resources, both downlink and uplink channel resources, with CUEs.

The channel considered in this paper is modeled as Rayleigh fading channel, and thus the channel response follows the independent complex Gaussian distribution. Besides, the distance-dependent path loss model is used to measure the signal power transmission loss. The channel gains contain the pathloss and the normalized small-scale fading, and we use  $G_{BS,UE_k(RUE)}$ ,  $G_{UE_k,BS(RUE)}$ , and  $G_{TUE_k,RUE(BS,UE_k)}$  to respectively represent the channel gains from BS to the  $k$ -th CUE (or the RUE) on the  $k$ -th channel resource, the channel gains from the  $k$ -th CUE to BS (or the RUE), and the channel gains from the TUE to the RUE (or BS, the  $k$ -th CUE) on the  $k$ -th channel resource.

$$\begin{aligned} G_{BS,UE_k(RUE)} &= PL_{BS,UE_k(RUE)} h_{BS,UE_k(RUE)} \\ G_{UE_k,BS(RUE)} &= PL_{UE_k,BS(RUE)} h_{UE_k,BS(RUE)} \\ G_{TUE_k,RUE(BS,UE_k)} &= PL_{TUE_k,RUE(BS,UE_k)} h_{TUE_k,RUE(BS,UE_k)} \end{aligned} \quad (1)$$

where  $PL_{ij}$  is the pathloss from  $i$  to  $j$ , and  $h_{ij}$  is the corresponding small-scale fading.

### B. Problem Formulation

In downlink, BS transmits signals to CUEs and the TUE transmits signals to the RUE. Accordingly, CUEs suffer from interferences caused by the TUE and the RUE is disturbed by BS.

We denote the situation of sharing resources between the D2D pair and cellular links in downlink by array  $rd(M)$ , where  $rd(k) = 1$  implies that the D2D pair shares the same resources with the  $k$ -th cellular link, and  $rd(k) = 0$  implies that they don't share the same resource.

The SINR of the  $k$ -th CUE can be represented as:

$$SINR_{UE_k}^{DL} = \frac{P_{BS,UE_k} G_{BS,UE_k}}{rd(k) P_{TUE_k,UE_k} G_{TUE_k,UE_k} + \sigma^2}. \quad (2)$$

where  $P_{BS,UE_k}$  is the transmit power of BS to the  $k$ -th CUE, and  $P_{TUE_k,UE_k}$  is the transmit power of the TUE to the  $k$ -th CUE.  $\sigma^2$  is the noise power.

Considering the resources allocation, the SINR of the RUE in the  $k$ -th channel resource can be represented as:

$$SINR_{RUE,k}^{DL} = rd(k) \frac{P_{TUE_k,RUE} G_{TUE_k,RUE}}{P_{BS,RUE_k} G_{BS,RUE_k} + \sigma^2}. \quad (3)$$

where  $P_{TUE,RUE_k}$  and  $P_{BS,RUE_k}$  are respectively the transmit power of the TUE and BS to the RUE in the k-th channel resource.

In uplink, CUEs transmit signals to BS and the TUE still transmits signals to the RUE. Accordingly, the RUE suffers from interferences caused by CUEs and BS is disturbed by the TUE.

We denote the situation of sharing resources between the D2D pair and cellular links in uplink by array  $ru(M)$ , where  $ru(k) = 1$  implies that the D2D pair shares the same resources with the k-th cellular link, and  $ru(k) = 0$  implies that they don't share the same resource.

The SINR of BS can be represented as:

$$SINR_{BS}^{UL} = \frac{\sum_{k=1}^M P_{UE_k,BS} G_{UE_k,BS}}{\sum_{k=1}^M (ru(k) P_{TUE_k,BS} G_{TUE_k,BS}) + \sigma^2}. \quad (4)$$

where  $P_{UE_k,BS}$  and  $P_{TUE_k,BS}$  are respectively the transmit power of the k-th CUE and the TUE to BS in the k-th channel resource.

The SINR of the RUE in the k-th channel resource can be represented as:

$$SINR_{RUE,k}^{UL} = ru(k) \frac{P_{TUE_k,RUE} G_{TUE_k,RUE}}{P_{BS,RUE_k} G_{BS,RUE_k} + \sigma^2}. \quad (5)$$

Then the system capacity is:

$$\begin{aligned} C &= C_{DL} + C_{UL} \\ C_{DL} &= \sum_{k=1}^M \log_2(1 + SINR_{UE_k}^{DL}) + \log_2(1 + SINR_{RUE,k}^{DL}) \\ C_{UL} &= \log_2(1 + SINR_{BS}^{UL}) + \sum_{k=1}^M \log_2(1 + SINR_{RUE,k}^{UL}) \end{aligned} \quad (6)$$

where  $C_{DL}$  and  $C_{UL}$  are the capacity in downlink and in uplink respectively.

Thus, the optimization objective can be transformed as finding array  $rd(M)$  and  $ru(M)$  to maximize the system capacity.

### III. INTERFERENCE MANAGEMENT SCHEME

In this section, the proposed interference management scheme based on ISA is introduced in details.

#### A. Power Control

In this system model, CUEs and the D2D pair coexist in the system and share whole channel resources. To guarantee the D2D communication, the transmit power of the TUE  $P_{TUE}$  should be promoted as much as possible. On the other hand, large  $P_{TUE}$  will cause severe interference to cellular communications which have higher priority. Hence, the  $P_{TUE}$  should be controlled at a proper level.

In this scheme, we set the  $P_{TUE}$  in downlink. On one hand, the interference from BS to the D2D communication in downlink can be offset effectively by power control. On the other hand, the interference from CUEs to the D2D pair in uplink is related to the distribution of CUEs, which has a great randomness and is not easy to be controlled.

To decide the  $P_{TUE}$ , BS sets up a SINR threshold  $\eta$  and the maximum transmit power of the TUE  $P_{TUEMAX}$ . To ensure the quality of the D2D communication,  $P_{TUE}$  should meet  $\eta$  as far as possible on the premise of not exceeding  $P_{TUEMAX}$ .

The SINR of the RUE in the k-th channel resource should satisfy:

$$SINR_{RUE,k} \geq \eta. \quad (7)$$

Here, we define the  $P_{min}$  as:

$$P_{min} = \eta \cdot \frac{\sum_{k=1}^M (rd(k) \cdot P_{BS,RUE_k} G_{BS,RUE_k}) + \sigma^2}{G_{TUE_k,RUE}}. \quad (8)$$

Hence, the  $P_{TUE}$  is

$$P_{TUE} = \min(P_{min}, P_{TUEMAX}). \quad (9)$$

#### B. ISA Setting

D2D communications have characteristics of short propagation distance and low transmit power. Thus the interference created by the D2D communication is quite limited. If resource scheduling was processed in the area with severe interference, which allocates the orthogonal resources to the D2D pair and the disturbed CUEs, the interference from the D2D communication to cellular networks would be reduced.

Therefore, we define an ISA for the D2D pair to indicate the area with severe interference. The CUEs and the D2D pair in same ISA will cause severe interference to each other, and require orthogonal resources.

In downlink, the receiving interference power of the k-th CUE is:

$$I_{UE_k}(d_{TUE,UE_k}) = \begin{cases} P_{TUE_k,UE_k} \cdot d_{TUE,UE_k}^{-\alpha} & d_{TUE,UE_k} > R_{DL} \\ 0 & d_{TUE,UE_k} \leq R_{DL} \end{cases}. \quad (10)$$

where  $d_{TUE,UE_k}$  is the distance between the TUE and the k-th CUE, and  $\alpha$  is the pathloss exponent.  $R_{DL}$  is the range of the ISA in downlink ( $ISA_{DL}$ ). Given a power threshold, indicated as  $P_{DL}$ , of the receiving signal from the D2D pair to CUEs. When the receiving power equals  $P_{DL}$ , the  $R_{DL}$  can be calculated as:

$$R_{DL} = \left( \frac{P_{DL}}{P_{TUE}} \right)^{\frac{1}{\alpha}}. \quad (11)$$

Similarly, we can obtain the range of the ISA in uplink (ISA<sub>UL</sub>)  $R_{UL}$ :

$$R_{UL} = \left( \frac{P_{UL}}{P_{UE_k}} \right)^{\frac{1}{\alpha}}. \quad (12)$$

where  $P_{UL}$  is the predetermined power threshold of the receiving signal from a CUE to the RUE.

The range of ISA affects the number of channel resources can be used by the D2D communication, i.e.,  $rd(M)$  and  $ru(M)$ . According to the formula (11) and formula (12), the range of ISA<sub>DL</sub> is related to  $P_{DL}$  and  $P_{TUE}$ , and the range of ISA<sub>UL</sub> is related to  $P_{UL}$  and  $P_{UE_k}$ . Therefore, BS can control the range of ISA by determining these parameters.

### C. Resources Allocation

When the interference between CUEs and the D2D pair is inevitable, the resources allocated to them should be orthogonal.

In this scheme, BS provides forbidden channel resources of the D2D communication in downlink, and then the TUE further determines ultimate channel resources in uplink.

In downlink, in order to minimize the interference caused by the D2D communication, the resources of the CUEs outside the ISA<sub>DL</sub> should be allocated to the D2D pair. For this purpose, we set up an exclusive channel D2DCH for the D2D communication to make a CUE estimate whether it is in ISA<sub>DL</sub> or not. In D2DCH, the TUE transmits an identifier to indicate that the D2D pair is communicating. All CUEs in community monitor this channel. When the receiving power is larger than  $P_{DL}$ , it indicates this CUE will be affected by the D2D communication heavily, i.e., in ISA<sub>DL</sub>. After receiving feedback information of the CUEs in ISA<sub>DL</sub>, BS notifies the TUE the resources allocation information of these CUEs. These resources are forbidden resources for the D2D communication.

Then, ISA<sub>UL</sub> is set to further determine the final used resources. The RUE monitors all channel resources. When the receiving power at certain channel resource is larger than  $P_{UL}$ , it indicates the CUE transmitting on this resource will affect the D2D communication seriously, i.e., in the ISA<sub>UL</sub>. The RUE reports these resources information to the TUE. According to the feedback information from BS and the RUE, the TUE obtains information about all forbidden resources and determines the available resources. If there are no available resources, the D2D communication should wait.

By this way, BS can control the range of ISA by setting related parameters, and through resource allocation, the interference between the D2D communication and cellular networks is reduced and system performance is improved.

Moreover, it is necessary to discuss the influence on system performance caused by ISA range. With the ISA

enlarges, the quantity of CUEs in ISA will increase and the cellular network performance will be improved because the interference caused by the D2D communication decreases. However, the D2D performance will decrease because the number of available resources reduces. Therefore, the system performance would not always be improved. There exist an appropriate range of ISA can make the system performance reach optimum. In this paper, we obtain the optimal ISA range by simulation.

## IV. SIMULATION RESULTS

In this section, the proposed scheme is simulated and the system performance is analyzed. The parameters are shown in Table I.

TABLE I. PARAMETERS FOR SIMULATION

Parameter	Value
Cell Radius	1000m
Distance between the TUE and the RUE	50m
Number of CUEs	30
Maximum Transmit Power of BS	46dBm
Maximum Transmit Power of CUEs and TUE	23dBm
Noise Power	-174dBm
Pathloss Exponent	4
SINR Threshold of the D2D Communication $\eta$	10dB

Firstly, we simulate the range of ISA with the change of corresponding power threshold. Due to the distribution of CUEs in ISA can reflect the range of ISA; we use the distribution of CUEs to illustrate the change of ISA. The D2D pair is located at the cell edge with coordinates (1000, 0). The simulation time is 200. Separately, scatter plot in Fig.2 illustrates the distribution of CUEs in the ISA<sub>DL</sub>, Fig.3 is about the CUEs distribution in ISA<sub>UL</sub>, and Fig.4 shows the combine situation of downlink and uplink.

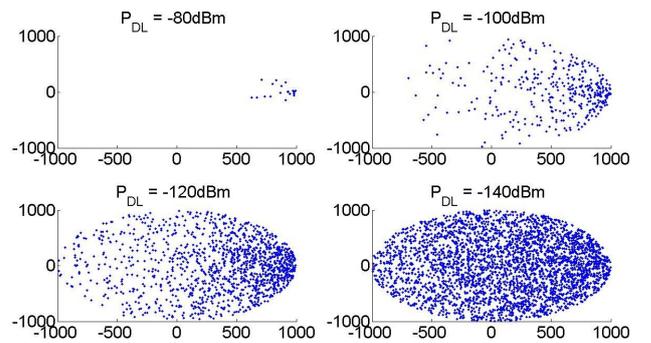


Figure 2. Distribution of CUEs in ISA<sub>DL</sub> with the  $P_{DL}$ .

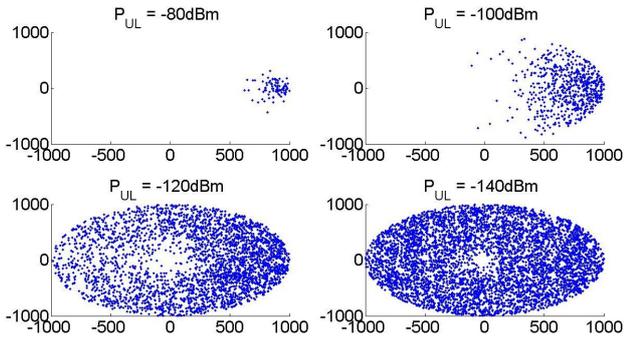


Figure 3. Distribution of CUEs in ISA<sub>UL</sub> with the  $P_{UL}$ .

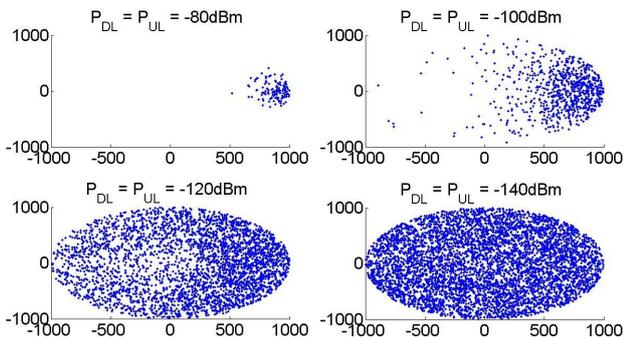


Figure 4. Distribution of CUEs in ISA with the  $P_{DL}$  and  $P_{UL}$ .

According to the above results, with the decrease of power threshold, ISA enlarges. In uplink, due to the higher signal power, the CUEs around BS are not easily affected by the D2D communication, so when the  $P_{UL}$  is small, there is a blank circle around BS in ISA<sub>UL</sub> as shown in Fig.3.

Fig.5 illustrates the system capacity changed with the  $P_{DL}$  and  $P_{UL}$ .

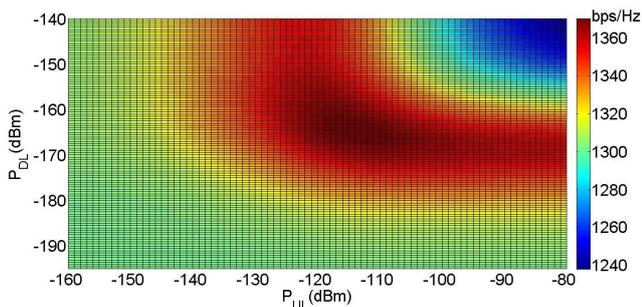


Figure 5. System capacity changed with the  $P_{DL}$  and  $P_{UL}$ .

We can draw a few conclusions from this simulation result. First of all, as we expected, there exists an optimal value of system capacity shown as the dark red areas. The optimal value is appeared when  $P_{DL}$  is around -130dBm and  $P_{UL}$  is around -115dBm. Secondly, when  $P_{DL}$  and  $P_{UL}$  set too large, the system capacity becomes small. This is because when the power threshold is too large, fewer CUEs will enter the ISA, the interference between the D2D communication and cellular networks can't be controlled efficiently. Finally, when the  $P_{DL}$  (or  $P_{UL}$ ) is fixed, the system capacity with the

increase of  $P_{UL}$  (or  $P_{DL}$ ) will first increases then decreases, there is an optimal value.

In order to assess the system performance of the proposed scheme, we compare the scheme with the random scheduling scheme which allocates the whole resources to the D2D communication as shown in Fig.6 and Fig.7. From above results, the maximum system capacity appears when  $P_{DL}$  is around -130dBm and  $P_{UL}$  is around -115dBm, therefore, we separately fix one of them and investigate the system capacity changed with another one.

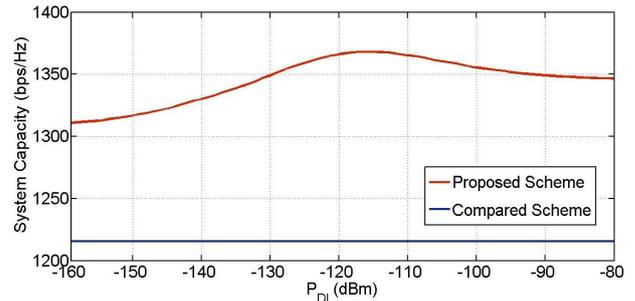


Figure 6. System capacity changed with the  $P_{DL}$ .

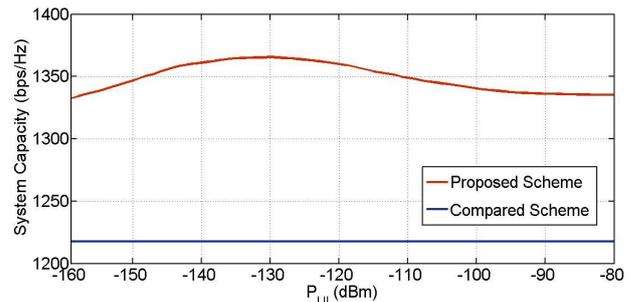


Figure 7. System capacity changed with the  $P_{UL}$ .

Fig.6 is the system capacity changed with  $P_{DL}$  when  $P_{UL}$  equals -115dBm and Fig.7 is the system capacity changed with  $P_{UL}$  when  $P_{DL}$  equals -130dBm. In these two graphs, the red curves represent the system capacity of the proposed scheme and the blue ones represent the system performance of the random scheduling scheme. It is obvious that the proposed scheme is superior to the compared scheme.

## V. CONCLUSION

In this paper, an interference management scheme based on ISA is proposed to manage the interference between the D2D communication at cell edge and cellular networks. BS controls the range of ISA by setting related parameters. On the premise of guaranteeing the D2D communication, BS controls the transmit power and helps determining the channel resources of the D2D communication. The simulation results show that this scheme can significantly promote the system capacity. Moreover, the range of ISA has great influence on the system performance, and the optimal system performance can be obtained by adjusting the range of ISA.

REFERENCES

- [1] C.-H. Yu, K. Doppler, C. B. Ribeiro, and O. Tirkkonen, "Performance impact of fading interference to device-to-device communication underlying cellular networks," IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Sep.2009, pp. 858-862.
- [2] J. Hong, S. Park, H. Kim, S. Choi, and K. B. Lee, "Analysis of Device-to-Device discovery and link setup in LTE networks," in Proceeding of IEEE 24th International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013, pp. 2856-2860.
- [3] C. Xu, L. Song, Z. Han, D. Li, and B. Jiao, "Resource Allocation Using A Reverse Iterative Combinatorial Auction for Device-to-Device Underlay Cellular Networks," IEEE Global Communications Conference (GLOBECOM), Dec. 2012, pp. 4542-4547.
- [4] S. Shalmashi, G. Miao, and S. Ben Slimane, "Interference management for multiple device-to-device communications underlying cellular networks," IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), 2013, pp. 223-227.
- [5] J. Shin, J. P. Choi, and J.-W. Choi, "An autonomous interference avoidance scheme for D2D communications through frequency overhearing," International Conference on ICT Convergence (ICTC), 2013, pp. 1074-1075.
- [6] H. Min, and J. Lee, "Capacity Enhancement Using an Interference Limited Area for Device-to-Device Uplink Underlying Cellular Networks," IEEE Transactions on Wireless Communications, vol. 10, no. 12, Dec. 2011, pp. 3995-4000.
- [7] X. Chen, L. Chen, M. Zeng, X. Zhang, and D. Yang, "Downlink Resource Allocation for Device-to-Device Communication Underlying Cellular Networks," IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Sept. 2012, pp. 232-237.
- [8] K. Doppler, M. Rinne, C. Wijiting, C. B. Ribeiro, and K. Hugl, "Device-to-Device Communication as an Underlay to LTE-Advanced Networks," Communications Magazine, IEEE, vol. 47, no. 12, Dec. 2009, pp 42-49.

# Optimization in Backoff Technique for IEEE 802.15.4/ ZigBee Networks

Muneer Bani Yassein, Yaser Khamayseh, Maged Refat Fakirah, Ghassan Ali Ali Mohsen

Computer Science Department, Jordan University of Science and Technology

Irbid, Jordan

Emails: {masadeh, yaser, mrfakerah12, gamohsen11}@cit.just.edu.jo

**Abstract**—Zigbee/IEEE 802.15.4 is a high level communication and short range standard in Wireless Sensor Networks (WSN), where each node can send and receive data with high security, reliability, low complexity, low power consumption and low cost. It is utilized by industry leaders, such as Samsung, Motorola, Siemens, Philips, etc., and used in industrial, medical purposes, control and monitoring applications. ZigBee supports two operation modes; beacon enabled mode and beaconless enabled mode. In this paper, we present the optimal values of two parameters used in the Zigbee standard, namely, BO (Beacon Order) and SO (Superframe Order), to enhance the network performance, which can be obtained by increasing the network throughput as well as decreasing the total energy consumption and end-to-end delay. The experiment was applied on a star topology, along with an improved backoff mechanism for ZigBee networks. The results showed that the best BO and SO values that increases the throughput, decreases the end-to-end delay as well as decreases the total energy consumption is (8, 8).

**Keywords**—IEEE 802.15.4; zigBee; backoff; beacon; superframe.

## I. INTRODUCTION

The IEEE 802.15.4 standard (Zigbee) [1] is a technological standard for high level communication in WSN, which approved to assign protocols for the data link, Mac and physical layers. This standard is essentially based on IEEE Low Rate Wireless Personal Area Networks. It connects between two devices in the lower level of communication that makes it useful for low data rate, low power and low cost networks, such as WSN [1][2]. It is developed by the ZigBee Alliance [5], and its name came from a zigzag type of dance, which is used by the honeybee to inform the other bee hive members that a new source of food is detected [6]. ZigBee protocols are directed towards earning remote control applications, like wearable health monitoring systems [4].

The MAC layer of IEEE 802.15.4 standard deals with two kinds of network topologies; star and peer to peer networks, and it supports two work modes; beacon mode, and beaconless mode [3].

To avoid a collision that appears in the network while transmitting in the same time and cause loss of data when using beacon mode, a popular approach is used, called slotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) that determines the devices which can rival with the other devices during the access of the channel [8].

There are two ways for channel access: the first way is Contention Free Period (CFP), where the Personal Area Network (PAN) coordinator assigns small period of time for

each node in the Superframe to do whatever it needed without contention or latency, this time is called Guaranteed Time Slot (GTS). The second way is Contention Access Period (CAP). In the beaconless mode the un-slotted CSMA/CA approach is preferred [3][13]. In both modes, the nodes which depend on IEEE 802.15.4 standard try to send their data, and, if the collision happens when more than one node sends in the same time, the user waits for a little period of time, called backoff time; then, it tries to resend its data again after that time is finished.

Backoff time is computed through a linear function selected randomly from the range of  $[0-2BE-1]$ , where BE symbolizes the backoff exponent that is required to determine the amount of time that a node have to wait before trying to send again [4].

The Beacon is a message with a particular form used to synchronize the clock of the node in the network. Two sequential beacons include among them one superframe, which is splatted into 16 active periods of time slots as well as an elective idle amount of time, where all nodes can weight if the sleep mode is enabled as shown below in Figure 1.

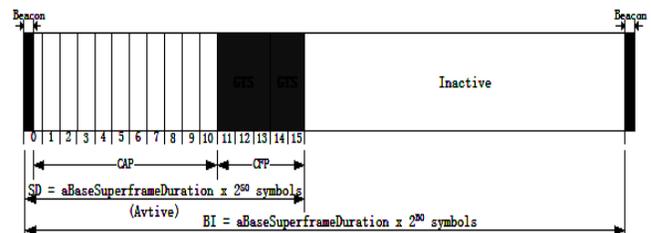


Figure 1. E SuperFrame Structure.

The process of using the beacon message can be clarified as follows: A coordinator sends the beacon signals to the node to start sending its data (beacon enabled PAN). On the other hand, the beaconless mode has no GTS time, but it has CFP, which makes an advantage for it because it saves the battery life unlike the beacon mode [2].

The paper presents the optimal BO and SO value to improve the network performance. The network improvements are gained by increasing network throughput, as well as decreasing the end-to-end delay and total energy consumption. The rest of the paper is organized as follows; Section II presents a literature review of several techniques used in ZigBee networks. Next, Section III shows the simulation and obtained results. Finally, the paper is concluded in Section IV.

## II. RELATED WORK

Many researches have been performed in the domain of the ZigBee networks; mostly, these researches are about how to improve the network throughput as well as decreasing the total energy consumption and average end-to-end-delay, that can be summarized as improving the network performance.

Rao et al. [4] have suggested an Adaptive Backoff Exponent (ABE) algorithm, where BE amount was administered depending on device's participation to the network passing. ABE executes a variable  $\text{macMinBE}$  and approved it just by devices participated in the transference activity. Moreover ABE leads to solve the potentiality of the devices to choose amounts as same as the BE amounts raised essentially by the little quantity of potential BEs. Thus, devices will backoff a corresponding period of time producing collisions. Hence ABE cannot back PANs with great scales.

Rohm et al. [6] have studied the impact of changing fundamental backoff values, such as  $\text{macMaxBEs}$ ,  $\text{macMinBEs}$ , and  $\text{macMaxCSMA}$ , on the efficiency of beaconless ZigBee networks; the study was performed under several traffic loads by suggesting a dynamic scheme for estimating the best backoff values. Where the new dynamic scheme continuously observes different network traffic indicators and change backoff parameters, according to it. Furthermore, they evaluated the traffic loads by observing the packet loss and latency averages at every node to avoid extra communications for nodes, which are needful in the centrally managed algorithm. Also, they ensured that there are no invisible nodes in the whole radio range of nodes. The simulations were implemented by using several packet sizes, such as 133 byte long packets.

Huang et al. [7] have combined two types of wireless technologies, which are: (WLAN) Wireless Local Area Network (IEEE 802.11b) and ZigBee (IEEE 802.15.4) in the health monitoring field. They have suggested a cooperative existence scheme based on dynamic GTS allocation with preemptive technique for the Wearable Health Monitoring systems. They suggested two-layer-architecture: the first layer was released by Body Area Network (BAN) that controls the communication between sensors on the body with the controller on the wrist. The second layer sends the information received from ZigBee devices to the infrastructure network through WLAN. The new pattern works as follows: at first it quantifies the RSSI and Channel Utilization (CU) to locate the status of the channel, so if ZigBee channel is empty the PAN coordinator will choose it, otherwise the new approach will be called and start by calculating GTS through active period in each superframe.

Ha et al. [10] have studied how to increase the throughput and decrease the energy consumption for Zigbee CSMA/CA. They have suggested two techniques; the first one was depended on the collision and Clear Channel Assessment (CCA) values, where they increases and decreases the backoff exponent (BE) value regarding to transmission status. The second technique benefits from the CCA results by shifting the scope of backoff counters. They performed their simulation using the NS-2 simulator with 25 nodes and one coordinator, where there are no invisible nodes and all

nodes forward data packets to the coordinator. And although that most Zigbee applications use a little amount of traffic load, each node may suffer from the saturating of the network and radio congestion. The simulation shows that the proposed techniques improved the throughput by 29.9% and energy efficiency by 86.8%.

Ko et al. [11] have presented a new State Transition Scheme to increase the network performance by modifying the BE values depending on the moving scheme base on the transmission conditions. In their simulation they have used the NS-2 simulator as well as modified NS-2 MAC simulator. From their experiment they observed that the throughput and network performance increased after changing the  $\text{minBE}$  values to 1 rather than the original values that were chosen from the range between 3 and 5.

Lee et al. [12] have adjusted the amount of BE in the slotted CSMA/CA to an effective EBE value that is being set regarding to the volume of devices available in the PAN in the network. The backoff delay time of the nodes decreased regarding to the lack of nodes that joined the network. So, they initialized the EBE to the least value in the range between 2 and  $\text{macMinBE}$  value, and setting the battery life extension (BLE) in the beacon frame to 1 that increases the necessity to long backoff periods due to the increase of nodes that joins to the network.

Khan et al. [14] have suggested the Improved BEB (IBEB) algorithm. Which decreases the potentiality that devices may choose amounts like BE amounts and waits for the same time of backoff. Devices uses a different value to compute the backoff time instead of the random choose of BE only. They pick temporary Backoff value greater than 10% and less than 40% of the computed backoff period of time, either they hired the unit temporary Period (IP) to reduce the potential of picking both IB and BE. The outcomes of simulation exposed that IBEB is better than the BEB scheme while examined on various network scales and loads.

## III. SIMULATION SETUP AND RESULTS DISCUSSION

The optimal BO and SO value that achieves the best network performance were investigated through an improved linear backoff mechanism, because using linear function is more suitable for ZigBee network to minimize the collision between different nodes in the network as well as minimizing the packets delay sustained by the backoff technique.

Simulation evaluation was performed using QualNet v5.2 simulator. QualNet is a simulator based on C/C++ programming language, derived from GloMoSim simulator that was first issued in 2000 by Scalable Network Technologies (SNT) [9]. QualNet is designed to back simulation of large scale and different networks. It has a smooth graphical user interface that supports a lot of tasks, such as network scenario creation, protocol design, and animated implementation.

Simulation was implemented using an IEEE 802.15.4/Zigbee network, depending on star topology that contains one PAN coordinator with 6 reduced functional devices (RFDs) one hop away from the PAN coordinator, where RFD nodes send CBR traffic of 70 bytes to the PAN coordinator during the simulation period. The nodes were

uniformly distributed in 25m×25m of the area as shown in Figure 2, in addition to an inter arrival time, which is equal to of 1 second, as shown below in table 1.

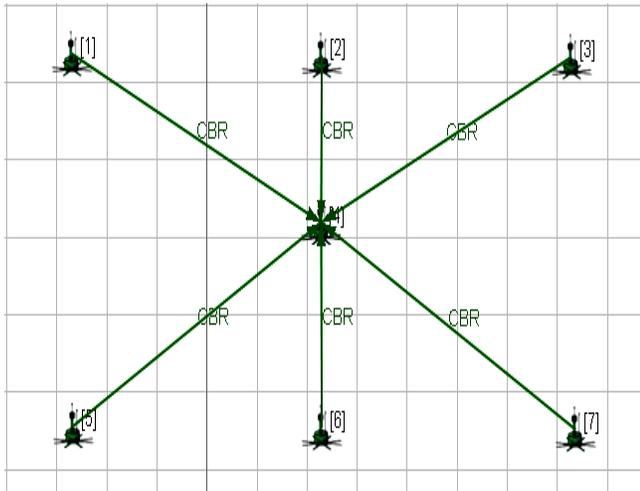


Figure 2. 7 nodes (star topology) implemented on QualNet 5.2 simulator.

TABLE I. QUALNET SIMULATION PARAMETERS USED IN THE SIMULATION.

Parameter	Value
Used Simulator	QualNet 5.2
Radio Type & MAC Type	IEEE 802.15.4
Area (x, y)	25m * 25m
Number of devices	7
Transmission range	5 m
Simulation time	1000 s
-Channel Frequency	2.4 GHZ
Start Time	15
End Time	0
Energy model	MICAZ
Antenna Height	0.08
Traffic	VBR
Item size	70 byte
Channel Access Mechanism	CMSA
Traffic Generator	CBR
Items to send	0
Interval	1 Second
BO and SO	Shown in Table 2

Many experiments were conducted to help researchers to pick up the best BO and SO value that represent the best network performance, where each experiment differed from the others in the value of BO and SO used, as shown below in table 2. BO and SO values were chosen from the range of [1-9] considering the following conditions:

- BO should be greater than or equal to SO.

- If BO is greater than 5, SO also should be greater than 5.

TABLE II. BO AND SO VALUES USED IN THE SIMULATION.

No	BO,SO	Throughput	Average End-to-End Delay	Total Energy Consumption
1	(1,1)	1529.2	0.1664944	12.663684
2	(2,1)	639.8	0.122474	8.298964
3	(2,2)	1993.4	0.2727686	9.0926402
4	(3,1)	330.2	0.508394	6.3243268
5	(3,2)	993.2	0.2942532	6.998513
6	(3,3)	2331	0.4144662	8.3540372
7	(4,1)	172.8	1.064789	5.2675614
8	(4,2)	507	0.394051	5.9119196
9	(4,3)	1170.4	0.4332626	7.4540344
10	(4,4)	2520.2	0.5172802	7.7634934
11	(5,1)	2592	0.6953788	8.4686058
12	(5,2)	361.4	1.3256918	5.0616906
13	(5,3)	620.6	1.440292	6.8974716
14	(5,4)	1301.2	1.409788	8.325229
15	(5,5)	2592	0.6953788	8.4686058
16	(6,6)	1978.4	3.068698	6.5669064
17	(7,6)	918	2.159328	6.077315
18	(7,7)	1666.2	2.043424	6.7881978
19	(8,6)	691.2	4.484266	7.455299
20	(8,7)	1025.6	2.346314	7.579415
21	(8,8)	1681.8	1.893714	4.7075258
22	(9,6)	620.4	16.44958	12.6877532
23	(9,7)	642	11.61832	11.7020778
24	(9,8)	757	4.912716	7.87586
25	(9,9)	910.6	1.2472382	4.5487218

The Performance metrics investigated in our study are:

- Throughput: volume of bits passes through the PAN per second.
- Average End-to-End delay: The average time, which the packet spends to move from the source device to the target PAN coordinator.
- Total Energy Consumption: The total amount of sending, receiving, sleeping and idle energy used by the PAN nodes.

In the following, there are presented three groups of results that show the variance of network performance depending on the BO and SO values. Where the first group of results shown in Figure 3 indicates that the throughput increases to reach its maximum value 2592 Bit/Sec at (BO, SO) =

(5, 1) as well as at (5, 5) and decreases to reach the minimum value 172.8 Bit/Sec when (BO, SO) = (4, 1).

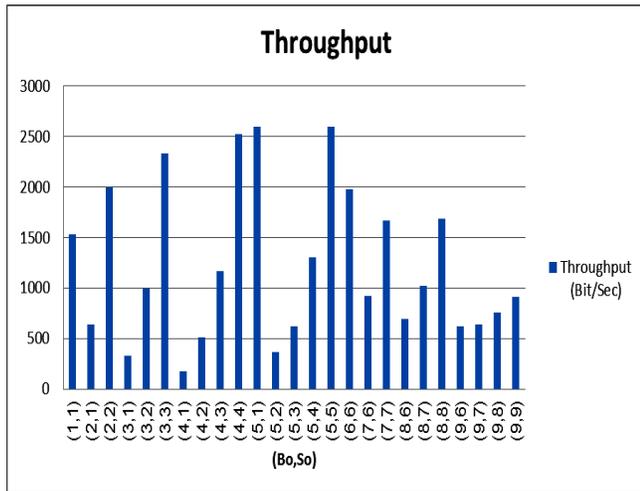


Figure 3. Throughput performance metric using different (BO,SO) values.

The second group of results shown in Figure 4 indicates that the average end-to-end delay increases to reach its maximum value, which is 16.44958 Sec when (BO, SO) = (9, 6) and decreases to reach the minimum value 0.122474 Sec when (BO, SO) = (2, 1).

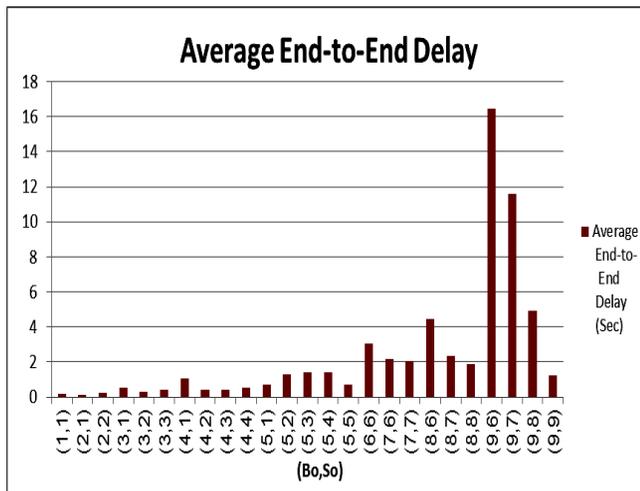


Figure 4. Average end-to-end delay performance metric using different (BO,SO) values.

The third group of results shown in Figure 5 indicates that the total energy consumption increases to reach its maximum value, which is 12.6877532 at (BO, SO) = (9, 6) and decreases to reach the minimum value 4.5487218 when (BO, SO) = (9, 9).

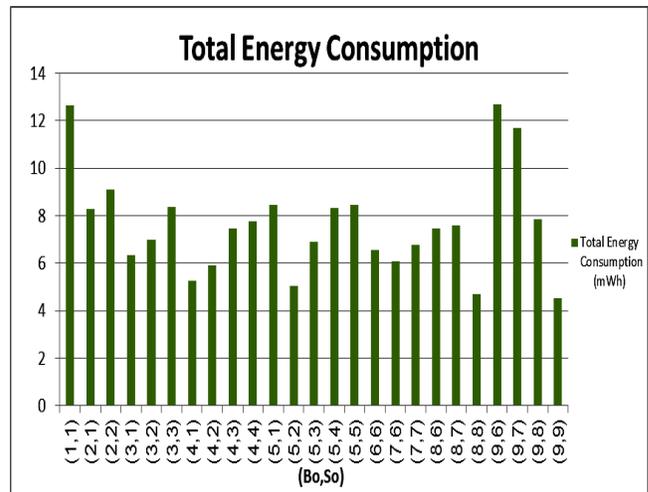


Figure 5. Total energy consumption performance metric using different (BO,SO) values.

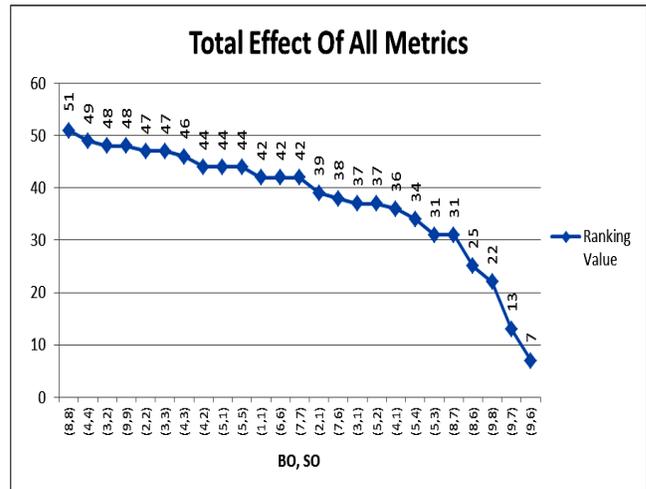


Figure 6. Total effect of all metrics using different (BO,SO) values.

Figure 5 best summaries the results obtained from the previous figures. It was constructed by giving the all BO and SO values suitable ranks according to the results obtained from each studied metric, i.e., (8, 8) has the highest value, that make it the best (BO, SO) values, where (4, 4) comes on the next stage by presenting high amount of throughput, balanced with low amount of end-to-end delay and total energy consumption. Furthermore, the value (9, 6) comes at the end of the list by presenting the worst amount of end-to-end along with the worst total energy consumption value.

#### IV. CONCLUSION AND FUTURE WORK

Zigbee is a high level communications and short-range standard in WSN, it consumes little power (small batteries) and low cost than alternative networks, such as Bluetooth or Wi-Fi. When more than one node tries to send data in the same time, collision and packets delay happens then, they have to wait for a little period of time, called backoff time. In

order to cope with these problems, the optimal BO and SO values that achieve best network performance on star topology were presented, as the network improvements were gained by increasing network throughput as well as decreasing the end-to-end delay and total energy consumption, considering the use of improved backoff mechanism for ZigBee network that minimizes the collision and packet delay between various nodes in the network. The results show that the best BO and SO value that increases the throughput, decreases the end-to-end delay as well as decreases the total energy consumption to is (8, 8). As a future work, we plan to study the effect of the BO and SO values on the same algorithm using other topologies instead of the star topology; and then to compare the network performance in all studied topologies to find out which one is the best. Another proposed work is to study the impact of the data traffic load and data traffic type on the performance of the Linear backoff algorithm by changing the number of packets sent from the nodes to the PAN coordinator per each experiment under different traffic types.

#### REFERENCES

- [1] G. Anastasi, M. Conti, and M. Di Francesco, "A comprehensive analysis of the MAC unreliability problem in IEEE 802.15.4 wireless sensor networks", IEEE Transactions on Industrial Informatics, Vol. 7, Issue 1, February, 2011, Pp. 52 - 6.
- [2] Y. Khamayseh, W. Mardini, R. Jaradat, and R. Hijawai, "Improved backoff technique for ZigBee networks", 3rd International Conference on Computer Modelling and Simulation, Brno, Czech Republic, 3 - 5 September 2012.
- [3] C. Buratti, "Performance analysis of IEEE 802.15.4 Beacon-Enabled mode", IEEE Transactions on Vehicular Technology, Vol. 59, Issue 4, May, 2010, Pp. 2031 - 2045.
- [4] Rao and D. Marandin, "Adaptive backoff exponent algorithm for Zigbee (IEEE 802.15.4)", Next Generation Teletraffic and Wired/Wireless Advanced Networking, Vol. 4003, 2006, Pp. 501 - 516.
- [5] S. Farahani, "ZigBee Wireless Networks and Transceivers", USA: Newnes: 339. ISBN: 978-0-7506-8393-7. 2008.
- [6] D. Rohm and M. Goyal, "Dynamic backoff for IEEE 802.15.4 Beaconless networks", IEEE Mini-Grants (National Science Foundation under Grant No. 0442313), University of Wisconsin Milwaukee, Milwaukee, WI 53201, March 2009.
- [7] M. L. Huang and S. C. Park, "A WLAN and ZigBee coexistence mechanism for wearable health monitoring system", In Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on IEEE, 2009, Pp. 555-559.
- [8] M. B. Yassien, M. Salayma, W. Mardini, and Y. Khamayseh, "Fibonacci backoff algorithm for IEEE 802.15.4/ZigBee", Network Protocols & Algorithms, Macrothink Institute, ISSN 1943-3581 2012, Vol. 4, No. 3, DOI: 10.5296/npa.v4i3.2094, 2012.
- [9] "<http://web.scalable-networks.com/content/qualnet>", retrieved: May, 2014.
- [10] J. Y. Ha, T. H. Kim, H. S. Park, S. Choi, and W. H. Kwon, "An enhanced CSMA-CA algorithm for IEEE 802.15.4 LR-WPANs", Communications Letters, IEEE 11(5), 461-463, May 2007.
- [11] J. G. Ko, Y. H. Cho, and H. Kim, "Performance evaluation of IEEE 802.15.4 MAC with different backoff ranges in wireless sensor networks." Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on IEEE, 2006, Pp. 1-5.
- [12] H. Lee et al. "An efficient slotted CSMA/CA algorithm for the IEEE 802.15.4 LR-WPAN." Information Networking (ICOIN), International Conference on. IEEE, 2011, pp. 488-493.
- [13] P. Kinney, "Zigbee technology: Wireless control that simply works." Communications design conference. Vol. 2. October 2003.
- [14] B. M. Khan, F. H. Ali, and E. Stipidis, "Improved backoff algorithm for IEEE 802.15.4 wireless sensor networks." Wireless Days (WD), 2010 IFIP. IEEE, October. 2010, Pp. 1-5.

# Circuit-Switched Voice over Scalable UMTS

Soumya Das, Yuheng Huang, Edwin Park, and Samir Salib Soliman  
 QUALCOMM Technologies, Inc.  
 San Diego, CA 92121  
 {soumyad, yuhengh, epark, ssoliman}@qti.qualcomm.com

**Abstract**— Scalable Universal Mobile Telecommunications System (S-UMTS) carriers occupy bandwidth different than regular UMTS carriers and may fit in available spectrum that cannot be used by UMTS carriers today. Due to the bandwidth scaling, the data rates in S-UMTS are scaled by the bandwidth scaling factor compared to regular UMTS for the same physical layer configuration. However, some services like Circuit-Switched (CS) voice needs the same data rate in S-UMTS as in regular UMTS. The modified configuration needed in S-UMTS to maintain the data rate of such services is proposed. Additionally, both uplink and downlink voice capacity in S-UMTS are shown to scale with the bandwidth scaling factor.

**Keywords**- S-UMTS; AMR; TTI; spreading factor; slot format.

## I. INTRODUCTION

UMTS has a chip rate of 3.84 Megachips per second (Mcps) and a nominal 5 MHz channel bandwidth. Many operators own spectrum that is non 5 MHz multiples, creating unusable narrow-bandwidth fragments where a 5 MHz carrier cannot be accommodated. Fig. 1 shows an example where a regular UMTS carrier may not fit the available spectrum between two UMTS carriers while a S-UMTS carrier occupying less than 5 MHz may fit.

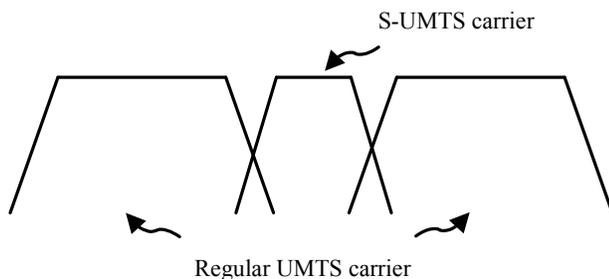


Figure 1. Scalable bandwidth UMTS carrier

This paper focuses on generating a S-UMTS carrier by scaling down the chip rate with respect to a regular UMTS carrier. Hence, such a S-UMTS carrier occupies smaller bandwidth than a regular UMTS carrier. Accordingly, the chip duration is increased or “dilated” by the same factor. For example, to generate a 2.5 MHz S-UMTS carrier, the chip rate is reduced by a factor of 2 from 3.84 Mcps to 1.92 Mcps. In S-UMTS due to the time dilation, chip duration, slot duration, frame duration, sub frame duration, and Transmission Time Interval (TTI) get scaled up, i.e., dilated by a factor of  $D_{cr}$  which is the chip rate divisor ( $D_{cr}=2$  for 2.5 MHz S-UMTS). Hence, the data rate in S-UMTS gets scaled down by the same factor  $D_{cr}$  compared to regular UMTS. The same Power Spectral Density (PSD) is assumed for S-UMTS and UMTS.

S-UMTS has the same link budget as UMTS and thus maintains the same coverage. S-UMTS parameters with respect to regular UMTS are shown in TABLE I.

TABLE I. SCALABLE UMTS PARAMETERS

	Regular UMTS ( $D_{cr}=1$ )	S-UMTS
Nominal Bandwidth (MHz)	5	$5/D_{cr}$
Chip Rate (Mcps)	3.84	$3.84/D_{cr}$
Radio Frame Duration (ms)	10	$10 \times D_{cr}$
Data Rate	D	$D/D_{cr}$
Power	P	$P/D_{cr}$
Range	R	R
Capacity	C	C

The scaling down of data rate in S-UMTS poses a challenge to support services like CS voice using the same vocoder where the service needs to maintain the same data rate as in regular UMTS. The support for 12.2 kbps Adaptive Multi-Rate (AMR) over 2.5 MHz ( $D_{cr}=2$ ) S-UMTS is discussed in this paper. The rest of the paper is organized as follows. Section II discusses the configurations for supporting AMR 12.2 kbps over regular UMTS; Section III discusses the changes in configurations needed for supporting AMR 12.2 kbps over  $D_{cr}=2$  S-UMTS; Section IV shows that the voice capacity of S-UMTS scales with the bandwidth while Section V concludes the paper.

## II. FULL RATE AMR 12.2KBPS VOICE OVER UMTS

Full rate AMR corresponds to bit rate 12.2 kbps, namely “Conversational / speech / UL:12.2 DL:12.2 kbps / CS RAB + UL:3.4 DL:3.4 kbps SRBs for DCCH”, as defined in [1], Section 6.10.2.2. Every 20 ms, the AMR vocoder produces a set of voice frames designated Class-A, Class-B, and Class-C. The Class-A bit sequence is the most important, is CRC protected and convolutionally encoded at rate 1/3. The Class-B does not have a CRC, but is encoded at the same rate of 1/3 as Class-A. The Class-C is the least important, which does not have a CRC and only encoded at rate 1/2. Each class is mapped to a separate Radio Access Bearer (RAB) subflow, which is then mapped to a separate Dedicated Traffic Channel (DTCH logical channel). The 3 DTCHs operate in Radio Link Control (RLC) Transparent Mode (TM) with a maximum Service Data Unit (SDU) size of 81, 103, and 60 bits respectively. The configuration for the DTCH carrying Class A bits also allows a SDU size of 39 bits for the Silence Indicator Descriptor (SID) and 0 bits for a Null frame. Medium Access Control (MAC) maps each DTCH directly

onto the associated Dedicated Channel (DCH transport channel) without adding a header. Using a TTI of 20 ms, half of the MAC Protocol Data Unit (MPDU) is transmitted during one radio frame (10 ms), and the other half is transmitted during the next radio frame.

There are four DCCHs allocated to carry control plane signaling, two for Radio Resource Control (RRC) and two for Non Access Stratum (NAS). MAC multiplexes the four DCCHs onto a single DCH, adding a header containing the C/T (Control/Traffic) field to identify the logical channel from which the data was sent. The DCH is CRC protected and convolutionally encoded at rate 1/3. Using a 40 ms TTI, the MAC PDU is distributed over four radio frames, concatenated with the blocks from the voice DCHs. Therefore, there are four Signaling Radio Bearers (SRBs) for RRC and NAS signaling and one RAB for voice. The mapping is as follows:

- 4 SRBs → 4 DCCHs → 1 DCH using 40 ms TTI
- 1 RAB → 3 DTCHs → 3 DCHs using 20 ms TTI

Fig. 2 shows the transport channel procedure for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in UMTS.

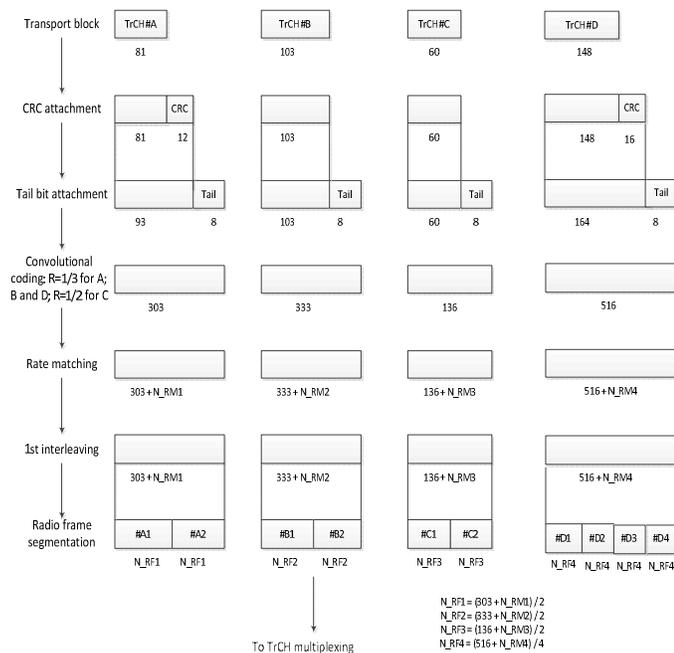


Figure 2. TrCH procedure for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in UMTS

The individual TrCHs are broken from their TTI into 10 ms frames. These frames are then serially multiplexed into a Coded Composite Transport Channel (CCTrCH). Thus each CCTrCH contains a portion of the AMR Class-A, Class-B, and Class-C bit sequences, plus a portion of the signaling information. A downlink Dedicated Physical Channel (DPCH) with 60 kbps is allocated to carry one CCTrCH. A second round of interleaving is done on the CCTrCh. The 510 coded bits per radio frame include CRC and tail bits, rate 1/2 or 1/3

bits-to-coded bits coding, and rate matching (RM). Then the coded bits go through serial-to-parallel conversion and are placed onto a 30 kbps Dedicated Physical Data Channel (DPDCH). Fig. 3 shows the transport channel multiplexing and mapping to physical channels for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in UMTS. It is to be noted that the Fig. 2 and 3 are for one of the six transport format combinations (TFCs) as specified in [1], Section 6.10.2.4.1.4: (RAB subflow#1, RAB subflow#2, RAB subflow#3, DCCH) = (TF2, TF1, TF1, TF1)

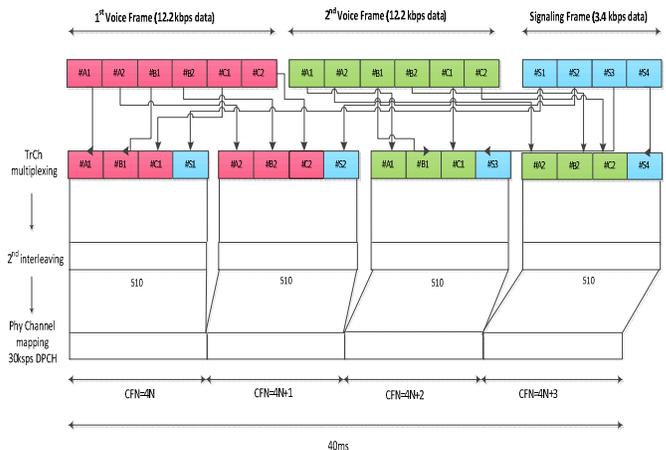


Figure 3. TrCH multiplexing and mapping to PhCH for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in UMTS

TABLE II. TRANSPORT CHANNEL PARAMETER FOR CONVERSATIONAL/SPEECH/UL/DL 12.2 KBPS/CS RAB

RAB/Signalling RB		RAB subflow #1	RAB subflow #2	RAB subflow #3
TB sizes, bit		39, 81 (alt. 0, 39, 81)	103	60
TFS	TF0, bits	0x81	0x103	0x60
	TF1, bits	1x39	1x103	1x60
	TF2, bits	1x81	N/A	N/A
TTI, ms		20	20	20
Coding type		CC 1/3	CC 1/3	CC 1/2
CRC, bit		12	N/A	N/A
Max number of bits/TTI after channel coding		303	333	136
Uplink: Max number of bits/radio frame before rate matching		152	167	68
RM attribute		180 to 220	170 to 210	215 to 256

TABLE II and TABLE III show the UL/DL transport channel parameters for CS RABs and SRBs respectively [1]. In regular UMTS, slot format 8 is typically used in DL, slot format 2 for UL DPDCH, and slot format 0 for UL Dedicated Physical Control Channel (DPCCH) for AMR 12.2 kbps. During compressed mode, slot format 8B is used in DL, slot format 3 for UL DPDCH, and slot format 0A or 0B for UL DPCCH. The DL DPDCH and DPCCH fields for different slot formats, UL DPCCH fields for different slot formats and UL DPDCH fields for different slot formats are defined in [2].

TABLE III. TRANSPORT CHANNEL PARAMETER FOR DL/UL 3.4 Kbps SRBs FOR DCCH

RAB/Signalling RB	SRB#1	SRB#2	SRB#3	SRB#4
User of radio bearer	RRC	RRC	NAS_DT High prio	NAS_DT Low prio
TB sizes, bit	148 (alt 0, 148)			
TFS	TF0, bits	0x148 (alt 1x0)		
	TF1, bits	1x148		
TTI, ms	40			
Coding type	CC 1/3			
CRC, bit	16			
Max number of bits/TTI after channel coding	516			
UL: Max number of bits/radio frame before RM	129			
UL RM attribute	155 to 185			
DL RM attribute	155 to 230			

TABLE IV. DL SLOT FORMAT – DPDCH AND DPCCH FIELDS FOR UMTS

Slot Format	Channel Bitrate (kbps)	Channel Symbol Rate (ksps)	SF	Bits/Slot	DPDCH Bits/Slot NTPC, NTFCL, NPilot	DPDCH Bits/Slot NData1, NData2	Transmitted slots per radio frame
8	60	30	128	40	2,0,4	6, 28	15
8B	120	60	64	80	4,0, 8	12, 56	8-14

TABLE V. UL SLOT FORMAT – DPDCH FIELDS FOR UMTS

Slot Format	Channel Bitrate (kbps)	Channel Symbol Rate (ksps)	SF	Bits/Frame	Bits/Slot	Ndata
2	60	60	64	600	40	40
3	120	120	32	1200	80	80
4	240	240	16	2400	160	160

TABLE VI. UL SLOT FORMAT – DPCCH FIELDS FOR UMTS

Slot Format	Channel Bitrate (kbps)	SF	Bits/Frame, Bits/Slot	NPilot, NTPC, NTFCL, NFBI	Transmitted slot per radio frame
0	15	256	150, 10	6, 2, 2, 0	15
0A	15	256	150,10	5, 2, 3, 0	10-14
0B	15	256	150, 10	4, 2, 4, 0	8-9

For the RM example in TABLE VII, the DL slot format is 8 with (6+28) x 15=510 (NData1 and NData2 in TABLE IV) data bits in a 10 ms radio frame. The RM attributes and the channel coded blocks for the four transport channels are shown in TABLE VII. After the radio frame segmentation, using the TTI of each of the four transport channels, the above bits match 510, i.e., 312/2 + 328/2 + 166/2 + 428/4 = 510.

TABLE VII. RATE MATCHING EXAMPLE

	TrCh# A	TrCh# B	TrCh# C	TrCh# D
RM Attribute	200	190	235	160
Channel Coded Blocks	303	333	136	516
Rate Matched	312	328	166	428

III. FULL RATE AMR 12.2KBPS VOICE OVER S-UMTS ( $D_{cr}=2$ )

The physical layer configuration for AMR voice service in regular UMTS (e.g., up to Rel-11) no longer meets the required data rate for AMR 12.2 kbps in S-UMTS. On the other hand, supporting full rate AMR (12.2kbps) is a requirement for many infra vendors and operators. The proposed solution to facilitate the support of CS voice over S-UMTS uses

- a) SF reduction
- b) TTI inverse scaling

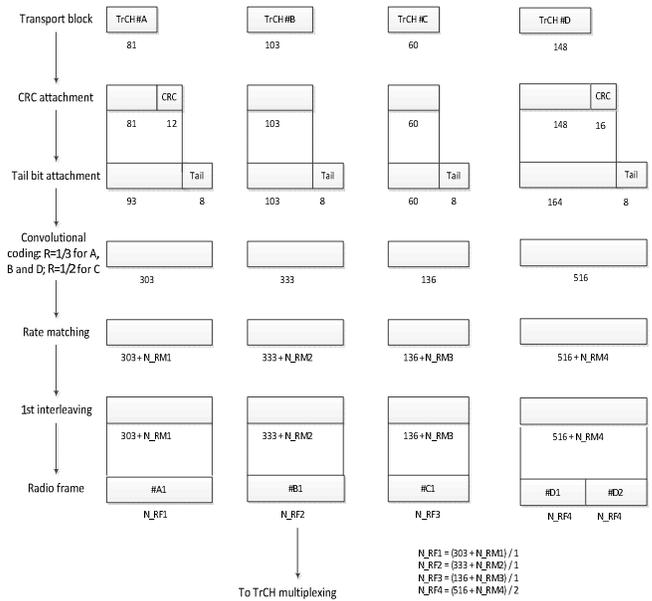


Figure 4. TrCH procedure for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in  $D_{cr}=2$  S-UMTS

Fig. 4 shows the transport channel procedure for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in  $D_{cr}=2$  S-UMTS. For 12.2 kbps AMR, one voice frame is still mapped to a 20 ms time window upon transmission, irrespective of  $D_{cr}$  in S-UMTS. To meet the requirements, SF reduction by  $D_{cr}$  in both DL and UL is proposed along with the following mapping

- 4 SRBs → 4 DCCHs → 1 DCH using  $20 \times D_{cr}$ , i.e.,  $20 \times 2$  ms = 40 ms TTI
- 1 RAB → 3 DTCHs → 3 DCHs using  $10 \times D_{cr}$ , i.e.,  $10 \times 2$  ms = 20 ms TTI

It is to be noted, that  $10 \times D_{cr}$  ms and  $20 \times D_{cr}$  ms TTI in S-UMTS correspond to 10 ms and 20 ms TTI respectively in regular UMTS except the time dilation in S-UMTS by  $D_{cr}$ .

TABLE VIII. TRANSPORT CHANNEL PARAMETER FOR CONVERSATIONAL/SPEECH/UL/DL 12.2 KBPS/CS RAB

RAB/Signalling RB		RAB subflow #1	RAB subflow #2	RAB subflow #3
TB sizes, bit		39, 81 (alt. 0, 39, 81)	103	60
TFS	TF0, bits	0x81	0x103	0x60
	TF1, bits	1x39	1x103	1x60
	TF2, bits	1x81	N/A	N/A
TTI, ms		10xDcr=10x2	10xDcr=10x2	10xDcr=10x2
Coding type		CC 1/3	CC 1/3	CC 1/2
CRC, bit		12	N/A	N/A
Max number of bits/TTI after channel coding		303	333	136
Uplink: Max number of bits/radio frame before rate matching		303	333	136
RM attribute		180 to 220	170 to 210	215 to 256

TABLE IX. TRANSPORT CHANNEL PARAMETER FOR DL/UL 3.4 KBPS SRBS FOR DCCH

RAB/Signalling RB		SRB#1	SRB#2	SRB#3	SRB#4
User of radio bearer		RRC	RRC	NAS_DT High prio	NAS_DT Low prio
TB sizes, bit		148 (alt 0, 148)			
TFS	TF0, bits	0x148 (alt 1x0)			
	TF1, bits	1x148			
TTI, ms		20xDcr=20x2			
Coding type		CC 1/3			
CRC, bit		16			
Max number of bits/TTI after channel coding		516			
Uplink: Max number of bits/radio frame before rate matching		258			
UL RM attribute		155 to 185			
DL RM attribute		155 to 230			

Let  $SF_{UMTS}$  be the SF for AMR 12.2 kbps in regular UMTS. According to [1], Section 6.10.2.4.1.4,  $SF_{UMTS,ULDPDCH} = 64$  and  $SF_{UMTS,DLDPDCH} = 128$ . For  $D_{cr} = 2$  S-UMTS, SF is calculated as in (1) for both UL and DL so that the number of channel bits per 20 ms stay unchanged.

$$SF_{S-UMTS} = \frac{SF_{UMTS}}{D_{cr}} \quad (1)$$

However,  $SF_{UMTS,ULDPDCH} = 256$  for both UMTS and S-UMTS. During the radio frame segmentation,  $10 \times D_{cr}$ , i.e.,  $10 \times 2 = 20$  ms AMR TTIs fit into  $10 \times D_{cr}$ , i.e.,  $10 \times 2$  ms = 20 ms radio frames while the  $20 \times D_{cr}$ , i.e.,  $20 \times 2$  ms = 40 ms DCCH TTI are broken into two 20 ms radio frames. TABLE VIII and TABLE IX show the transport channel parameters in UL and DL for CS RABs and SRBs respectively. In DL, DPCCH and DPDCH are time multiplexed as DPCH, and use the same SF.

The modified version of slot format 8 for  $D_{cr} = 2$  S-UMTS in TABLE X corresponds to slot format 8B for UMTS in

TABLE IV for DL DPCH in terms of spreading factor and bits per slot. For  $D_{cr} = 2$  S-UMTS, as there is less bandwidth while the timing and bit requirements of AMR 12.2 kbps are the same, the SF is reduced to increase the bits/slot and bits/frame accordingly. However, the channel bitrate or channel symbol rate stays unchanged as time is dilated for  $D_{cr} = 2$  S-UMTS. The RM tuning method also stays unchanged in  $D_{cr} = 2$  S-UMTS as compared to regular UMTS. The slot format 8B in TABLE X does not correspond to any slot format in regular UMTS.

TABLE X. DL SLOT FORMAT – DPDCH AND DPCCH FIELDS FOR S-UMTS (DCR=2)

Slot Format	Channel Bitrate (kbps)	Channel Symbol Rate (ksps)	SF	Bits/Slot	DPCCH Bits/Slot NTPC, NTFCI, NPilot	DPDCH Bits/Slot NData1, NData2	Transmitted slots per radio frame
8	120/Dcr=60	60/Dcr=30	64	80	4,0,8	12, 56	15
8B	240/Dcr=120	120/Dcr=60	32	160	8,0,16	24, 128	8-14

TABLE XI. UL SLOT FORMAT – DPDCH FIELDS FOR S-UMTS (DCR=2)

Slot Format	Channel Bitrate (kbps)	Channel Symbol Rate (ksps)	SF	Bits/Frame	Bits/Slot	Ndata
2	120/Dcr=60	120/Dcr=60	32	1200	80	80
3	240/Dcr=120	240/Dcr=120	16	2400	160	160

TABLE XII. UL SLOT FORMAT – DPCCH FIELDS FOR UMTS

Slot Format	Channel Bitrate (kbps)	SF	Bits/Frame, Bits/Slot	NPilot, NTPC, NTFCI, NFBI	Transmitted slot per radio frame
0	15/Dcr=7.5	256	150, 10	6, 2, 2, 0	15
0A	15/Dcr=7.5	256	150, 10	5, 2, 3, 0	10-14
0B	15/Dcr=7.5	256	150, 10	4, 2, 4, 0	8-9

The slot formats 2 and 3 for  $D_{cr} = 2$  S-UMTS in TABLE XI correspond to slot formats 3 and 4 respectively for UMTS in TABLE V for UL DPDCH. For  $D_{cr} = 2$  S-UMTS, as there is less bandwidth while the timing and bit requirements of AMR 12.2 kbps are the same, the SF is reduced to increase the bits/slot and bits/frame accordingly. However, the channel bitrate or channel symbol rate stays unchanged as time is dilated for  $D_{cr} = 2$  S-UMTS. The RM tuning method also stays unchanged. The slot formats 0, 0A and 0B for  $D_{cr} = 2$  S-UMTS in TABLE XII correspond to slot formats 0, 0A and 0B respectively for UMTS in TABLE VI for UL DPDCH. For UL DPCCH, all slot formats have SF 256. Hence, the bits/slot and bits/frame stays the same and channel bitrate and channel symbol rates are scaled down by  $D_{cr}$ . As a result, the Transmit Power Control (TPC) rate is reduced from 1500Hz in UMTS to  $1500/D_{cr}$ , i.e., 750 Hz in  $D_{cr} = 2$  S-UMTS.

Fig. 5 shows the TrCH multiplexing and mapping to physical channels for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” in  $D_{cr} = 2$  S-UMTS. The TrCHs for AMR fit into radio frames entirely while the TrCH for signaling is broken

from its TTI into  $10 \times D_{cr}$  ms, i.e., 20 ms frames. These frames are then serially multiplexed into a Coded Composite Transport Channel (CCTrCH). Thus, each CCTrCH contains an AMR Class-A, B, and C bit sequences, plus a portion of the signaling information. A downlink DPCH with  $120/D_{cr}$ , i.e.,  $120/2=60$  kbps is allocated to carry one CCTrCH. A second round of interleaving is done on the CCTrCh. The  $510 \times D_{cr}$ , i.e., 1020 coded bits per radio frame include CRC and tail bits, rate 1/2 or 1/3 bits-to-coded bits coding, and RM. Then the coded bits go through serial-to-parallel conversion and are placed onto a  $60/D_{cr}$ , i.e.,  $60/2=30$  kbps DPDCH.

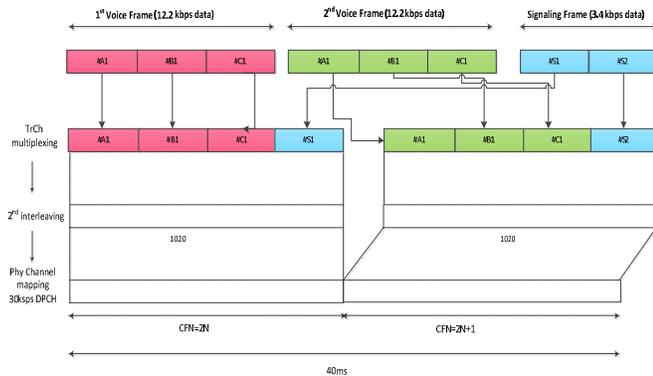


Figure 5. TrCH multiplexing and mapping to PhCH for “AMR DL: 12.2kbps RAB + DL: 3.4 kbps SRB” in  $D_{cr}=2$  S-UMTS.

It is to be noted that Figs. 4 and 5 are for one of the six TFCs as specified in [1], Section 6.10.2.4.1.4: ( $RAB$  subflow#1,  $RAB$  subflow#2,  $RAB$  subflow#3,  $DCCH$ ) = ( $TF2$ ,  $TF1$ ,  $TF1$ ,  $TF1$ )

RM is done before radio frame segmentation in the DL for UMTS. For  $D_{cr}=2$ , radio frame segmentation is not needed for TrCh# A, TrCh# B and TrCh# C. At this point, the number of bits are exactly equal for UMTS and  $D_{cr}=2$  S-UMTS as there is no difference in processing till radio frame segmentation. As a result, RM attributes stay the same for  $D_{cr}=2$  S-UMTS as in UMTS. For the RM example earlier shown in TABLE VII, the DL slot format is 8 (modified) with SF 64 (TABLE X). There are  $(12+56) \times 15=1020$  data bits in a  $10 \times D_{cr}$ , i.e., 20 ms radio frame. The RM attributes, the channel coded bits and bits after RM for the four transport channels stay the same as in TABLE VII. Using the TTI of each of the four transport channels, the above bits match 1020, i.e.,  $312 + 328 + 166 + 428/2 = 1020$ .

#### A. Transport Format Combination Indicator (TFCI)

##### a) Blind Transport Format Detection for $D_{cr}=2$ S-UMTS DL

In UMTS, Blind Transport Format Detection (BTDF) is used for “AMR DL: 12.2 kbps RAB + DL: 3.4 kbps SRB” as DL slot format 8 with SF 128 as shown in TABLE IV has no TFCI. BTDF is used in UMTS DL for AMR voice because if

the TFCI field is sent, the SF would have to be lower and lower SFs for voice provide a greater potential to run out of Orthogonal Variable Spreading Factor (OVSF) codes. For  $D_{cr}=2$ , the modified slot format 8 as shown in TABLE X also has no TFCI. Section 4.3.1 of [3] puts a number of restrictions that need to be fulfilled for BTDF. In order to allow BTDF for AMR 12.2 kbps for  $D_{cr}=2$  S-UMTS, the following restriction needs to be changed.

- the number of CCTrCH bits received per radio frame is  $600 \times D_{cr}$  or less instead of 600.

##### b) UL TFCI

There are 32 TFCI encoded bits that needs to be transmitted at least once every 20ms, i.e., every voice frame. TFCI is encoded using a (32, 10) sub-code of the second order Reed-Muller code [3]. If the TFCI consists of less than 10 bits, it is padded with zeroes to 10 bits, by setting the most significant bits to zero. The bits of the 32 bit TFCI code word are directly mapped to the slots of the radio frame. Within a slot the bit with lower index is transmitted before the bit with higher index. The coded bits  $b_k$  are mapped to the transmitted TFCI bits  $d_k$  according to the following formula:

$$d_k = b_k \text{ mod } 32 \quad (2)$$

For uplink physical channels regardless of the SF, bits  $b_{30}$  and  $b_{31}$  of the TFCI code word are not transmitted. Using slot format 0 in TABLE XII, there are  $2 \times 15 = 30$  TFCI encoded bits every frame duration of  $10 \times D_{cr}$ , i.e., 20ms. Therefore, bits  $b_{30}$  and  $b_{31}$  of the TFCI code word are not transmitted. For uplink compressed mode, the slot format is changed so that no TFCI coded bits are lost [3]. The different slot formats in compressed mode do not match the exact number of TFCI coded bits for all possible Transmission Gap Lengths (TGL). Repetition of the TFCI bits is therefore used.

#### B. TPC and Pilot

For the TPC and pilot bits in DL, the patterns defined in [2], are used. Similarly, for the TPC and pilot bits in UL, the patterns defined in [2], are used.

#### C. Transmission Power

Generally, a reduction in SF has impacts on the link budget. In order to maintain the same voice service coverage, the UE and NodeB need to increase the transmission power according to the chosen SF. The baseline assumption is that the Power Spectral Density (PSD) remains the same in both uplink and downlink for UMTS and S-UMTS. Since the S-UMTS bandwidth is  $1/D_{cr}$  of that of regular UMTS, the total transmit power of S-UMTS is also  $1/D_{cr}$  of the transmit power of regular UMTS. However, for supporting AMR voice, this has been relaxed. In the DL, the total transmit power of S-UMTS

is still assumed to  $1/D_{cr}$  of the transmit power of regular UMTS and hence, same PSD remains for S-UMTS. However, the transmit power for each voice channel is same in S-UMTS and in regular UMTS. In the UL, the UE transmit power in S-UMTS is assumed to be the same as that in regular UMTS. In other words, the transmit power in S-UMTS for AMR in both UL and DL is increased by  $10\log_{10}(D_{cr})$  dB w.r.t. same PSD to compensate for SF reduction.

#### D. Latency

Once the first voice frame (equivalently 20 ms long) is available at the MAC, it will be delivered to the PHY. After some PHY layer processing (assuming processing time does not scale with  $D_{cr}$ ), the over-the-air transmission is allowed to start only at the next radio frame boundary due to the current specification restriction. However, for  $D_{cr}=2$  S-UMTS, as the TTI is  $10 \times D_{cr} = 20$  ms and the radio frame is also  $10 \times D_{cr} = 20$  ms, there is no additional latency compared to regular UMTS. For  $D_{cr} > 2$ , there would be additional latency due to TTI ( $10 \times D_{cr}$ ) being greater than 20 ms.

### IV. VOICE CAPACITY

The voice capacity in the uplink and downlink are compared between UMTS and S-UMTS in *A* and *B* respectively.

#### A. Uplink Capacity

The upper bound of the uplink capacity, referred to as the pole capacity ( $N_{pole}$ ) of a UMTS carrier, can be estimated using the standard uplink capacity equation [8]

$$N_{pole} = \frac{W/R_b}{E_b/N_t * v * (1+\alpha)} \quad (3)$$

where

$W$  = Spreading bandwidth of the system

$R_b$  = RAB rate for selected application, e.g., AMR 12.2 kbps

$E_b/N_t$  = Required energy per bit to total noise spectrum density ratio

$v$  = Voice Activity Factor (VAF)

$\alpha$  = Interference factor (ratio between the other cell interference power and the total received signal power of users in the cell)

Note: VAF depends on the vocoder, channel coding etc. and is typically considered as 60%

The pole capacity is obtained by assuming that the user equipment (UE) has infinite transmission power and the interference at the Node-B receiver goes to infinity. For a practical system, both the UE transmission power and the node-B receiver allowed interference level are limited and the operating point is set well below the pole capacity. This operating point is referred to as the uplink loading and is

defined as the percentage of the pole capacity. From the pole capacity, a practical cell capacity ( $N_{user}, \eta$ ) for a system can be calculated after the uplink loading ( $\eta$ ) has been determined, as shown in (3).

$$N_{user} = N_{pole} * \eta \quad (4)$$

The link efficiency for UMTS is given by the following equation

$$\frac{E_b}{N_t} = \frac{E_c}{N_t} + 10\log_{10}\left(\frac{R_c}{R_b}\right) \quad (5)$$

where

- $R_b$  is the UMTS traffic rate
- $R_c$  is the UMTS chip rate

For AMR over S-UMTS,

- Chip rate is  $R_c/D_{cr}$
- $R_b$  is same as in UMTS

$$\begin{aligned} \left(\frac{E_b}{N_t}\right)_{S-UMTS} &= \left(\frac{E_c}{N_t} * \frac{R_c}{R_b}\right)_{S-UMTS} \\ &= \left(\frac{D_{cr} * E_c}{N_t} * \frac{R_c/D_{cr}}{R_b}\right)_{UMTS} = \left(\frac{E_b}{N_t}\right)_{UMTS} \end{aligned} \quad (6)$$

The link efficiency for S-UMTS is same as the link efficiency for UMTS as seen from (6). Hence, S-UMTS maintains the coverage of UMTS. Therefore, in (3) for pole capacity in uplink, everything remain same for S-UMTS as in UMTS except  $W$  that changes from 3.84 MHz in regular UMTS to 1.92 MHz for  $D_{cr}=2$  S-UMTS. The uplink loading ( $\eta$ ) is also assumed to be same for S-UMTS. As a result, for the same loading factor, the uplink cell capacity for voice services is inversely proportional to the value of  $D_{cr}$  according to (3) and (4).

#### B. Downlink Capacity

In the downlink as well, the pole capacity can be interpreted as the maximum capacity with infinite base station power. The following equation is used or the downlink pole capacity

$$N_{pole} = \frac{(1-n_{OH}) * W/R_b}{E_b/N_t * v * (\delta + \beta)} \quad (7)$$

where

$n_{OH}$  = percentage of overhead channel power

$W$  = Spreading bandwidth of the system

$R_b$  = RAB rate for CS voice

$E_b/N_t$  =Required energy per bit to total noise spectrum density ratio

$v = \text{VAF}$

$\beta$  = Interference factor (ratio between the other cell interference power and the total downlink cell transmission power at UE receiver)

$\delta$  = Orthogonally factor (percentage of the serving cell signal that becomes the interference at the UE receiver due to the multipath effect and the limitations of the rake receiver)

Note: Cell geometry is the inverse of  $\beta$

The overhead channel power is the power for the overhead channels in UMTS and S-UMTS, e.g., Primary and Secondary Synchronization Channels, Primary Common Pilot Channel, Primary Common Control Physical Channel, Secondary Common Control Physical Channel, Paging Indicator Channel etc. In S-UMTS, the overhead scales with the bandwidth and hence, S-UMTS has the same signaling overhead percentage as UMTS.

TABLE XIII. SCALABLE UMTS PARAMETERS FOR AMR

	Regular UMTS (Dcr=1)	Half BW UMTS (Dcr=2)
Nominal Bandwidth (MHz)	5	5/Dcr=2.5
Chip Rate (Mcps)	3.84	3.84/Dcr=1.92
Radio frame duration (ms)	10	10xDcr=20
Data Rate	D	D
Power	P	P
Range	R	R
Capacity	C	C/Dcr=C/2

In (7), everything except W remains same for S-UMTS as in UMTS. Hence, the downlink cell capacity for voice services is inversely proportional to the value of  $D_{cr}$ . Thus, a trade-off can be achieved in S-UMTS between data rate, power and capacity. For services that need to maintain the same data rate over S-UMTS as in UMTS, the capacity is scaled down by  $D_{cr}$

in S-UMTS while for other services, the capacity is unchanged but data rate is scaled down by  $D_{cr}$  in S-UMTS as shown TABLE I and TABLE XIII.

V. CONCLUSION

S-UMTS carriers occupy different bandwidth than regular UMTS carriers and may fit in spectrum where regular UMTS carrier cannot fit. Solutions for supporting full rate AMR (12.2 Kbps) over  $D_{cr}=2$  S-UMTS have been proposed with no additional delay. As S-UMTS uses the same vocoder, it offers the same voice quality as regular UMTS. The proposed solution can be extended for other AMR rates and also for other S-UMTS  $D_{cr}$  values. It is also shown that the voice capacity for S-UMTS scales with the bandwidth. As the system bandwidth in S-UMTS is scaled by the bandwidth scaling factor  $D_{cr}$ , S-UMTS UL and DL voice capacity are also scaled by the same factor  $D_{cr}$  compared to regular UMTS.

REFERENCES

- [1] 3GPP TS34.108, Common test environments for user equipment (UE); conformance testing.
- [2] 3GPP TS 25.211, Physical channels and mapping of transport channels onto physical channels (FDD).
- [3] 3GPP TS 25.212, Multiplexing and channel coding (FDD).
- [4] 3GPP TS 25.213, Spreading and modulation (FDD).
- [5] 3GPP TS 25.331 Radio Resource Control (RRC) protocol specification.
- [6] 3GPP TS25.101 User Equipment (UE) radio transmission and reception (FDD).
- [7] 3GPP TS 25.104 Base Station (BS) radio transmission and reception (FDD).
- [8] H. Holma and A. Toskala, "WCDMA for UMTS - HSPA Evolution and LTE", 5<sup>th</sup> ed, Wiley, pp. 313-314, 2010.

# Non-frequency-selective I/Q Imbalance in Zero-IF Transceivers for Wide-Band mmW Links

Ainhoa Rezola, Juan Francisco Sevillano,  
Roc Berenguer, Igone Vélez

Centro de Estudios e Investigaciones Técnicas (CEIT)  
San Sebastián, Spain  
Email: {argarciandia,jfsevillano,  
rberenguer,ivelez}@ceit.es

Martin Leyh, Moises Lorenzo,  
Aharon Vargas

Fraunhofer Institute for Integrated Circuits (IIS)  
Erlangen, Germany  
Email: {martin.leyh,moises.lorenzo,  
aharon.vargas}@iis.fraunhofer.de

**Abstract**—Millimeter wave (mmW) links are an attractive solution for backhaul of mobile networks. In order to cope with the requirements of future networks, these mmW links should achieve Gigabit data rates. These data rates can be achieved using wide-band and high order modulations in E-Band. Zero-IF architectures are good candidates for integrated transceivers. However, the design of integrated transceivers at these frequencies is a challenging issue. An important source of degradation is I/Q imbalance, which can significantly reduce the performance of a communication system with zero-IF transceivers if it is not appropriately compensated. In this paper, the impact of the I/Q imbalance impairment on the transmitted and received signal is analyzed and suitable digital signal processing techniques are evaluated for I/Q imbalance compensation at the receiver for a 64-QAM system using 2GHz bandwidth.

**Keywords**—Mobile backhaul; millimeter-wave; transceivers; RF impairments.

## I. INTRODUCTION

The growing demand for ubiquitous broadband communication, e.g., fourth-generation (4G) wireless, has motivated the deployment of ultra high-speed communication systems. Especially in backhauling networks, optical fiber is required to transport very high data rates. However, optical fiber exhibits important drawbacks, such as high costs, long deployment times, and low flexibility. Recently, point-to-point wireless communication systems have been proposed as an attractive alternative to optical fiber. In order to achieve comparable data rates as the optical fiber, these communication systems demand a very high bandwidth to transport enough data. Although the frequency spectrum is congested, the regulation of the E-band facilitates the deployment of high-speed communication systems. The European Telecommunications Standards Institute (ETSI) is carrying out a standardization process for this frequency band [1][2].

Commercial off-the-shelf communication systems operating in the E-band support data rates up to 2.5 Gbit/s. However, new applications demand even higher data rates, which necessitate both wide-band and high-order modulations to utilize the spectrum efficiently. Higher order modulations require high Carrier-to-Interference (C/I) ratios at the receiver involving careful analysis of the degradation effects introduced by the analog Radio-Frequency (RF) impairments and evaluation of corresponding compensation algorithms in the digital baseband processing [3].

In this paper, we focus on the I/Q imbalance impairment caused by the local oscillators used for quadrature modulation and demodulation in zero-IF transceivers. In an ideal quadrature modulator or demodulator, the mirror images are completely suppressed. I/Q imbalance entails a degradation in the Image Rejection Ratio (IRR) and causes interfering images at mirror frequencies [4]. The paper analyzes the impact of these local oscillator imbalances in the transmitted and received signal. Building upon the results of this analysis, an approach to compensate for both transmitter and receiver induced I/Q imbalance by digital signal processing techniques is selected and its performance for a 64-QAM transceiver operating with a signal bandwidth of 2GHz is evaluated.

Section II introduces the I/Q imbalance issue identifying and modeling the source of this impairment. The mitigation of the I/Q imbalance impairment by digital signal processing at the receiver is described in Section III. In Section IV, the selected mechanism to cope with the I/Q imbalance is simulated and analyzed. Finally, some conclusions are drawn in Section V.

## II. SYSTEM ANALYSIS

### A. Transceiver architecture

In order to address new applications for the future backhauling networks, a point-to-point microwave link in the E-Band using a 64-QAM modulation with a signal bandwidth of 2GHz is considered. Figure 1 shows the proposed transceiver (TRX) architecture for a point-to-point microwave link in the E-Band. As shown, the transmitter (Tx) front-end consists of an I/Q up-converting modulator that up-converts the baseband I and Q channels to an Intermediate Frequency (IF). After combining the I and Q channels, the IF signal is up-converted by means of the millimeter-wave (mmW) mixer. The receiver (Rx) front-end consists of a wideband Low Noise Amplifier (LNA), which receives and amplifies the signal at the E-Band. After the LNA, a first mixer down-converts the mmW signal to the same IF as in the Tx. This way, the same PLL can be re-used for the Tx and the Rx. Finally, an I/Q demodulator down-converts the IF signal to 0-Hz.

This architecture presents a good balance between different design aspects and enables to minimize the sampling frequency of the Digital-to-Analog (DAC) and Analog-to-Digital (ADC) Converters. Nowadays, we can find commercial DACs and

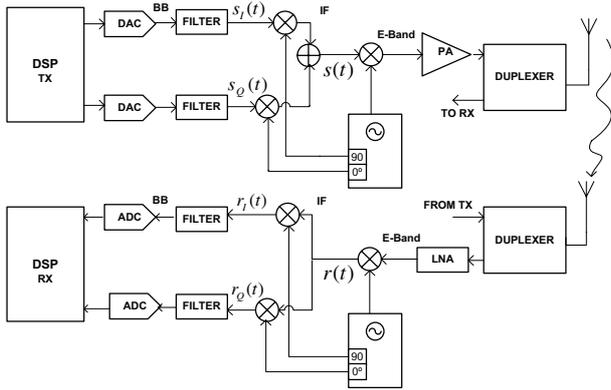


Figure 1. Architecture of the transceiver.

ADCs able to provide sampling rates close to 3Gps, which is enough for practical implementation of the zero-IF architecture. The use of other architectures such as low-IF would require very high performance DACs or ADCs, with sampling rates well above 4Gps to achieve a practical implementation of base-band and image rejection filters in the analog front-end.

However, this zero-IF architecture is subject to the corruption due to I/Q imbalances at the transmitter and receiver quadrature modulator and demodulator respectively. The resulting system performance degradation can be important, specially for high-order modulation schemes [5].

### B. I/Q Imbalance Analysis

The goal of the I/Q modulator in Figure 1 is to perform a frequency translation of the signal. That is, if the base-band input signal to the I/Q modulator is

$$\tilde{s}(t) = s_I(t) + js_Q(t), \quad (1)$$

where  $s_I(t)$  is the signal in the I-datapath and  $s_Q(t)$  is the signal in the Q-datapath, a perfect I/Q modulation mixes the base-band input signal with

$$l_{tx}(t) = e^{j\omega_{tx}t} = \cos(\omega_{tx}t) + j \sin(\omega_{tx}t) \quad (2)$$

producing an output signal

$$s(t) = \text{Re}[\tilde{s}(t)l_{tx}(t)] = s_I(t) \cos(\omega_{tx}t) - s_Q(t) \sin(\omega_{tx}t). \quad (3)$$

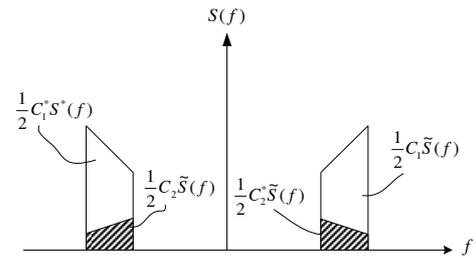
However, when implementing an I/Q modulator with actual electronic circuits, the signals produced by the Local Oscillator (LO) will present some difference in their amplitudes and will not have a phase difference of  $\pi/2$ . The effect of this imbalance can be modeled as the mixing of the base-band input signal with

$$l_{tx}(t) = \cos(\omega_{tx}t) + jg_{tx} \sin(\omega_{tx}t + \phi_{tx}) \quad (4)$$

to yield the output signal

$$s(t) = (s_I(t) - g_{tx} \sin(\phi_{tx})s_Q(t)) \cos(\omega_{tx}t) \quad (5a)$$

$$- s_Q(t)g_{tx} \cos(\phi_{tx}) \sin(\omega_{tx}t). \quad (5b)$$


 Figure 2. Spectrum of  $s(t)$  with imbalance.

In I/Q imbalance analysis, it is common to rewrite the signal produced at the transmitter LO with imbalance of (4) in the form [6]

$$l_{tx}(t) = C_1 e^{j\omega_{tx}t} + C_2 e^{-j\omega_{tx}t}, \quad (6)$$

with

$$C_1 = \frac{1 + g_{tx} e^{j\phi_{tx}}}{2} \quad (7a)$$

$$C_2 = \frac{1 - g_{tx} e^{-j\phi_{tx}}}{2} \quad (7b)$$

and the transmitted signal is

$$s(t) = \text{Re}[\tilde{s}(t) (C_1 e^{j\omega_{tx}t} + C_2 e^{-j\omega_{tx}t})]. \quad (8)$$

In (8), the desired term is the one multiplied by  $e^{j\omega_{tx}t}$  and the term multiplied by  $e^{-j\omega_{tx}t}$  is considered an undesired image. Working on (8) we can rewrite it as

$$s(t) = \frac{1}{2} (C_1 \tilde{s}(t) e^{j\omega_{tx}t} + C_1^* \tilde{s}^*(t) e^{-j\omega_{tx}t}) \quad (9a)$$

$$+ \frac{1}{2} (C_2^* \tilde{s}^*(t) e^{j\omega_{tx}t} + C_2 \tilde{s}(t) e^{-j\omega_{tx}t}), \quad (9b)$$

where  $(\cdot)^*$  denotes the complex conjugate.

Let  $X(\omega)$  denote the Fourier Transform of a signal  $x(t)$ , then from (9) we have

$$S(\omega) = \frac{1}{2} (C_1 \tilde{S}(\omega - \omega_{tx}) + C_1^* \tilde{S}^*(\omega + \omega_{tx}) \quad (10a)$$

$$+ C_2^* \tilde{S}^*(\omega - \omega_{tx}) + C_2 \tilde{S}(\omega + \omega_{tx})) \quad (10b)$$

Figure 2 illustrates the spectrum of  $s(t)$ . The first line in the above equation is the desired term and the second line is an image that aliases on the desired signal. A measure of performance of the I/Q modulator is the Image Rejection Ratio (IRR)

$$\text{IRR}_{tx} = \frac{|C_1|^2}{|C_2|^2} = \frac{1 + g_{tx}^2 + 2g_{tx} \cos(\phi_{tx})}{1 + g_{tx}^2 - 2g_{tx} \cos(\phi_{tx})}. \quad (11)$$

Similarly, the task of the Rx I/Q demodulator in Figure 1 is to mix the input signal with

$$l_{rx}(t) = e^{-j\omega_{rx}t} = \cos(\omega_{rx}t) - j \sin(\omega_{rx}t), \quad (12)$$

so that after low-pass filtering, we get the base-band equivalent of the signal in the frequency band of interest  $\tilde{z}(t) = z_I(t) + jz_Q(t)$ . Note that  $\tilde{z}(t)$  is the base-band equivalent referred to a carrier frequency  $\omega_{rx}$ . The actual implementation of the I/Q

demodulator will introduce similar imbalances to the actual implementation of the I/Q modulator, which can be modeled as the mixing of  $r(t)$  with

$$l_{rx}(t) = \cos(\omega_{rx}t) - jg_{rx} \sin(\omega_{rx}t + \phi_{rx}). \quad (13)$$

The signal at the output of the I/Q demodulator after low-pass filtering can be written as [6]

$$\tilde{r}(t) = r_I(t) + jr_Q(t) = K_1\tilde{z}(t) + K_2\tilde{z}^*(t) \quad (14)$$

with

$$K_1 = \frac{1 + g_{rx}e^{-j\phi_{rx}}}{2} \quad (15a)$$

$$K_2 = \frac{1 - g_{rx}e^{j\phi_{rx}}}{2}. \quad (15b)$$

In this case, the first term in the sum of (14) is the desired term and the second one is the image that aliases on the desired signal. The IRR for the I/Q demodulator is defined as

$$\text{IRR}_{rx} = \frac{|K_1|^2}{|K_2|^2} = \frac{1 + g_{rx}^2 + 2g_{rx} \cos(\phi_{rx})}{1 + g_{rx}^2 - 2g_{rx} \cos(\phi_{rx})}. \quad (16)$$

Note that the above model for the impairment affects in the same way to the whole information bearing signal. Thus, the I/Q imbalance is said to be Non-Frequency-Selective (NFS). Other imbalances in the in-phase and quadrature datapaths of the information bearing signals can introduce Frequency-Selective (FS) I/Q imbalance. In this paper, we restrict our attention to the NFS component of the I/Q imbalance. Although the model has been developed from the point of view of the local oscillators of the I/Q modulator and demodulator, it can also be used to include the mean imbalances between the in-phase and quadrature datapaths.

In order to gain some insight, we assume enough linearity and proper filtering in the remaining stages of the transmitter and receiver analog chain and noise-less operation. Using (9), it can be seen that

$$\tilde{z}(t) = (C_1\tilde{s}(t) + C_2^*\tilde{s}^*(t)) e^{j(\Delta\omega t + \theta)} \quad (17)$$

where  $\Delta\omega$  and  $\theta$  account for the overall carrier frequency and phase offset between the transmitter and the receiver. Thus, we have

$$\tilde{r}(t) = K_1C_1\tilde{s}(t)e^{j(\Delta\omega t + \theta)} \quad (18a)$$

$$+ K_1C_2^*\tilde{s}^*(t)e^{j(\Delta\omega t + \theta)} \quad (18b)$$

$$+ K_2C_1^*\tilde{s}^*(t)e^{-j(\Delta\omega t + \theta)} \quad (18c)$$

$$+ K_2C_2\tilde{s}(t)e^{-j(\Delta\omega t + \theta)} \quad (18d)$$

The desired term in (18) is the one in the first line and the terms in the second to fourth line represent undesired images at the receiver due to transmitter and receiver I/Q imbalances.

When  $\Delta\omega = 0$ , (18) simplifies to

$$\tilde{r}(t) = J_1\tilde{s}(t) + J_2\tilde{s}^*(t) \quad (19)$$

where  $J_1$  and  $J_2$  are constants given by

$$J_1 = K_1C_1e^{j\theta} + K_2C_2e^{-j\theta} \quad (20a)$$

$$J_2 = K_1C_2^*e^{j\theta} + K_2C_1^*e^{-j\theta}. \quad (20b)$$

Comparing (19) with (14), it can be concluded that when  $\Delta\omega = 0$ , the effect observed at the output of the receiver's

I/Q demodulator due to the I/Q imbalance introduced at the transmitter is the same as the one due to an I/Q imbalance introduced by the I/Q demodulator. In a real transmission system there will be some carrier frequency offset between the transmitter and the receiver. However, the former observation suggests that the I/Q imbalance introduced at the transmitter may be addressed after carrier frequency recovery using approaches designed to address the I/Q imbalance introduced at the receiver.

### III. I/Q IMBALANCE COMPENSATION

#### A. Tx I/Q Imbalance Compensation

Using (1) and (3) with  $\tilde{s}'(t) = s'_I(t) + js'_Q(t)$  denoting the equivalent baseband signal of the Tx I/Q distorted signal  $s(t)$  with respect to the transmitter carrier frequency  $\omega_{tx}$ , we can derive the following matrix equation for Tx I/Q imbalance distortion:

$$\begin{bmatrix} s'_I(t) \\ s'_Q(t) \end{bmatrix} = \begin{bmatrix} 1 & -g_{tx} \sin(\phi_{tx}) \\ 0 & g_{tx} \cos(\phi_{tx}) \end{bmatrix} \begin{bmatrix} s_I(t) \\ s_Q(t) \end{bmatrix}. \quad (21)$$

If the matrix in (21) is invertible ( $g_{tx} \neq 0$  and  $\phi_{tx} \neq \pm\pi/2$ ), what is the case for practical cases, the Tx NFS I/Q compensation can ideally be achieved by performing a digital predistortion based on the inverse operation. In this case, we feed the I/Q modulator with the predistorted signal  $\zeta(t) = \zeta_I(t) + j\zeta_Q(t)$ , which is obtained as

$$\begin{bmatrix} \zeta_I(t) \\ \zeta_Q(t) \end{bmatrix} = \begin{bmatrix} 1 & \tan(\phi_{tx}) \\ 0 & 1/(g_{tx} \cos(\phi_{tx})) \end{bmatrix} \begin{bmatrix} s_I(t) \\ s_Q(t) \end{bmatrix}, \quad (22)$$

Compensation by inverse transformation requires knowledge of the gain and phase imbalance values,  $g_{tx}$  and  $\phi_{tx}$ .

Techniques for compensation of the transmitter I/Q imbalance at the transmitter have been proposed in the literature using tones as test signals (e.g., [7][8][9][10]) or even from random data (e.g., [8][11]). The compensation using test tones is very powerful and can be used for initial calibration. During normal full-duplex operation of the transceiver, I/Q imbalance compensation from the random transmitted data would be preferred. All these techniques of compensation require additional circuitry (including an extra ADC) at the transmitter to feedback measurements performed in the analog front-end.

Under certain conditions Tx I/Q imbalance can also be compensated in the receiver [6]. For this to be possible, an important issue is that the spectral images caused by Tx I/Q imbalance have to be emitted. This condition is fulfilled in the analyzed system. Receiver-based compensation of Tx I/Q imbalance is the approach further investigated in this paper, because it does not require any additional circuitry in the analog front-end of the transceiver and all the compensation can be performed by digital signal processing.

#### B. Rx I/Q Imbalance Compensation

The down-converted complex-valued base-band signal  $\tilde{r}(t)$  can be written as a function of  $\tilde{z}(t)$  using the following matrix equation

$$\begin{bmatrix} r_I(t) \\ r_Q(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -g_{rx} \sin(\phi_{rx}) & g_{rx} \cos(\phi_{rx}) \end{bmatrix} \begin{bmatrix} z_I(t) \\ z_Q(t) \end{bmatrix}. \quad (23)$$

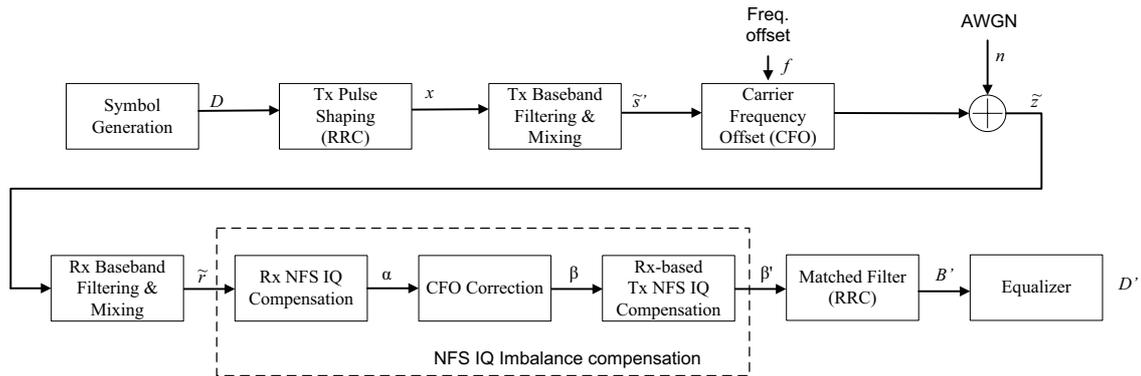


Figure 3. System Model for I/Q Imbalance Simulations.

In case the matrix in (23) is invertible ( $g_{rx} \neq 0$  and  $\phi_{rx} \neq \pm\pi/2$ ), what is the case for practical cases, the Rx I/Q imbalance can be ideally compensated by performing the inverse operation

$$\begin{bmatrix} \alpha_I(t) \\ \alpha_Q(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \tan(\phi_{rx}) & 1/(g_{rx} \cos(\phi_{rx})) \end{bmatrix} \begin{bmatrix} z'_I(t) \\ z'_Q(t) \end{bmatrix}, \quad (24)$$

where  $\alpha(t) = \alpha_I(t) + j\alpha_Q(t)$  is the output of the Rx I/Q imbalance compensator and ideally would yield the desired baseband signal  $\tilde{z}(t)$ . Compensation by inverse transformation requires knowledge of the gain and phase imbalance values  $g_{rx}$  and  $\phi_{rx}$ , which can be derived by using statistics and correlation properties of the I and Q signals as proposed in [12].

#### IV. SIMULATION RESULTS

##### A. System Model

Figure 3 depicts the simulation model developed in Matlab for the analysis of the I/Q imbalance in the transceiver performance. Random data information is generated as a sequence of I/Q symbols,  $D$ , by using a 64-QAM mapper.  $D$  is then filtered through an appropriate Root-Raised-Cosine (RRC) filter to create a pulse-shaped base-band signal  $x$ . The Tx Baseband Filtering and Mixing models the data processing in the transmitter analog front-end shown in Figure 1, including the NFS-I/Q imbalance at the I/Q modulator.

Considering an Additive White Gaussian Noise (AWGN) channel, the signal  $\tilde{z}$  at the output of the channel model is given by:

$$\tilde{z}(t) = \tilde{s}'(t) \cdot e^{j2\pi ft} + n(t), \quad (25)$$

where  $\tilde{z}(t)$  is the low-pass equivalent of the transmitted signal, with  $f$  the Carrier Frequency Offset (CFO) in Hz. Finally,  $n(t)$  corresponds to a complex-valued white Gaussian noise process.

On the receiver side, the Rx Baseband Filtering and Mixing models the receiver analog front-end structure shown in Figure 1 including Rx I/Q imbalance. NFS I/Q Imbalance Compensation is performed prior to matched filtering following a multi-stage approach that compensates for both Tx and Rx I/Q Imbalance:

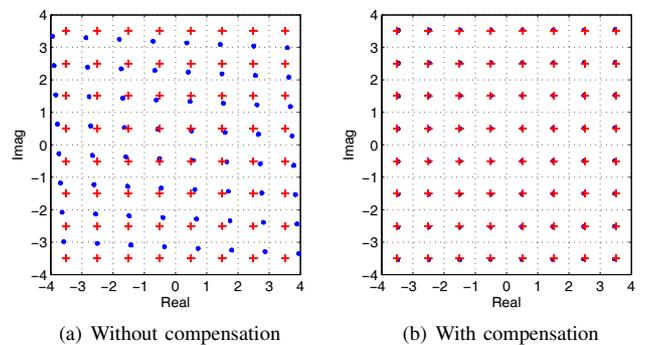


Figure 4. NFS I/Q imbalance

- 1) The first stage compensates for the Rx NFS I/Q imbalance based on the algorithms in Section III.
- 2) A CFO correction stage compensates for the CFO.
- 3) A third stage is employed for performing Tx NFS I/Q imbalance compensation prior to the matched filter by applying again the algorithms in Section III.

The NFS I/Q compensated signal,  $\beta'$ , is then filtered with a matched RRC filter in order to recover the transmitted information. Due to the high bandwidth of the signal and limited sampling rates of the DACs and ADCs, there is a frequency selectivity in the analog chain of the system. So as to compensate for this frequency response, an equalizer modelled as a Wiener filter [13] has been included at the receiver. The received symbols  $D'$  are then compared with the transmitted symbols  $D$  to quantify the performance of the whole system.

##### B. Results

Figure 4(a) illustrates the effect of NFS I/Q imbalance. For illustration purpose, a noiseless transmission is considered. Crosses correspond to the constellation when perfect transmission takes place, and the dots correspond to a transmission with I/Q imbalances both at transmitter and receiver assuming zero carrier frequency offset. The I/Q imbalances considered were 0.5 dB and 3 degrees in gain and phase, respectively, in the transmitter and 1 dB and 3 degrees in gain and phase, respectively, in the receiver. The signs of the imbalances

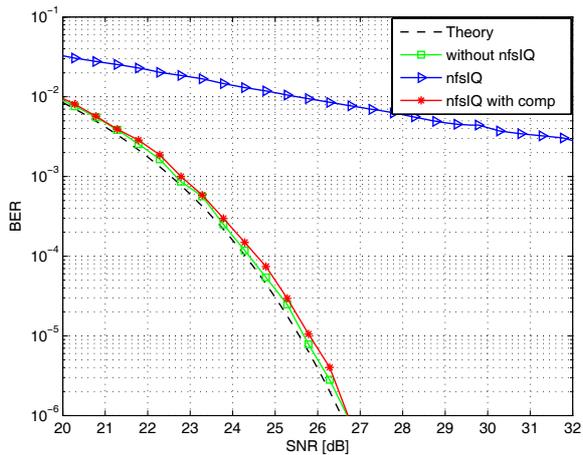
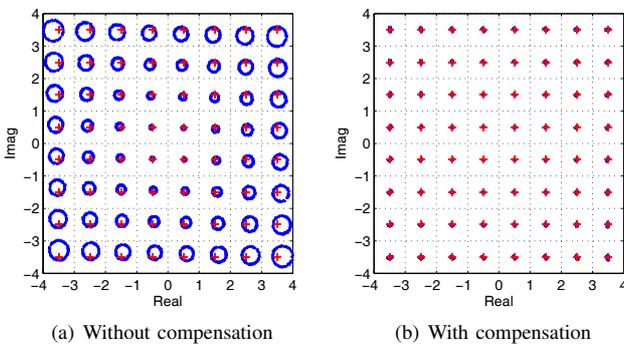


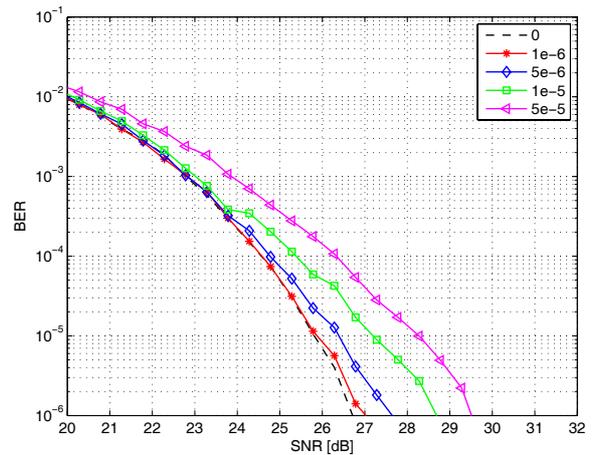
Figure 5. System performance in the presence of NFS I/Q imbalance.


 Figure 6. NFS I/Q imbalance with normalized residual carrier frequency offset  $\Delta\omega T/(2\pi) = 5 \cdot 10^{-6}$ .

were selected in the transmitter and the receiver so that they combine in the worst possible distortion. Figure 4(b) shows the constellation at the receiver when the NFS I/Q imbalance compensation is active. As it can be seen, the NFS I/Q imbalance compensation approach is able to remove the distortion of the constellation.

Figure 5 shows the impact of the NFS I/Q imbalance in the performance of the transceiver. The curve labeled ‘without NFS IQ’ is the performance of the transceiver when there is no I/Q imbalance. As it is shown in the figure the equalizer used to compensate the frequency selectivity of the analog chain introduces minor losses. The curve labeled ‘NFS IQ’ is the performance of the transceiver with I/Q imbalance at both transmitter and receiver. The curve labeled ‘NFS IQ with comp’ is the performance when the I/Q imbalance compensation is active. The NFS I/Q imbalance compensation is able to reduce the losses to a few tenths of a dB.

Figure 6(a) illustrates the effect of NFS I/Q imbalance when there is a residual carrier frequency offset equal to  $\Delta\omega = 2\pi \cdot 5 \cdot 10^{-6}/T$  after the CFO correction in Figure 3, with  $T$  the symbol period. For proper symbol detection and BER estimation, the residual carrier frequency offset has been compensated in the simulations at the input of the receiver’s matched-filter. The figure shows the corrected constellations


 Figure 7. Performance of NFS I/Q imbalance compensation approach for different values of  $\Delta\omega T/(2\pi)$ .

after this final residual carrier frequency offset compensation. The transmitter’s I/Q imbalance manifests itself as rotations of the constellation around the distorted constellation due to the receiver’s I/Q imbalance.

Figure 6(b) shows the constellation when the I/Q compensation algorithms are active for the same residual carrier frequency offset. It can be seen that a reduction in the distortion of the constellation has been achieved despite some residual carrier frequency offset.

Figure 7 shows the impact of residual carrier frequency offset on the performance of the NFS I/Q imbalance compensation approach presented in Section III. The different curves correspond to different values of the normalized residual carrier frequency offset  $\Delta\omega T/(2\pi)$ . It can be seen that a very accurate carrier frequency offset correction is needed for proper compensation of the transmitter I/Q imbalance with residual CFO below  $\Delta\omega T/(2\pi) = 5 \cdot 10^{-6}$ , what can be achieved via application of state of the art coarse and fine frequency synchronization algorithms [14].

## V. CONCLUSION

In this paper, a theoretical analysis of the NFS I/Q imbalance in a zero-IF transceiver has been presented. Digital signal processing at the receiver has been evaluated based on a multi-stage approach for the compensation of both transmitter and receiver I/Q imbalance, as well as carrier frequency offset.

Simulation results were presented for a zero-IF transceiver using 64-QAM with a signal bandwidth of 2GHz. The simulation results show that receiver-based Tx I/Q imbalance compensation can be achieved with negligible degradation in overall system performance. This is achieved when accurate frequency synchronization is performed in the receiver that reduces the residual carrier frequency offset below specified limits.

## ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community’s Framework Programme

FP7/2007-2013 under grant agreement no. 317957. Consortium: Ceit, FhG, ALU-I, CEA-Leti, INCIDE, SiR, ST-I, Silvers IMA, OTE.

#### REFERENCES

- [1] Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas; Part 1: Overview and system-independent common characteristics, ETSI EN 302 217-1, Sept. 2012.
- [2] Fixed Radio Systems; Characteristics and requirements for point-to-point equipment and antennas; Part 2-2: Digital systems operating in frequency bands where frequency co-ordination is applied; Harmonized EN covering the essential requirements of article 3.2 of the R&TTE Directive, ETSI EN 302 217-2-2, Sept. 2012.
- [3] G. Fettweis et al., "Dirty RF: a new paradigm," in Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on, vol. 4, 2005, pp. 2347–2355 Vol. 4.
- [4] M. Valkama, M. Renfors, and V. Koivunen, "Advanced methods for I/Q imbalance compensation in communication receivers," IEEE Transactions on Signal Processing, vol. 49, no. 10, 2001, pp. 2335–2344.
- [5] B. Razavi, "Design considerations for direct-conversion receivers," Circuits and Systems II: Analog and Digital Signal Processing, IEEE Transactions on, vol. 44, no. 6, 1997, pp. 428–435.
- [6] L. Antilla, "Digital Front-End Processing with Widely-Linear Signal Models in Radio Devices," Ph.D. dissertation, Tampere University of Technology, 2011.
- [7] J. Cavers and M. Liao, "Adaptive compensation for imbalance and offset losses in direct conversion transceivers," Vehicular Technology, IEEE Transactions on, vol. 42, no. 4, 1993, pp. 581–588.
- [8] J. Cavers, "New methods for adaptation of quadrature modulators and demodulators in amplifier linearization circuits," Vehicular Technology, IEEE Transactions on, vol. 46, no. 3, 1997, pp. 707–716.
- [9] A. Nassery, S. Byregowda, S. Ozev, M. Verhelst, and M. Slamani, "Built-in-self test of transmitter i/q mismatch using self-mixing envelope detector," in VLSI Test Symposium (VTS), 2012 IEEE 30th, 2012, pp. 56–61.
- [10] S. D'Souza et al., "A 10-bit 2-gs/s dac-ddfs-iq-controller baseband enabling a self-healing 60-ghz radio-on-chip," Circuits and Systems II: Express Briefs, IEEE Transactions on, vol. 60, no. 8, 2013, pp. 457–461.
- [11] X. Huang and M. Caron, "Efficient transmitter self-calibration and amplifier linearization techniques," in Circuits and Systems, 2007. ISCAS 2007. IEEE International Symposium on, 2007, pp. 265–268.
- [12] H.-J. Jentschel, "Direct conversion receivers - expectations and experiences," in RF Front End Architectures, IEEE MTT-S 2000, Boston, Workshop, June 2000.
- [13] S. O. Haykin, Adaptive Filter Theory. Prentice Hall, 2001.
- [14] H. Meyr, M. Moeneclaey, and S. A. Fechtel, Digital Communication Receivers. Wiley, 1996.

# A Distributed Algorithm for In-Network Adaptive Estimation Using Incremental Aggregated Gradient

Wael Bazzi

Amir Rastegarnia, Azam Khalili

Saeid Sanei

Electrical Engineering Department  
American University in Dubai  
Dubai, UAE  
wbazzi@aud.edu

Department of Electrical Engineering  
Malyer University  
Malyer, 65719-95863, Iran  
{a\_rastegar, a.khalili}@ieee.org

Department of Computing  
University of Surrey  
Surrey, GU2 7XH, U.K.  
s.sanei@surrey.ac.uk

**Abstract**—In this paper, we consider the distributed estimation problem where a set of nodes is deployed to estimate a parameter of interest when the statistical model (or information) for the underlying processes is not available, or it varies in time. Such a scenario appears in many real-world applications e.g. in sensor networks. The estimation problem can be expressed mathematically as the minimization of a cost function, which is the sum of continuously differentiable local cost functions. The paper aims to develop an iterative, fully distributed and adaptive solution for the optimization problem. Similar to the existing Incremental Least Mean-Squares (ILMS) algorithm, in the proposed algorithm we use steepest-descent method to generate an approximation of the descent direction at every node. However, unlike the ILMS, the proposed algorithm uses the aggregate gradient at each node which is the average of the previously computed gradients by other nodes. The resultant algorithm which is called Incremental Aggregated Gradient-LMS (IAG-LMS) outperforms the ILMS algorithm in terms of the steady-state error. Moreover, its stability bound (in terms of the step-size parameter) is also wider than the ILMS algorithm. We present numerical simulations to support the mentioned claims and illustrate the results.

**Keywords**— Adaptive networks; incremental; least mean square; estimation.

## I. INTRODUCTION

Consider the following unconstrained optimization problem

$$\underset{\mathbf{w}}{\text{minimize}} \quad f^g(\mathbf{w}) = \sum_{k=1}^N f_k(\mathbf{w}), \quad \mathbf{w} \in \mathbb{R}^M \quad (1)$$

where every  $f_k : \mathbb{R}^M \rightarrow \mathbb{R}$  is a continuously differentiable scalar function on  $\mathbb{R}^M$  ( $M \in \mathbb{Z} \geq 1$ ). The motivation for considering optimization models of the form in (1) stems from practical problems that arise in a variety of applications such as sensor networks, precision agriculture, environment monitoring, disaster relief management, smart spaces, target localization, neural networks training, and a number of medical applications [1],[2]. As an example, in sensor networks, each  $f_k(\mathbf{w})$  corresponds to the data collected by the  $k$ th sensor in the network. There are already several useful strategies for estimation over distributed networks such as consensus strategies and adaptive networks. In the consensus strategy [3],[4],[5], each node estimates the source signal (or parameter of interest) and shares the information with its neighbors. As the number of iterations increases, the local estimates, computed by the nodes may converge to the same final estimate. The initial implementations suggested for the consensus strategy relied on the use of two time-scales [6]: one

for the collection of measurements and another to iterate over the collected data to attain agreement before the process is repeated. The main problem in the mentioned methods is inability of the network to undertake continuous learning and optimization [7]. Therefore, consensus-based algorithms that rely on two time-scales can not be used in the problems that this paper seeks to provide a solution to them.

Single time-scale consensus strategies, on the other hand, have also been developed in the literature [4],[5],[8],[9]. Although this kind of consensus algorithms can deliver continuous learning, as it is shown in [10], they suffer from stability problems. Indeed, consensus networks can become unstable even if all of the nodes in the network are stable. This problem, along with the need for continuous learning in distributed networks, motivated the development of adaptive networks [11],[12]. An adaptive network is a collection of agents (nodes) that collaborate with each other through in-network local processing rules in order to estimate and track the parameters of interest [12]. The two major classes of adaptive networks are incremental strategy [13],[14],[15],[16] and diffusion strategy [17],[18],[19]. In the incremental mode, a cyclic path through the network is required, and nodes communicate with neighbors within this path, while in diffusion mode, each node can communicate with all of its neighboring nodes. Although incremental networks are less robust to node and link failures, they do offer excellent estimation performance since every node uses data from the entire network to update the local estimate of the desired parameter.

In the available incremental-based adaptive networks such as [13],[15],[16],[20],[21],[22],[23],[24] the estimation problem is solved by starting from steepest-descent solution, splitting the global cost function into  $N$  local cost functions ( $N$  is the number of nodes), and applying a suitable local stochastic approximations (such as LMS-type learning rule) at every node. The resultant algorithms are able to handle the unavailability or variation of statistical information. However, they use only one of the  $N$  components (local cost functions) in order to generate an approximate descent direction [25]. Thus, in this paper we employ the aggregate gradient at each node which is the average of the  $N$  previously computed gradients. The resultant algorithm outperforms the existing incremental LMS (ILMS) algorithm in terms of the steady-state error. Moreover, the stability bound (in terms of the step-size parameter) is also wider than of the ILMS algorithm. We also present some simulation results to support the mentioned claims.

**Notation:** We adopt boldface letters for random quantities. The symbol  $*$  denotes conjugation for scalars and Hermitian transpose for matrices.  $y = \text{floor}(x)$  rounds the elements of  $x$  to the nearest integers less than or equal to  $x$ . Moreover,  $(a)_N$  denotes  $a$  modulo  $N$  with representative classes  $1, 2, \dots, N$ . We use the notation  $\|\mathbf{x}\|_{\mathbf{A}}^2 =$

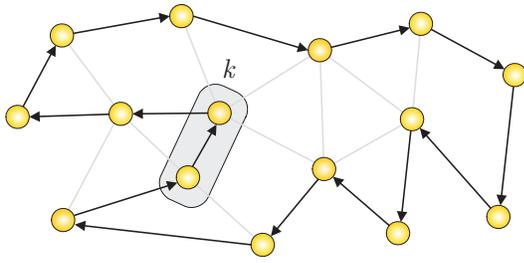


Fig. 1. A distributed network with incremental cooperation among the nodes. The neighbor of node  $k$  is defined as its preceding node in Hamilton cycle.

$\mathbf{x}^* \mathbf{A} \mathbf{x}$  for the weighted square norm of  $\mathbf{x}$ .

## II. PROBLEM STATEMENT

To further explain the problem that this paper seeks to provide a solution to it, let us consider a set of nodes  $\mathcal{N} = \{1, 2, \dots, N\}$  that are distributed over a domain in space. Each node  $k \in \mathcal{N}$  communicates with its neighbors (denoted by  $\mathcal{N}_k$ ). The set of neighbors for node  $k$  for incremental cooperation mode is shown in Fig. 1. At time instant  $i$ , node  $k$  records the scalar measurement  $d_k(i)$  and  $1 \times M$  regression data  $\mathbf{u}_{k,i}$ . The purpose of the network is to solve the following optimization problem at every node in the network

$$\underset{\mathbf{w}}{\text{minimize}} \sum_{k=1}^N \mathbb{E}[|d_k(i) - \mathbf{u}_{k,i} \mathbf{w}|^2] \quad (2)$$

Comparing (1) with (2) reveals that

$$f_k(\mathbf{w}) = \mathbb{E}[|d_k(i) - \mathbf{u}_{k,i} \mathbf{w}|^2]$$

**Remark 1.** In many practical applications, we can assume a linear regression model between the measurements  $\{d_k(i)\}$  and  $\{\mathbf{u}_{k,i}\}$ . According to this model, we have

$$d_k(i) = \mathbf{u}_{k,i} \mathbf{w}^o + v_k(i) \quad (3)$$

where  $\mathbf{w}^o \in \mathbb{R}^M$  is an unknown parameter and  $v_k(i)$  is the observation noise term with variance  $\sigma_{v,k}^2$ . Based on the intended application, the vector  $\mathbf{w}^o$  may represent different physical quantities, e.g. location of a target and parameter of an auto-regressive (AR) model.

The optimal solution for (2) is  $\mathbf{w}^o$  and can be expressed in terms of the statistics of recorded data  $\{d_k(i), \mathbf{u}_{k,i}\}$  via the following normal equation [13],[14]

$$\left( \sum_{k=1}^N \mathbf{R}_{u,k} \right) \mathbf{w}^o = \sum_{k=1}^N \mathbf{p}_{du,k} \quad (4)$$

where

$$\mathbf{R}_{u,k} = \mathbb{E}[\mathbf{u}_{k,i}^* \mathbf{u}_{k,i}], \quad \mathbf{p}_{du,k} = \mathbb{E}[d_k(i) \mathbf{u}_{k,i}^*] \quad (5)$$

**Remark 2.** The unknown vector  $\mathbf{w}^o$  in the linear model (3) is the same as the optimal solution in (4).

This solution needs the statistical information  $\{\mathbf{p}_{du,k}, \mathbf{R}_{u,k}\}_{k \in \mathcal{N}}$  to be available at every node, which are not available in many applications or they may vary in time. In the next section, we introduce our proposed distributed estimation algorithm which relies on the incremental aggregated gradient method.

## III. PROPOSED ALGORITHM

To solve the unconstrained problem in (1), we start with the standard steepest-descent method. Let  $\mathbf{w}_i^g$  be the global estimate for  $\mathbf{w}^o$  at iteration  $i$ . Then, using the standard steepest-descent method we have

$$\mathbf{w}_i^g = \mathbf{w}_{i-1}^g - \mu (\nabla_{\mathbf{w}} f^g(\mathbf{w}_{i-1}^g))^* \quad (6)$$

where  $\mu > 0$  is a step-size parameter and  $\nabla_{\mathbf{w}} f^g$  represents the complex gradient of the  $f^g(\mathbf{w}^g)$ . The required gradient in (6) is given by

$$(\nabla_{\mathbf{w}} f^g(\mathbf{w}_{i-1}^g))^* = \sum_{k=1}^N (\mathbf{R}_{u,k} \mathbf{w}_{i-1}^g - \mathbf{p}_{du,k}) \quad (7)$$

Substituting (7) into (6) yields the steepest-descent solution as

$$\mathbf{w}_i^g = \mathbf{w}_{i-1}^g + \mu \sum_{k=1}^N (\mathbf{p}_{du,k} - \mathbf{R}_{u,k} \mathbf{w}_{i-1}^g) \quad (8)$$

As we mentioned before, this is not a practical solution as it needs the statistical information  $\{\mathbf{p}_{du,k}, \mathbf{R}_{u,k}\}_{k \in \mathcal{N}}$ . This issue can be addressed by replacing the statistical averages by time averages assuming that the process is ergodic, i.e.

$$\mathbf{R}_{u,k} \approx \mathbf{u}_{k,i}^* \mathbf{u}_{k,i}, \quad \mathbf{p}_{du,k} \approx d_k(i) \mathbf{u}_{k,i}^* \quad (9)$$

Replacing (9) in (8) yields

$$\mathbf{w}_i^g = \mathbf{w}_{i-1}^g + \mu \sum_{k=1}^N \mathbf{u}_{k,i}^* (d_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{i-1}^g) \quad (10)$$

The above recursive equation can be implemented in a distributed manner by splitting the update equation into  $N$  steps and resorting to the incremental cooperation which is established among the nodes. By doing the above steps, the ILMS algorithm is obtained which can be written as follows

$$\begin{cases} \mathbf{w}_{1,i} \leftarrow \mathbf{w}_{N,i-1} \\ \mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} + \mu_k \mathbf{u}_{k,i}^* (d_k(i) - \mathbf{u}_{k,i} \mathbf{w}_{k-1,i}) \end{cases} \quad (11)$$

where  $\mathbf{w}_{k,i}$  is a local estimate of  $\mathbf{w}^o$  at node  $k$  and time (iteration)  $i$ . Note that we have

$$\mathbf{w}_{N,i} = \mathbf{w}_i^g$$

It is shown in [11], [13] that as  $i \rightarrow \infty$ , we have  $\mathbf{w}_{k,i} \rightarrow \mathbf{w}^o$  in the mean, for every node  $k$ , and for an appropriately chosen set of step sizes  $\{\mu_k : k = 1, \dots, N\}$ .

It is clear from (6) that the ILMS algorithm uses only the available information at node  $k$  in order to generate an approximate for the steepest-descent direction. We can improve the approximate direction if we use the average of the  $N$  previously computed gradients [25]. Thus, we use the concept of aggregated gradient to develop our proposed algorithm. To begin with, let us define the sequence of  $\{\mathbf{x}_j\}$ ,  $j = 1, 2, \dots$  as

$$\mathbf{x}_j \triangleq \mathbf{w}_{k,i}, \quad j = (i-1) \times N + k \quad (12)$$

Equivalently, we have

$$k = (j)_N, \quad \text{and} \quad i = \text{floor}(j/N) + 1 \quad (13)$$

Using (12), we can define the aggregated gradient at iteration  $i$  as

$$\mathbf{g}_{k,i} = \left( \sum_{\ell=0}^{N-1} \nabla_{\mathbf{w}} f_{(k-\ell)_N}(\mathbf{x}_{k-\ell}) \right)^* \quad (14)$$

Note the right hand side of (14) is a function of  $i$  since according to (12)  $\mathbf{x}_{k-\ell}$  can be expressed in terms of  $\mathbf{w}_{k,i}$ .

**Remark 3.** To calculate the aggregated gradient in (14) we need  $N$  initial points  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$  which according to (12) can be selected as

$$\mathbf{x}_1 = \mathbf{w}_{1,1}, \quad \mathbf{x}_2 = \mathbf{w}_{2,1}, \quad \dots \quad \mathbf{x}_N = \mathbf{w}_{N,1}$$

That is why we have defined  $\mathbf{g}_{k,i}$  for  $i \geq 2$ . Possible initialization strategies include setting  $\mathbf{x}_1 = \mathbf{x}_2 = \dots = \mathbf{x}_N$ .

Now, we can modify the standard steepest-descent (14) using (6) as follows

$$\mathbf{x}_j = \mathbf{x}_{j-1} - \mu \frac{1}{N} \mathbf{g}_{k-1,i}, \quad \text{for } j \geq N+1 \quad (15)$$

or in terms of  $\mathbf{w}_{k,i}$  as

$$\mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} - \mu \frac{1}{N} \mathbf{g}_{k-1,i} \quad (16)$$

where we add the factor  $1/N$  in order to make  $\mathbf{g}_{k,i}$  comparable to the one used in the standard incremental gradient method [25]. We can also rewrite (14) in an equivalent form as

$$\mathbf{g}_{k,i} = \mathbf{g}_{k-1,i} - (\nabla_{\mathbf{w}} f_{(k)_N}(\mathbf{x}_{k-L}))^* + (\nabla_{\mathbf{w}} f_{(k)_N}(\mathbf{x}_k))^* \quad (17)$$

As we will discuss later, this form enables us to implement equations (15) and (17) in a distributed manner.

**Remark 4.** Note that for  $i \geq 2$  we can merge the equations (16) and (17) to obtain an equivalent update equation as

$$\mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} - \mu \frac{1}{N} \left( \sum_{\ell=0}^{N-1} \nabla_{\mathbf{w}} f_{(k-\ell)_N}(\mathbf{w}_{k',i'}) \right)^* \quad (18)$$

where

$$k' = (k-1-\ell)_N, \quad \text{and} \quad i' = \text{floor}((k-1-\ell)/N) + 1.$$

Although the update equation given by (16) is a distributed solution for (1), however, it is not an *adaptive* solution because we still need to use the second-order moments  $\{p_{d_{u,k}}, \mathbf{R}_{u,k}\}$  to evaluate (16). To obtain an adaptive solution, we replace the mentioned moments with the local instantaneous approximations as

$$(\tilde{\nabla}_{\mathbf{w}} f_k(\mathbf{w}))^* = -(\mathbf{u}_{k,i}^* (d_k(i) - \mathbf{u}_{k,i} \mathbf{w})) \quad (19)$$

where we used  $\tilde{\nabla}_{\mathbf{w}} f_k(\mathbf{w})$  to denote the *approximate* gradient. Using (19), we can rewrite

$$\mathbf{w}_{k,i} = \mathbf{w}_{k-1,i} - \mu \frac{1}{N} \tilde{\mathbf{g}}_{k-1,i} \quad (20a)$$

$$\tilde{\mathbf{g}}_{k,i} = \tilde{\mathbf{g}}_{k-1,i} - (\tilde{\nabla} f_{(k)_N}(\mathbf{x}_{k-L}))^* + (\tilde{\nabla} f_{(k)_N}(\mathbf{x}_k))^* \quad (20b)$$

We can implement (20) in a fully distributed manner as follows: at iteration  $i$ , upon receiving  $\mathbf{x}_{(i-1) \times N + k-1} = \mathbf{w}_{k-1,i}$  and  $\tilde{\mathbf{g}}_{k-1,i}$  from node  $k-1$ , node  $k$  updates the local estimate  $\mathbf{w}_{k,i}$  and the aggregate estimate  $\tilde{\mathbf{g}}_{k,i}$  according to (20) and sends them to the next node  $k+1$ . The pseudo code of the proposed algorithm (IAG-LMS algorithm) is shown in Algorithm. 1.

**Algorithm 1:** The pseudo code of the IAG-LMS algorithm.

---

```

Initialize  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N = \mathbf{w}_{1,1}, \mathbf{w}_{2,1}, \dots, \mathbf{w}_{N,1} = \mathbf{0}$ 
foreach  $i = 1, 2, \dots$  do
    foreach  $k = 1, 2, \dots, N \dots$  do
        receive  $\mathbf{w}_{k-1,i}$  and  $\tilde{\mathbf{g}}_{k-1,i}$  from node  $k-1$ 
        update  $\mathbf{w}_{k,i}$  according to (20a)
        update  $\tilde{\mathbf{g}}_{k,i}$  according to (20b)
        send  $\mathbf{w}_{k,i}$  and  $\tilde{\mathbf{g}}_{k,i}$  to the next node (node  $k+1$ ).
    end
end
    
```

---

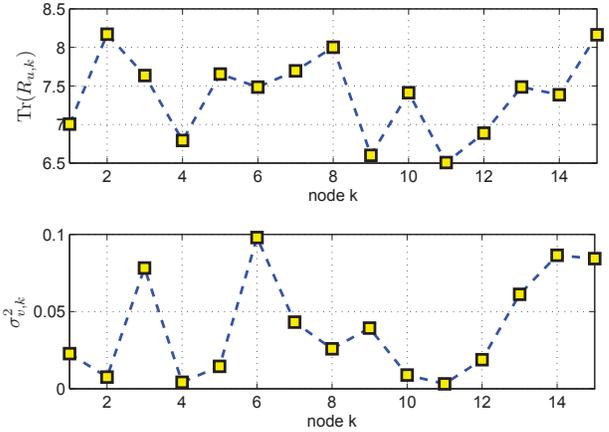


Fig. 2. Node profile  $\text{Tr}(R_{u,k})$  (up) and  $\sigma_{v,k}^2$  (down).

#### IV. SIMULATION RESULTS

In this section, we present some simulation results to evaluate the performance of the proposed algorithm. To this end, we assume a network with  $N = 15$  nodes, where the nodes are connected via a ring topology. The regressors  $\mathbf{u}_{k,i}$  are generated as independent realizations of a Gaussian distribution with a covariance matrix  $\mathbf{R}_{u,k}$  where the eigenvalue spread for every  $\mathbf{R}_{u,k}$  is 2. The measurement data  $d_k(i)$  at each node  $k$  is generated by using the data model (3) where the parameter  $\mathbf{w}^\circ$  is chosen to be  $[1 \ 1 \ 1 \ 1]^T$ . The observation noise  $v_k(i)$  in (3) is drawn from a Gaussian distribution with variance  $\sigma_{v,k}^2 \in (0, 0.1)$ . Fig. 2 shows the node profile including the trace of every covariance matrix  $\mathbf{R}_{u,k}$  and  $\sigma_{v,k}^2$ .

We consider the mean-square deviation as the performance metric which is defined at node  $k$  as

$$\text{MSD}_k = \mathbb{E}[\|\mathbf{w}^\circ - \mathbf{w}_{k,i}\|^2] \quad (21)$$

To evaluate the performance of ILMS and the proposed algorithm, we use *network* MSD that is given by

$$\text{MSD} = \frac{1}{N} \sum_{k=1}^N \text{MSD}_k \quad (22)$$

Fig. 3 shows the learning curve, in terms of network MSD, for the ILMS algorithm and the proposed algorithm. For both ILMS and the proposed algorithm we select  $\mu = 0.1$ . The results are obtained by averaging over 100 independent runs. We observe that the proposed algorithm outperforms the ILMS algorithm in terms of the steady-state error at the expense of a slight increase in the computations (due to the need for updating  $\tilde{\mathbf{g}}_{k,i}$ ) and communications (due to the need for sending  $\tilde{\mathbf{g}}_{k,i}$ ). In Fig. 4, we have plotted the trajectory of the proposed algorithm. The dash straight line corresponds to the  $\mathbf{w}^\circ(1) = 1$ . It is seen that the proposed algorithm rapidly converges to the desired parameter.

When we increase the step-size value, we observe that the ILMS algorithm diverges and can not provide an acceptable estimate for  $\mathbf{w}^\circ$ . This behavior does not occur for the proposed algorithm as it is shown in Fig. 5, where we have plotted the network MSD for both algorithms for a bigger step-size  $\mu = 0.25$ . Fig. 6 shows steady-state values of the network MSD for both algorithms in terms of the step-size value. The steady-state values are obtained by averaging over 100 runs 50 time samples after the convergence of algorithms. We observe that the stability range for the proposed algorithm is wider than the ILMS algorithm, which leads to a more robust implementation.

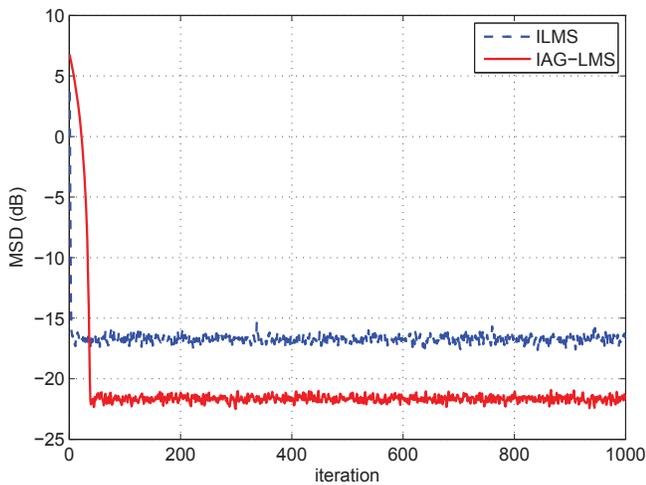


Fig. 3. Learning curve for ILMS and the proposed algorithm in terms of network MSD ( $\mu = 0.1$ ).

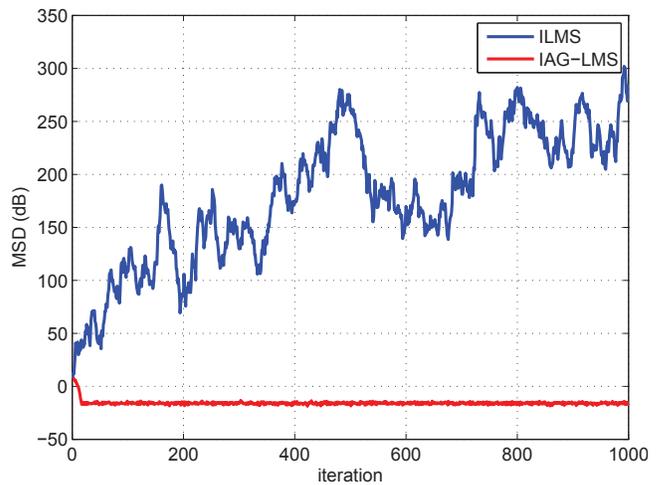


Fig. 5. Learning curve for ILMS and the proposed algorithm in terms of network MSD ( $\mu = 0.25$ ).

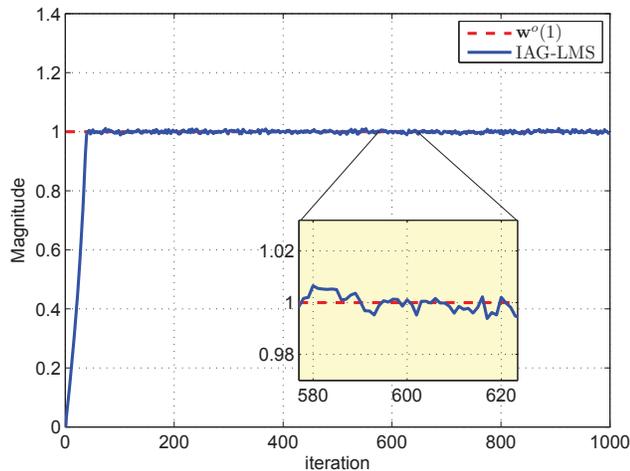


Fig. 4. The trajectory of the proposed algorithm. The dash straight line corresponds to  $w^o(1) = 1$ , ( $\mu = 0.1$ ).

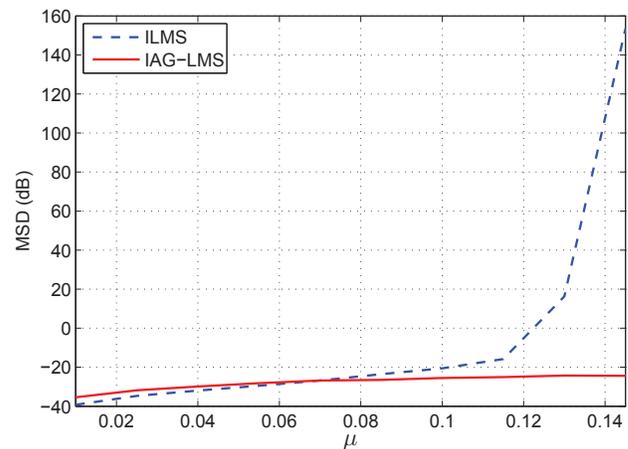


Fig. 6. Steady-state performances of the ILMS algorithm and the proposed algorithm in terms of the step-size parameter.

### V. CONCLUSIONS AND FUTURE WORK

In this paper, we proposed an incremental adaptive network based on the aggregated gradient method. The main advantages of the proposed algorithm, (in comparison with the existing ILMS algorithm) are its steady-state performance error and its wider stability range. However, the cost of this improvement is a slight increase in computation and communication per integration per node. In addition, the convergence rate for the proposed algorithm is slower than ILMS algorithm, which in turn affects its tracking performance spatially in fast-changing environment. We presented some simulation results to support our claims in this paper. This work can be extended by considering the effects of noisy links, which is a more practical assumption than the ideal links. Moreover, we can further improve the performance of the proposed algorithm with tuning the step-size parameter at every node according to its measurement quality. Our future work will address the mentioned issues.

### REFERENCES

[1] N. Bulusu and S. Jha, *Wireless Sensor Networks: A Systems Perspective*. Artech House, 2005.

[2] I. F. Akyildiz, T. Melodia, and K. R. Chowdury, "Wireless multimedia sensor networks: A survey," *IEEE Wireless Commun. Mag.*, vol. 14, no. 6, pp. 32–39, 2007.

[3] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE/ACM Trans. Netw.*, vol. 14, no. SI, pp. 2508–2530, June 2006.

[4] I. Schizas, G. Mateos, and G. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *Signal Processing, IEEE Transactions on*, vol. 57, no. 6, pp. 2365–2382, 2009.

[5] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: Fundamental limits and tradeoffs," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 4, pp. 674–690, 2011.

[6] L. Xiao, S. Boyd, and S. Lai, "A space-time diffusion scheme for peer-to-peer least-squares estimation," in *Information Processing in Sensor Networks, 2006. IPSN 2006. The Fifth International Conference on*, 2006, pp. 168–176.

[7] J. Chen and A. H. Sayed, "Diffusion adaptation strategies for distributed optimization and learning over networks," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4289–4305, August 2012.

[8] A. Nedic and A. Ozdaglar, *Cooperative distributed multi-agent optimization*, Y. Eldar and D. Palomar, Eds. Cambridge, U. K.: Cambridge Univ. Press, 2009.

- [9] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *Automatic Control, IEEE Transactions on*, vol. 31, no. 9, pp. 803–812, 1986.
- [10] S.-Y. Tu and A. Sayed, "Diffusion strategies outperform consensus strategies for distributed estimation over adaptive networks," *Signal Processing, IEEE Transactions on*, vol. 60, no. 12, pp. 6217–6234, 2012.
- [11] A. H. Sayed, "Diffusion adaptation over networks," in *E-Reference Signal Processing*, R. Chellapa and S. Theodoridis, Eds. Elsevier, 2013, to appear.
- [12] A. H. Sayed, S.-Y. Tu, J. Chen, X. Zhao, and Z. Towfic, "Diffusion strategies for adaptation and learning over networks," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 155–171, May 2013.
- [13] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Trans. Signal Processing*, vol. 55, no. 8, pp. 4064–4077, August 2007.
- [14] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategies over adaptive networks," in *Proc. Asilomar Conf. Signals, Systems, Computers*, Monterey, CA, October 2006, pp. 233–237.
- [15] N. Takahashi and I. Yamada, "Incremental adaptive filtering over distributed networks using parallel projection onto hyperslabs," *IEICE Technical Report*, vol. 108, pp. 17–22, 2008.
- [16] F. Cattivelli and A. H. Sayed, "Analysis of spatial and incremental LMS processing for distributed estimation," *IEEE Trans. on Signal Process.*, vol. 59, no. 4, pp. 1465–1480, April 2011.
- [17] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Trans. on Signal Process.*, vol. 56, no. 7, pp. 3122–3136, July 2008.
- [18] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion recursive least-squares for distributed estimation over adaptive networks," *IEEE Trans. on Signal Process.*, vol. 56, no. 5, pp. 1865–1877, May 2008.
- [19] F. S. Cattivelli and A. H. Sayed, "Multilevel diffusion adaptive networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Taipei, Taiwan, April 2009.
- [20] C. G. Lopes and A. H. Sayed, "Distributed processing over adaptive networks," in *Proc. Adaptive Sensor Array Processing Workshop*. Lexington, MAMIT Lincoln Lab., Lexington, MA: MIT Lincoln Lab, June 2006.
- [21] L. Li, J. A. Chambers, C. G. Lopes, and A. H. Sayed, "Distributed estimation over an adaptive incremental network based on the affine projection algorithm," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 151–164, January 2010.
- [22] C. G. Lopes and A. H. Sayed, "Randomized incremental protocols over adaptive networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Dallas, TX, March 2010, pp. 3514–3517.
- [23] A. Khalili, M. A. Tinati, and A. Rastegarnia, "Amplify-and-forward scheme in incremental lms adaptive network with noisy links: Minimum transmission power design," *AEU - International Journal of Electronics and Communications*, vol. 66, no. 3, pp. 262 – 265, 2012.
- [24] A. Rastegarnia and A. Khalili, "Incorporating observation quality information into the incremental lms adaptive networks," *Arabian Journal for Science and Engineering*, vol. 39, no. 2, pp. 987–995, 2014.
- [25] D. Blatt, A. Hero, and H. Gauchman, "A convergent incremental gradient method with a constant step size," *SIAM Journal on Optimization*, vol. 18, no. 1, pp. 29–51, 2007.

# Channel Inversion CoMP Technique in Cellular System: A User-Selection Algorithm

J. Gómez, M.C. Aguayo-Torres, F. Blázquez-Casado, F.J. Martín-Vega  
 Departamento de Ingeniería de Comunicaciones  
 Universidad de Málaga  
 Málaga, Spain  
[jesusgomezgalvan@gmail.com](mailto:jesusgomezgalvan@gmail.com), {aguayo, fbc, [fjmvega](mailto:fjmvega@ic.uma.es)}@ic.uma.es

**Abstract**—Network coordination techniques are very promising for improving spectral efficiency over conventional cellular networks limited by inter-cell interference. Thus, Coordinated MultiPoint (CoMP) transmission and reception techniques are under consideration for LTE-Advanced to meet 4G requirements. In this work, a coordination technique for inter-cell interference cancellation is evaluated over a packetized multiuser adaptive OFDMA (Orthogonal Frequency-Division Multiple Access) system. As a result of our network coordination setup, a new question on which pair of users should be served at each physical resource comes up. A low power user selection algorithm is proposed and evaluated, and its performance is compared to that of random user selection.

**Keywords**—; CoMP; LTE; MIMO; 4G; QoS

## I. INTRODUCTION

Data access over mobile networks has caused an increasing demand for high data rate wireless communications, a trend which is expected to continue. Nowadays, a typical user consumes 1 Gigabyte of data per month, approximately 200% percent more than one year ago [1] [2] [3]. The current mobile networks and wireless technologies need to be improved in order to support the traffic demand.

To satisfy the high demand for wireless services, 3<sup>rd</sup> Generation Partnership Project (3GPP) has been developing a new mobile standard, referred to as Long Term Evolution (LTE). The first release of LTE (Release 8) is labeled as 3.9G as it does not meet the International Mobile Telecommunications-Advanced (IMT-Advanced) requirements for 4G. However, the evolved version of LTE (Release 10), referred to as LTE-Advanced (LTE-A), meets or exceeds the requirements of the International Telecommunication Union (ITU) for the fourth generation (4G) radio communication standard, known as IMT-Advanced. In October 2009, the 3GPP Partners formally submitted LTE-A to the ITU Radiocommunication sector (ITU-R) as a candidate for 4G IMT-Advanced [4].

In a cellular communication system such as 3GPP LTE,

throughput is limited by co-channel interference, either within a cell (intra-cell interference) or from nearby cells (inter-cell interference).

Downlink intra-cell interference can be eliminated by orthogonal channel allocation. LTE radio transmission is based on Orthogonal Frequency-Division Multiple Access (OFDMA). Over frequency selective channels, OFDMA divides the transmitted bit flow into many different substreams and sends these over many different subchannels. Data from multiple users are being transmitted on parallel narrow-band subcarriers. The advantage is that each subcarrier is relatively narrowband, which decreases the effect of delay spread.

Adaptive Quadrature Amplitude Modulation (AQAM) is an adaptive digital modulation that provides the system with the ability to match the service rate to the channel quality while meeting a target Bit Error Rate (BER) [2]. Adaptive QAM and OFDMA are combined by performing adaptive modulation in each OFDM subcarrier. Although OFDMA achieves improved spectral efficiency within one cell, inter-cell interference is still preventing these technologies from coming close to the theoretical rates for multi-cell networks [5].

Inter-cell interference mainly influences the data rates of those users at the cell-edge affecting the average spectral efficiency of the cell. Using different transmission frequencies between neighboring cells, soft handoff or beam-forming multiple antennas, inter-cell interference can be reduced. Using more advanced techniques by coherently coordinating transmissions of signals across base stations, interference could be completely eliminated. Roughly speaking, multiple base stations, geographically distributed, transmit user's signal simultaneously. These signals are weighted and pre-processed so that inter-cell interference is cancelled [6] [7].

In the 3GPP standardization activities concerning LTE-Advanced [8], the coordinated base stations transmissions fall into the term of Coordinated Multi-Point (CoMP) transmission and reception. CoMP refers to a wide range of techniques with the common characteristic of dynamic

coordination of transmission and/or reception at multiple geographically separated sites with the aim to enhance system performance and end-user quality [1]. 3GPP has proposed CoMP techniques as a performance enhancement for LTE-A Release 12 and beyond, to meet 4G requirements [8] [9]. These techniques are the key to achieve the needed requirements in terms of spectral efficiency and cell-edge throughput.

In this paper, we present network coordination as a means to provide high spectral efficiency in cellular systems. This study presents a basic scenario where two base stations, connected via high speed backbone, coordinate their transmissions and pre-process the signals for cancelling inter-cell interference. Each user receives its data from both base stations improving the signal strength and with no interference, thus reaching a higher spectral efficiency. The performance of coordinated networks is compared to that of conventional networks without coordination and, as expected, it is shown that coordinated base station transmissions offers better spectral efficiency values.

Our coordinated experimental setup is not power efficient, so a new challenge comes up for selecting the users to be served at each physical resource in order to minimize the system transmission power. A user selection algorithm, based on simulation results, is proposed. This algorithm specifies resources allocation to users in order to minimize system transmission power. The proposed algorithm is evaluated and their results are compared to that of a random user selection algorithm.

The remainder of this paper is organized as follows. Section II briefly describes the fundamentals of coordinated networks, presenting the studied CoMP technique for cancelling inter-cell interference. In Section III, the system model is described. In Section IV, a user selection algorithm is proposed for system transmission power reduction. In Section V, the coordinated scenario using inter-cell interference cancellation and the proposed user selection algorithm is then evaluated. Finally, some concluding remarks are given in Section VI.

## II. COMP TRANSMISSION

The basic system model is as shown in Fig. 1. Base Stations (BSs) BS1 and BS2 will operate together to transmit signals to Mobile Stations (MSs) MS1 and MS2. Both base stations will be geographically separated but connected via high speed backbone link making coordination possible. Both BSs and MSs are equipped with single antennas. Knowledge of the Channel State Information (CSI) is assumed to be perfectly available at the transmitters, as it is measured by the receivers and fed back to the BSs.

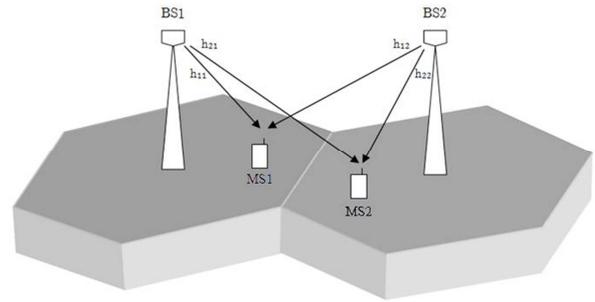


Fig. 1. System model

In a classical cell system, for BS1, BS2 is an interfering base station, and in the same way, for BS2, BS1 is an interfering base station. The received signals at MS1 and MS2,  $y_1$  and  $y_2$ , respectively, are given by

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \quad (1)$$

where  $h_{ij}$  is the channel between terminal  $i$  and BS  $j$ ,  $x_1$  is transmitted signal by BS1 and  $x_2$  is that by BS2.  $n_i$  is additive white Gaussian noise [10].

A basic system model for a downlink network with  $M$  single antenna base stations and  $N$  single antenna mobiles is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (2)$$

where  $\mathbf{H} = [h_{ij}]_{N \times M}$  denotes the channel matrix,  $h_{ij}$  denotes the complex channel gain between mobile  $i$  and base station  $j$ ,  $\mathbf{x} = [x_1, x_2, \dots, x_M]^T$  denotes the complex antenna inputs to the channel,  $\mathbf{n} = [n_1, n_2, \dots, n_N]^T$  denotes additive white noise and  $\mathbf{y} = [y_1, y_2, \dots, y_M]^T$  are the inputs to the receiver.

In a coordination scenario, where all  $M$  base stations act together, each mobile may receive useful signals from all base stations. If a linear spatial precoding matrix  $\mathbf{A} \in \mathbb{C}^{M \times N}$  is used to map the data symbols to the antenna outputs, i.e.,

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (3)$$

where  $\mathbf{s}$  denotes the vector of data symbols given by  $\mathbf{s} = [s_1, s_2, \dots, s_N]^T$  then the antenna output at the  $j$ th base station is a linear combination of  $N$  data symbols,

$$x_j = \sum_{i=1}^N A_{ji}s_i \quad (4)$$

Selecting a proper precoding matrix  $\mathbf{A}$  allows interference cancellation. A simple form of coordination, called Channel Inversion (CI) or Zero-Forcing (ZF) [1] [2] [6] [9], uses a pseudo-inverse precoding matrix given by

$$\mathbf{A} = \mathbf{H}^\dagger (\mathbf{H}\mathbf{H}^\dagger)^{-1} \quad (5)$$

According to this, when  $N = M$ , (3) becomes simply  $\mathbf{x} = \mathbf{H}^{-1}\mathbf{d}$  [6] [9].

The received signal according to (2), (3) and (5) is given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = \mathbf{H}\mathbf{H}^\dagger (\mathbf{H}\mathbf{H}^\dagger)^{-1}\mathbf{s} + \mathbf{n} = \mathbf{s} + \mathbf{n} \quad (6)$$

Thus the  $i$ th mobile receives  $y_i = s_i + n_i$ . All network antennas in range can help in the transmission of each message, but the message is received only by the intended user with no interference [6].

Notice that according to (6) the system model corresponds to an Additive White Gaussian Noise (AWGN) channel where inter-cell interference has been cancelled. To achieve this is not free: depending on the actual channel the precoder should compensate the channel's effects causing fluctuations on the system transmission power. These fluctuations will require a large normalization factor in a real system, which will dramatically reduce the Signal-to-Noise Ratio (SNR) at receivers.

### III. SYSTEM MODEL

Orthogonal Frequency Division Multiplexing (OFDM) is a modulation technique widely used to counteract the effects of Inter Symbol Interference (ISI) in frequency selective channels [2]. OFDM divides the transmission band in a large number of sub-bands narrow enough to be considered flat. An Inverse Fast Fourier Transform (IFFT) efficiently performs the modulation process. Its reciprocal process, the forward Fast Fourier Transform (FFT), is used to recover the data as a cyclic extension of the OFDM symbol eliminating the residual ISI. In this way, OFDM can be considered as a time-frequency squared pattern, where each bin can be addressed independently.

Modulation of the OFDM subcarriers is analogous to that of the conventional single carrier systems. Supported downlink data-modulation schemes are Binary Phase-Shift Keying (BPSK), Quadrature Phase-Shift Keying (QPSK), 16QAM, and 64QAM (as those in LTE). The number of bits allocated to each subcarrier can be modified on a symbol basis to simultaneously track the time variant frequency response of the channel and fulfill the BER service requirements.

When OFDM is also used as multiplexing technique, the term OFDM Access (OFDMA) is preferred. In this case, a block of bins is assigned to a single user in what can be considered a hybrid Time Division Multiple Access-Frequency Division Multiple Access (TDMA-FDMA) technique.

In this work, the physical resources to be assigned to users are formed by one subcarrier in one symbol time, thus OFDMA systems allow subsets of the subcarriers to be allocated dynamically among the different users on the

channel. A Round Robin (RR) scheduling algorithm has been used in order to allocate resources to users. Round Robin is a fixed no opportunist scheduling algorithm which dispenses system resources equally among users following a cyclic order. Users are located in an imaginary line between BS1 and BS2 and the distance to their respective BSs is defined for each user and simulation.

Perfect channel state information (CSI) is assumed at the transmitter. As a consequence, results obtained here are upper bounds for those in a practical system.

The setup is according to system model of Fig. 1. Two hexagonal cells with a base station located at the center and where each base is loaded with one mobile. The cell radius considered is set to  $R_{cell}$  meters.

As channel model, frequency selective Rayleigh fading channel parameters have been employed. We assume path loss Hata model [2] in urban areas including lognormal shadowing with a standard deviation of  $\sigma$  dB. Propagation losses and shadowing have been included to channel matrix  $\mathbf{H} = [h_{ij}]_{N \times M}$ . Being  $h_{ij}$  independent normal random variables with zero mean and unit variance modeling the normalized complex channel gain between mobile  $i$  and base station  $j$ , and  $g_{ij}$  being the propagation losses including shadowing between mobile  $i$  and base station  $j$ , channel matrix  $\mathbf{H}$  is given by

$$\mathbf{H} = \begin{bmatrix} g_{11}h_{11} & g_{12}h_{12} \\ g_{21}h_{21} & g_{22}h_{22} \end{bmatrix} \quad (8)$$

In our two base stations OFDMA system model, each BS will serve to the mobile in its own cell coverage, thus a pair of users is served in each resource unit, formed by one subcarrier in one symbol time. Furthermore, in our coordinated scenario according to (4) the precoder makes system transmission power variable. This brings us up a new challenge: a user selection algorithm could be defined, which allows to choose the best pair of users available for resource allocation, in terms of system transmission power reduction.

When  $N = M$  and according to (3) and (5) complex antenna inputs to the channel from BS1 and BS2 are given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} g_{11}h_{11} & g_{12}h_{12} \\ g_{21}h_{21} & g_{22}h_{22} \end{bmatrix}^{-1} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\det(\mathbf{H})} \begin{bmatrix} g_{22}h_{22} & -g_{12}h_{12} \\ -g_{21}h_{21} & g_{11}h_{11} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (10)$$

where  $x_i$  denotes the complex antenna inputs to the channel and  $s_i$  is the  $i$ th mobile's complex data symbol.

According to (10), system transmission power for BS1 and BS2 depends on  $\det(\mathbf{H})$  given by

$$\det(\mathbf{H}) = g_{11}g_{22}h_{11}h_{22} - g_{12}g_{21}h_{12}h_{21} \quad (11)$$

Notice according to (10) that, when an inverse precoding matrix is applied to the input symbols, the transmission power for BS1 and BS2 is modified. Thus, precoding causes system transmission power fluctuations. An ill-conditioned channel matrix will cause enormous final transmission power. This is the price that must be paid in order to achieve inter-cell interference be identically to zero maintaining SNR level at the receivers.

#### IV. USER SELECTION ALGORITHM

In this section, a user selection algorithm for system transmission power reduction is proposed.

As we mention before, according to (10) and (11) the precoder causes fluctuations on system transmission power in order to compensate the channel's effects.

In a real scenario, a power constraint should be applied in order to maintain a fixed transmission power. Then, a large normalization factor should be used that would dramatically reduce the SNR at the receivers. Our theoretical study is based on how channel inverse precoding technique affects to system transmission power, so no power constraint will be applied. It will conduct us to get an unreasonable amount of transmission power but it will be essential to not cover up transmission power results. In this way, the results will be able to be analyzed in order to define a more efficient transmission power user selection algorithm. Once the algorithm has been defined it could be applied in a real scenario where a power constraint has been used.

According to (11) it can be deduced that minimum system transmission power corresponds to a situation as showed in Fig. 2. In this scenario, is very probable that  $\det(\mathbf{H})$ , has a maximum value and, as a consequence, system transmission power according to (10) will be minimum.

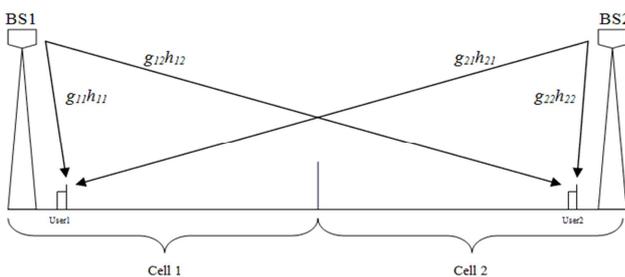


Fig. 2. Minimum sytem transmission power scenario

In the same way, maximum system transmission power corresponds to a scenario as showed in Fig. 3. In this situation,  $\det(\mathbf{H})$  has probably a minimum value and therefore according to (10) system transmission power will be maximum.

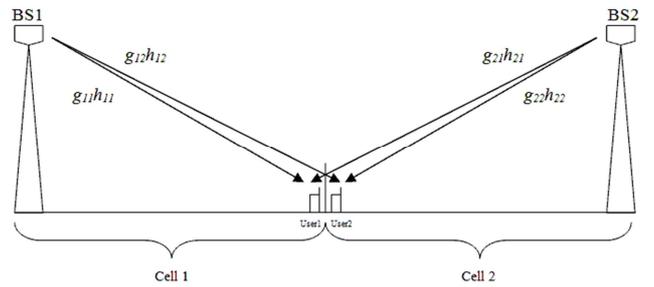


Fig. 3. Maximum system transmission power scenario

Given a fixed position for user1 and variable position for user2 as shown in Fig. 4

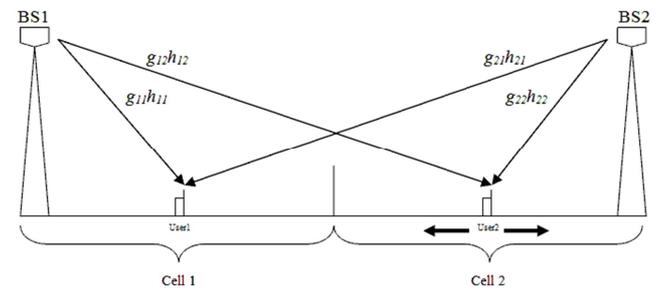


Fig. 4. System transmission power dependence scenario

It can be deduced according to (10) and (11) that if user 2 moves close to cell edge  $\det(\mathbf{H})$  value is probably reduced and therefore transmission power is increased. Otherwise, if user 2 moves close to its BS,  $\det(\mathbf{H})$  value is probably increased and therefore transmission power is reduced.

We can conclude that we need high values for  $\det(\mathbf{H})$  in order to have a lower system transmission power. To achieve this, user 1 and user 2 should be the most far away between them as possible. So for a fixed position of one user its pair, to be served in the same resource block, should be selected as far as possible.

Thus, taking into account the above and based on system transmission power results showed in Fig. 7, we propose the following criterion for user's pair selection:

$$d_2 \leq R_{cell} - d_1 \quad (12)$$

being  $d_1$  and  $d_2$  distances MS1 to BS1 and MS2 to BS2 respectively.  $R_{cell}$  denotes cell radius as defined above. Selecting user pairs, to be served in the same resource block, following (12) ensures a lower system transmission power, as we will show in the next section. The lower is  $d_2$  respect to  $R_{cell} - d_1$ , the lower will be system transmission power.

In the next section, the proposed algorithm is compared to a random user pair selection.

## V. SIMULATION RESULTS

In this section, we compare the performance of the coordinated networks to that of conventional cellular networks with inter-cell interference. After that, we show how the proposed algorithm for user pairs to be served at each physical resource obtains better results in terms of transmission power. First of all, we evaluate the performance of network coordination and we show that coordination is successful in eliminating inter-cell interference. Second, based on simulations results, we compare user selection algorithm with random user selection. Then, we analyze the results and we show how the proposed algorithm minimizes the system transmission power.

The basic experimental setup is based on Fig. 1, two hexagonal cells with a base station located at the center and where each base is loaded with one mobile. The SUI-4 (Modified Stanford University Interim) [12] channel model parameters have been used. Two base stations antennas allow two simultaneously transmissions over 16 subcarriers ( $N_c = 16$ ). As previously described, path loss Hata model in urban areas [2] including lognormal shadowing with a standard deviation of 10 dB is assumed. Cell's radius is 500m and noise power spectral density is -174dBm/Hz [13].

Main configuration parameters are summarized in Table I.

TABLE I  
SIMULATION CONDITIONS

Parameter	Value
Carrier frequency	5 GHz
System bandwidth	1.75 MHz
Sampling frequency	2 MHz
Noise PSD	-174 dBm/Hz
Channel model	SUI-4 [12]
MSs Velocity	5 Km/h
Cell's Radius	500m

Fig. 5 shows system spectral efficiency in bit/s/Hz for coordinated transmission applying channel inversion technique versus conventional not coordinated network.  $P_{tx}$  corresponds to base stations transmission powers,  $P_n$  is the noise power and  $L_{pmax}$  corresponds with the maximum losses that an UE located at cell-edge would suffer. Simulation results are given for a fixed location for both MS1 and MS2. MS1 has been located to  $d_1 = 150$ m and MS2 to  $d_2 = 250$ m, being, as we mention above,  $d_1$  and  $d_2$  distances MS1 to BS1 and MS2 to BS2 respectively.

For the uncoordinated scenario at low  $P_{tx}/P_n - L_{pmax}$  values the noise is the determining factor, while at high  $P_{tx}/P_n - L_{pmax}$  values is the inter-cell interference. Thus, we can observe how at high  $P_{tx}/P_n - L_{pmax}$  values the efficiency is limited by inter-cell interference.

For the coordinated scenario, the determining factor is

only the noise; here inter-cell interference has been cancelled, so the efficiency depends only on SNR values at receivers. Furthermore, it can be observed at low  $P_{tx}/P_n - L_{pmax}$  values the steeped form of the graph corresponding to adaptive modulation transmission. In this case, at high  $P_{tx}/P_n - L_{pmax}$  values, the efficiency is limited for the maximum constellation used (64QAM).

The simulation results show that coherent coordination transmission improves the spectral efficiency by about a factor of 1.875, that is, an enhancement of 87.5% is obtained when coordinated transmission is used.

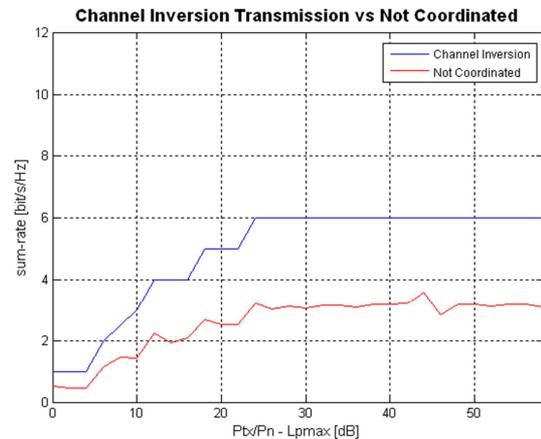


Fig. 5. System spectral efficiency when coordinated Channel Inversion transmission technique is used vs. not coordinated transmission

Fig. 6 shows bit error rate (BER) versus SNR for coordinated transmission when channel inversion precoding is applied.

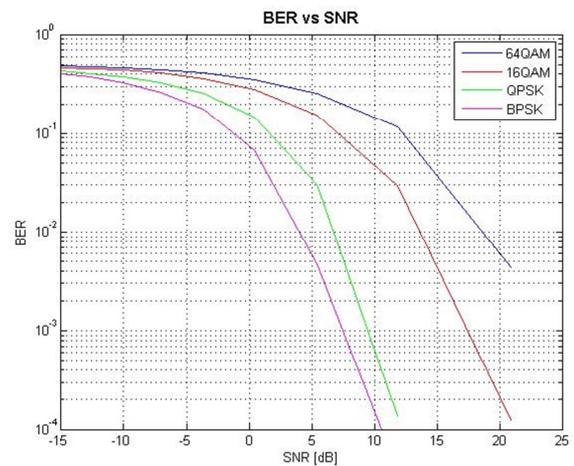


Fig. 6. BER vs. SNR for coordinated transmission

We can observe how inter-cell interference has been cancelled and BER values are the typical of an AWGN channel, as expected according to (6).

Simulation results for system transmission power

according to (10) vs  $d_1$  and  $d_2$  are shown in Fig. 7. The base stations transmission powers configured for both BS1 and BS2 are fixed to  $P_t = 40\text{dBm}$ . MS1 and MS2 locations have been configured in steps of 50m (distance from MS to BS) in order to obtain the grid of possible locations.

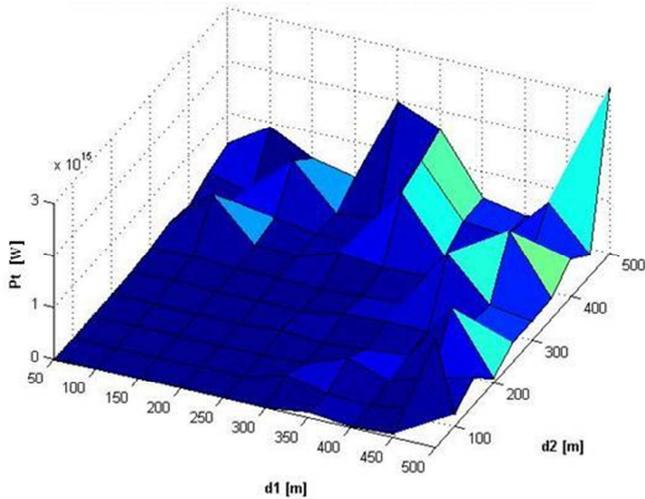


Fig. 7. System Transmission Power vs.  $d_1$  and  $d_2$

Fig. 7 shows that lower system transmission power is obtained when MS1 and MS2 are very close to their respective base stations. The worst or higher system transmission power values are obtained when both MSs are in the cell-edge. It is distinguished approximately two zones separated by those values  $d_1=d_2$ . In order to have lower system transmission power values, pairs of users must be selected in the region area  $d_2 \leq R_{cell} - d_1$  as proposed according to (12). Notice that system transmission power values are unreasonably high. As we commented in Section IV, this is because power normalization factor has not been used in order to study how channel inverse precoding technique affects to system transmission power.

Applying the user selection algorithm proposed for resource allocation in Section IV, in the worst case, that is, when the equality is fulfilled in (12), i.e.,  $d_2 = R_{cell} - d_1$ , ensures keeping in the limit of lower transmission power values zone. A comparison has been done in this case between the proposed user selection algorithm and a random algorithm selecting user pairs randomly. The results have been obtained for ten users at each cell. Five significant different realizations using the same users group and served in the same cyclic order have been used. For user selection algorithm proposed the same user pairs from the users group that fulfill the algorithm in the worst case, i.e.,  $d_2 = R_{cell} - d_1$ , have been used for each realization. From the point of view of the random algorithm, user pairs have been selected randomly for each realization from users group. Results are given in Table II.

TABLE II  
TRANSMISSION POWER

No.	Proposed Algorithm [dBm]	Random Algorithm [dBm]
1	175.2276	176.9869
2	175.2276	178.2176
3	175.2276	175.2295
4	175.2276	176.1282
5	175.2276	177.3151

Table II shows that in all occasions we get better results applying the proposed algorithm in the worst case than applying a random algorithm for selecting user pairs. Approximately, an enhancement of 46.76% on average, computed in linear with Watt, is obtained when user pairs are selected following the algorithm instead of randomly. Obviously, this enhancement is in the worst case; it will be higher if pairs of users are selected inside the zone of lower system transmission power instead of on the limit.

## VI. CONCLUSIONS

In this paper, we have shown that using coordinated transmission in a cellular network improves spectral efficiency. In spite of its simplicity, base station coordination technique of channel inversion achieves its goal of cancelling inter-cell interference. However, it is not free as precoder causes system transmission power fluctuations due to the stringent requirement that the interference at the receivers be identically to zero.

A user selection algorithm has been proposed for resource allocation, where user pairs are selected in the region area  $d_2 \leq R_{cell} - d_1$ . The user pair criterion proposed solely depends on the distance between mobile stations, but it is shown that results with the proposed algorithm are always better, in terms of system transmission power, than applying an algorithm where user pairs are randomly selected.

In future works, the user selection algorithm may be improved. An algorithm also dependent on the users' channel gains not only on their mutual distance could be proposed.

## ACKNOWLEDGMENT

This work has been partially supported by the Spanish Government and FEDER under project TEC2010-18451.

## REFERENCES

- [1] E. Dahlman, S. Parkvall, and J. Sköld, 4G LTE/LTE-Advanced for Mobile Broadband, Elsevier 2011.
- [2] A. J. Goldsmith, Wireless Communications, New York: Cambridge University Press, 2005.
- [3] N. Seifi, Base Station Coordination in Multicell MIMO Networks, Chalmers University of Technology, 2012.
- [4] 3GPP press release, "3GPP Partners propose IMT-Advanced radio," Geneva, October 8, 2009.

- [5] R. Irmer et al., "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Communications Magazine*, Feb. 2011, pp. 102-111.
- [6] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, August 2006, pp. 56-61..
- [7] F. Boccardi and H. Huang, "Limited Downlink Network Coordination in Cellular Networks," *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'07)*, Sept. 2007, pp. 1-5.
- [8] TR 36.814 V1.0.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA Physical layer aspects (Release 9)," Mar 2010.
- [9] M. Sawahashi, Y. Kishiyama, A. Morimoto, D. Nishikawa, and M. Tanno, "Coordinated Multipoint Transmission/Reception Techniques for LTE-Advanced," *IEEE Wireless Communications*, June 2010, pp. 26-34..
- [10] M. R. Ramesh Kumar, S. Bhashyam, and D. Jalihal "Downlink Performance of 2-Cell Cooperation Schemes in a Multi-Cell Environment," 11<sup>th</sup> International Symposium on Wireless Personal Multimedia Communications (WPMC'08).
- [11] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A Vector-Perturbation Technique for Near-Capacity Multi-Antenna Multi-User Communication-PartI: Channel Inversion and Regularization," *IEEE Transactions on Communications*, vol.53, Jan. 2005, pp. 195-202.
- [12] 7-IEEE 802.16.3c-01/29r4, "Channel Models for Fixed Wireless Applications," 2001.
- [13] G. Piro, L. Grieco., G. Boggia., F. Capozzi, and P. Camarda, "Simulating LTE Cellular Systems: An Open-Source Framework," *IEEE Transactions on Vehicular Technology*, vol. 60, no. 2, Feb. 2011 , pp.498,513.

## Enhancement of A5/1 Stream Cipher Overcoming its Weaknesses

Mahdi Madani, Salim Chitroub

Signal and Image Processing Laboratory

Electronics and Computer Science Faculty, USTHB

Algiers, Algeria

e-mail: {mmadani49@gmail.com, s\_chitroub@hotmail.com}

**Abstract**—The Global system for Mobile (GSM) communication is still the most widely used cellular system in the world, with over the billion customers around the world; even the fourth and fifth generations are now operated in some countries. However, GSM bears numerous security vulnerabilities and for that reason it has seriously considered security threats. Although GSMs architecture is designed in such a way to provide various security features like authentication, data and signaling confidentiality, and the user secrecy, the GSM channel is yet susceptible to replaying, interleaving and man-in-the-middle attacks. The GSM voice calls are encrypting a family of algorithms collectively called A5. A5/1 is the stream cipher which encrypts the information transmitted from a mobile user. Initially, A5 algorithm was kept secret to ensure the security, but its algorithm was disclosed many cryptanalytic attacks that have made in evidence the weakness of A5 algorithm. In this paper, after the discussion on the A5/1 encryption algorithm, and the attacks that it has suffered, we will put in evidence the weakness of A5/1. Although attacks that can be used to break the A5/1 algorithm require incredible computing power so that not only certain people with certain computer could break the A5/1 algorithm, however, this does not preclude the necessity of improve safety A5/1. We propose here to enhance the A5/1 stream cipher by overcoming its weakness.

**Keywords**—GSM Networks; Stream Cipher; Mobile Security; Encryption and Cryptography; Cryptanalysis; Non-Linear Boolean Functions; Linear Feedback Shift Register.

### I. INTRODUCTION

Mobile communications have become now more popular and easier. Nowadays, people can communicate with each other on any place at any time. However, the openness of wireless communications poses serious security threats of communicating parties. How to provide secure communication channels is essential to the success of a mobile communication network. Encryption in wireless communication is essential to protect sensitive information, and to prevent fraud. Stream ciphers are symmetric-key ciphers that generate pseudo-random binary sequences, which are used to encrypt the message signals and data wireless communications. Stream ciphers can be designed to be exceptionally fast, much faster than any block cipher. While block ciphers operate on large blocks of data, stream ciphers typically operate on smaller units of plaintext,

usually bits. A stream cipher generates a key-stream and encryption is provided by combing the key-stream with the plaintext, usually with the bitwise XOR operation.

GSM is a widely used mobile standard in the world that uses A5/1 stream cipher for protecting the privacy and secrecy of the subscriber's information over the air interface. However, recent research and studies show that it has some limitations owing to which it is cryptanalyzed by a number of cryptographic attacks. One of the weaknesses of A5/1 is fixed feedback polynomial of Linear Feedback Shift Registers (LFSRs); and other is the weak clocking mechanism. A5/1 was initially cryptanalyzed by Golic [1], when only a rough outline of A5/1 was leaked. Then, A5/1 algorithm was cryptanalyzed by Biryukov, Shamir, and Wagner [2], Bihan and Dunkelman [3], Ekdahl and Johansson [4], Maximov et al. [5], and recently, by Barkan and Bihan [6]. Most of the attacks against A5/1 are known plaintext attacks and use security weakness in the clock-controlling unit [7].

The remainder of the paper is organized as follow. The following section summarizes the design of the A5/1 stream cipher. The third section describes how the algorithm works. The fourth section reviews some of the attacks that can be used against A5/1 algorithm. The fifth section shows the weakness of the A5/1. Our proposed method to enhance the security of the A5/1 stream cipher algorithm is exposed in the sixth section. The last section concludes this work.

### II. DATA ENCRYPTION IN GSM

The GSM standard has its specific algorithms for data encryption and authentication that are grouped in the A5/1 algorithms family. Only the A5/0, which is no an encryption algorithm, A5/1 and A5/2 are the two encryption algorithms stipulated by this standard, where the stream cipher A5/1 is used within Europe and most other countries. A5/2 is the internationally weaker version of A5/1 which has been developed (due to export restrictions) for deploying GSM outside of Europe. A5/3 is a new algorithm based on the Universal Mobile Telecommunications System (UMTS)/Wide Code Division Multiple Access (WCDMA) and Kasumi algorithm [8]. All of these algorithms use 64-bits key [9]. Through terminal of both ciphers were kept secret, their designs were disclosed in 1999 by means of reverse engineering [12]. In this work, we focus on the A5/1, its performances and its weakness.

A5/1 is a synchronous stream cipher that uses a 64 bits session key and an initial vector of 2 bits derived from the 22 bits frame number, which is publicly known. It uses three LFSRs called R1, R2, and R3 of the length 19, 22, and 23 bits respectively, as it is shown in Figure 1. The combined length of the three LFSRs is then 64 bits. The rightmost bit in each register is labeled as bit zero. Three primitive feedback polynomials are used for the three LFSRs R1, R2, and R3, which are:  $x^{19} + x^5 + x^2 + x + 1$ ,  $x^{22} + x + 1$ , and  $x^{23} + x^{15} + x^2 + x + 1$ , respectively. Each register contain a set of bits called taps. The taps of the LFSRs correspond to primitive polynomials and, therefore the registers give sequences of maximum periods. The taps of R1 are at bit positions 13, 16, 17, 18; the taps of R2 are at bit positions 20, 21; and the taps of R3 are at bit positions 7, 20, 21, 22. R1, R2, and R3 are clocked irregularly based on the values of the clocking bits that are at positions 8, 10, and 10 of registers R1, R2, and R3, respectively.

III. HOW A5/1 IS WORKING?

A. Clocking

The stream cipher produces the key-stream by generating one bit at each clock cycle until reached the size of such key-stream. The registers are clocked in a stop/go clock control using a majority rule, which uses three clocking bits C1, C2 and C3 of registers R1, R2 and R3, respectively, and determines the value of the majority bit m using the formula:  $m = \text{maj}(C1, C2, C3)$  among the clocking bits, if two or more are 0 then the value of majority bit m is 0. Similarly, if two or more bits are 1, then majority bit m is 1. After that, if  $C_i = m$  then  $b_i = 1$ , otherwise  $b_i = 0$ . For example, let  $(C1, C2, C3) = (1, 0, 1)$  then, according to majority rule,  $m = 1$ . Now, if  $C_i = m$ , then register  $R_i$  will be clocked (shifted left by 1 bit), where  $i = 1, 2, 3$ . The probability of an individual LFSR being clocked is 3/4. At each clocking cycle, each LFSR generates one bit  $x_i$  which are then combined by a linear combining function  $z(t)$ , to produce one bit of the output key-stream, defined as follows:

$$Z(t) = x_1 \oplus x_2 \oplus x_3. \tag{1}$$

All the possible cases of clocking process are summarized in Table 1.

B. Process

To generate the 228 bits of key-stream, initially, the 64 bits of the session key  $K_c$  and the 22 bits of frame number are used, then the process below is follow-up [11]:

- The first step consists to zeroed the three registers, then for 64 cycles (ignoring the stop/go clock control), the 64 bits of the secret key  $K_c$  (from LSB to MSB) are mixed using the XOR operator, in parallel with the least significant bit of each registers.
- Similarly, the second step consists to clock the registers for 22 additional cycles (ignoring the stop/go clock control). Like the first step, during this

TABLE I. MAJORITY RULE RESULTS

b1	b2	b3	m	c1	c2	c3	Register(s) clocked
0	0	0	0	1	1	1	R1, R2, R3
0	0	1	0	1	1	0	R1, R2
0	1	0	0	1	0	1	R1, R3
0	1	1	1	0	1	1	R2, R3
1	0	0	0	0	1	1	R2, R3
1	0	1	1	1	0	1	R1, R3
1	1	0	1	1	1	0	R1, R2
1	1	1	1	1	1	1	R1, R2, R3

period the 22 bits of Frame number (Fn) (from LSB to MSB) are again mixed in parallel with the least significant bit of each register. The contents of the three registers at the end of this step are called the initial states of the frame.

- In the third step, the three register are clocked for 100 additional cycles with the stop/go clock control, but the output is discarded for these cycles. After this is completed, the cipher is now ready to produce two 114 bit sequences at output, the first 114 bits are for downlink, and the last 114 bits are for uplink.
- The last step consists to clock the three registers for 228 additional cycles with the stop/go clock control in order to produce the 228 output bits. At each clock cycle, one bit is produced as the result of the XOR operation of the most significant bit of the three registers.

As we said above, the GSM conversations are send as sequence of the frames, and the objective is to encrypt each frame before its transmission. So, after generating this pseudo random of 228 bits, they will be mixed using the XOR operator with one frame of the plaintext and then sent the result (one frame of cipher-text) to the other part of the network. Then, the receiver can decrypt the messages using the same process.

IV. ATTACKS AGAINST A5/1

A certain number of serious weaknesses in its coding mechanism were identified. Thereafter, it was targeted by several attacks which we can classify in two categories. First, Brute force attacks are based on searching to decrypt the conversations by trying all the  $2^{64}$  computations of  $K_c$ .

To decrypt one frame of 228 bits, the operation tacks about 3-6 years. Since the decryption is not in real-time, the conversation maybe meaningless after 3-6 years. Second, Look Up Table Attacks need in total of  $2^{64}$  memories to store the look up table in total. That cost about 18446744 terabytes in total, and we also need a strong computing power to process such big data like that.

The security of the A5/1 encryption algorithm was analyzed in several papers [1][2][12][13]. The known attacks can be summarized as this:

- Briceno [12] found out that in all the deployed versions of the A5/1 algorithm, the ten least significant of the 64 bits of the keys were always set

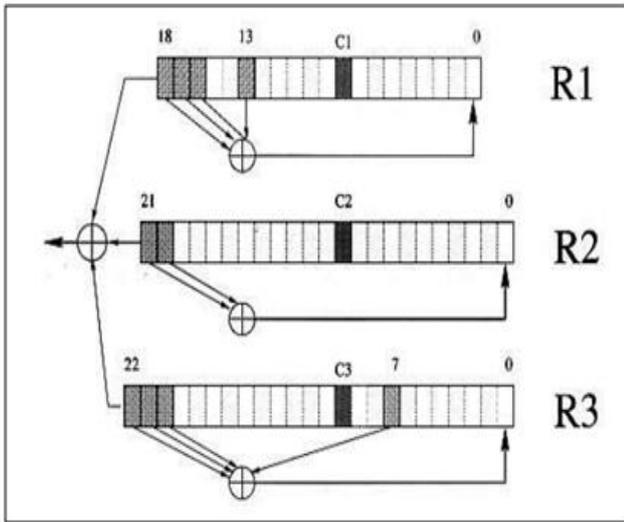


Figure 1. A5/1 stream cipher

to zero. The complexity of exhaustive search is thus reduced to  $O(2^{54})$ .

- Briceno [12] found out that in all the deployed versions of the A5/1 algorithm, the ten least significant of the 64 bits of the keys were always set to zero. The complexity of exhaustive search is thus reduced to  $O(2^{54})$ .
- Anderson and Roe [13] proposed an attack based on guessing the 41 bits in the shorter registers, R1 and R2, and deriving the 23 bits from the longer register, R3, of the output. However, they occasionally have to guess additional bits to determine the majority-based clocking sequence, and thus the total complexity of the attack is about  $O(2^{45})$ . This attack needs more than one month to find one key.
- Golic [1] described an improved attack which requires  $O(2^{40})$  steps. This attack is based on the solution of a system of linear equations. It takes, however, more time than the previous algorithm.
- Golic [1] describes a general time-memory tradeoff attack on stream ciphers. It was independently discovered by Babbage [1] two years earlier. Golic [1] concludes that it is possible to find the key, of the A5/1 algorithm, in  $O(2^{22})$  probes into random locations of a computed table with 242128 bit entries. Since such a table requires a hard disk of 64 terabytes, the space requirement is unrealistic. Alternatively, it is possible to reduce the space requirement to 862 gigabytes, but then, the number of probes increases to  $O(2^{28})$ . Since random access to the fastest commercially available PC disks requires about 6 milliseconds, the total probing time is almost three weeks. And it can only be used to attack GSM phone conversations which last more than 3 hours, and so it is unrealistic.

- Biryukov et al. [2] proposed two new attacks cryptanalytic on A5/1, in which a simple PC can extract the key from conversation in real time starting from some output generated. The first attack (called the biased birthday attack) requires two minutes of data and one second of processing time, whereas the second attack (called the random subgraph attack) requires two seconds of data and several minutes of processing time.

### V. WEAKNESS OF A5/1

The major problem of stream ciphers, as the algorithm A5/1, is the difficulty of generating a long unpredictable bit pattern (key-stream). According to the results reached by the operations and the cryptanalysis made by the researchers, in addition to the various attacks listed in preceding section, the algorithm A5/1 suffers from several weaknesses. For this reason, the A5/1 stream cipher was classified in the column of low level of security algorithms. The set of huge weaknesses is the following.

- The first weakness is in the generation of the output sequence which is ensured by a linear function (simple XOR), which is considered weak to crack by a linear cryptanalysis.
- The second one is about the short period problem: without stop/go operation, the period clocking of sum of the three LFSRs is given by:

$$(2^{19}-1)(2^{22}-1)(2^{23}-1). \tag{2}$$

However, the experiments show that the period of A5/1 is around  $(4/3)(2^{23}-1)$ .

- The third weakness is caused by the collision problem: the different seeds (i.e., different initial states of three LFSRs) may result in the same key-stream and so, only about 70% of seeds that really generate different key-streams.
- The fourth weakness is the majority function that is the worst function in terms of correlation with all affine functions.

### VI. IMPROVED MODEL FOR A5/1

Before describing in detail our proposed improved model for A5/1, we recall that the basic model suffers from fore large weaknesses. In this paper, we propose two solutions. The first one is for improving the clocking mechanism, which is considered as one of the major weaknesses of A5/1. The second solution is concerning the problem of the linear output function that is reversible.

Therefore, the proposed model aims at improve those two weaknesses. The clocking mechanism is improved by a new function and the linear combining function of A5/1 is replaced by a more cryptographically better nonlinear function to strengthen the cipher. The architecture of the proposed scheme is shown in the Figure 2, and the model is detailed as follows:

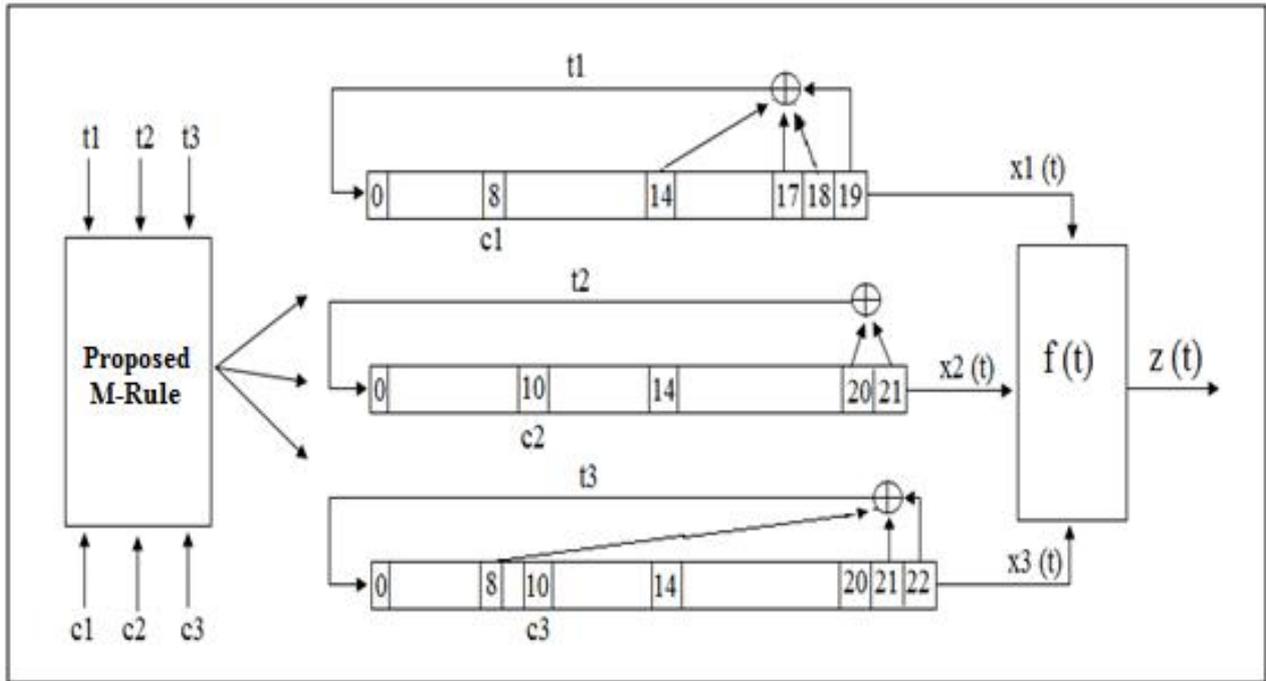


Figure 2. Proposed modified A5/1

#### A. Improved Unit of Clock-Controlling

As we mentioned in the previous section, we propose a new model to ameliorate the flaw of the clocking mechanism. The clock-controlling unit of the proposed scheme uses a new different rule to clock the registers. Instead of the existing M-rule, which is based on the one clocking taps bit (only one bit by register), the proposed rule consists of clocking each register by the output of its taps of the bits in combination with its clocking taps as well.

This mechanism can be governed by our proposed M-rule, which is defined by the following formula:

$$m = C1 \cdot t1 \oplus C2 \cdot t2 \oplus C3 \cdot t3. \quad (3)$$

where “.” represents the AND operation, and  $\oplus$  represents the XOR operation.  $t1$ ,  $t2$ ,  $t3$  are the outputs of the taps of the bits of the three registers (see Figure 2). They are defined by the three following formulas:

$$t1 = R1(13) \oplus R1(16) \oplus R1(17) \oplus R1(18).$$

$$t2 = R2(19) \oplus R2(20).$$

$$t3 = R3(7) \oplus R3(19) \oplus R3(20) \oplus R3(21).$$

where  $C1$ ,  $C2$ , and  $C3$  are the clocking taps of the registers  $R1$ ,  $R2$  and  $R3$ , respectively. Once the majority bit is defined, we compare this bit with the  $t$  bit of each register, and then the decision will be taken. The register will be clocked only if the two bits are much. In other words, if  $m = t(i)$ , for  $i = 1, 2$  or  $3$ ,  $R(i)$  will be clocked. The proposed M-rule is illustrated in the Figure 3.

#### B. New Non-linear Combination Function

A5/1 stream cipher generates one bit at each clocking cycle. As we mentioned in the previous sections, the output of A5/1 is ensured by combining each the output bits of the three LFSRs using the XOR operation (see Figure 4). But the attackers prove that this function is weak and it is easy to recover the inputs data. In fact, this function is weak like all the linear functions because any linear function is reversible. This problem is treated by the searchers in the literature. Sarkar and Maitra [15] say that the linear combination functions are cryptographically weak functions.

Thus, it is necessary to replace these functions with another one more secure and more complex. Thus, to increase the linear complexity of the A5/1 stream cipher and to overcome the weaknesses, due to the use of the linear function for combining, a new non linear combination function, which is more cryptographically better, is proposed here. The scheme of the proposed function is shown in Figure 5. Such proposed function is defined by the following formula:

$$Z(t) = f(x1, x2, x3) = (x1 \cdot x2) \oplus (x1 \oplus x3) \cdot (x2 \cdot x3). \quad (4)$$

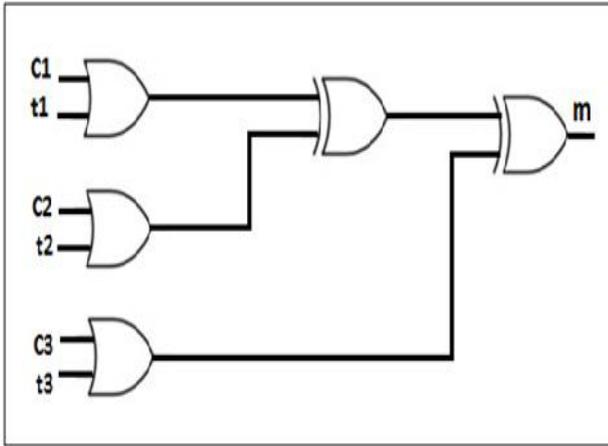


Figure 3. Proposed M-rule

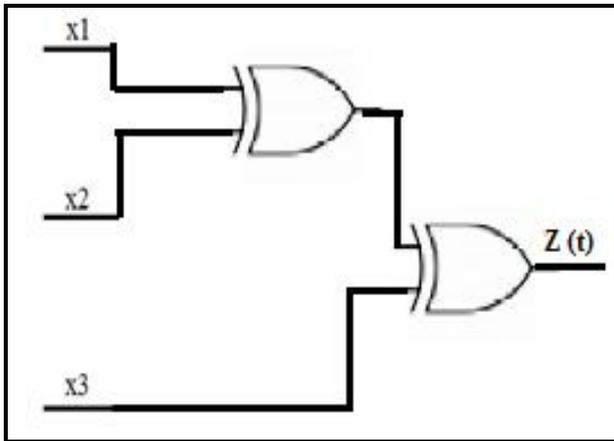


Figure 4. Original linear Function

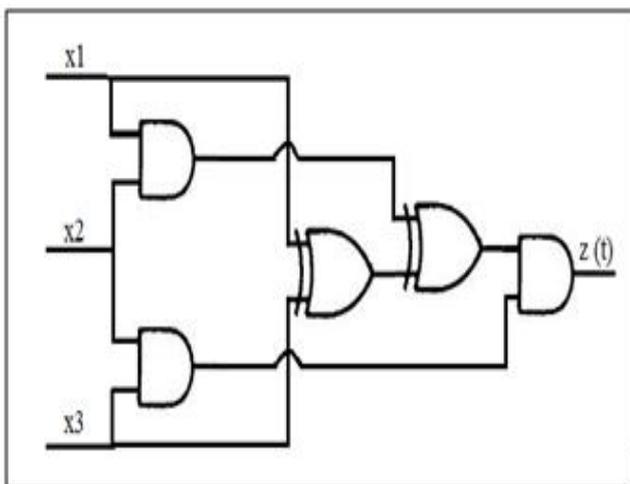


Figure 5. Proposed Non-linear Combination Function

where “.” represents the AND operation, and  $\oplus$  represents the XOR operation.  $x_1, x_2, x_3$  are the output bits of the three registers after clocking cycle, and  $z(t)$  is the output of the proposed scheme.

Like in the original algorithm, one bit is generated in each clocking cycle. Our proposed scheme is also based on the output of the three LFSRs, but the difference appears in the combining function, when the linear XOR operation is replaced by a new non-linear function  $f(x)$ .

Now, we will show that the proposed scheme is more secure than the existing one because of its complexity and its non linearity. In fact, because of the clocking mechanism that becomes more irregular, the new nonlinear combining function is more secure than the existing one. However, because of the restriction of the number of the pages for this paper, we cannot detail this point here, but it will be included in the future works.

### VII. SIMULATION RESULTS

This section presents the simulation results of the system that we have implemented using the proposed new non-linear combination function that we have detailed in Section VI. Note that for implementing our proposed system, we must go through several steps that we have described in details in the preceding sections. We have the preliminary results are shown in Figure 6. The difference between the linear output signal (above) and non-linear output signal (below), in terms of irregularities and complexities, is remarkable. It is due to the random aspect of the new non-linear function. In the future works, we intend to make the test of the effectiveness of our proposed system by securing data in GSM networks.

### VIII. CONCLUSION AND DISCUSION

The aim behind the proposed enhancements of the A5/1 encryption algorithm, used in GSM standard, is to increase the linear complexity of the generated output sequence, and to accelerate the clocking mechanism. So, we have resolved two from the four problems listed in this paper that from which the A5/1 stream cipher suffers.

The proposed scheme can also retain the speed and it can be used for transmitting data in real time. Even though the algorithm became complex, it is easy in its implementation. We note that there is no explanation regarding the reason for selecting bits in the LFSR registers as tap bits because of the original architecture of A5/1, which is until this time unknown. There is no official description of A5/1 in the literature. The details of the algorithm have been published anonymously on a mailing list on the Internet, presumably due to a reverse engineering. So, we can conclude that the proposed scheme is better against to the cryptographic attacks compared to the conventional A5/1 stream cipher, since it generates cryptographically better binary sequence than the conventional one with slight increase in the hardware implementation.

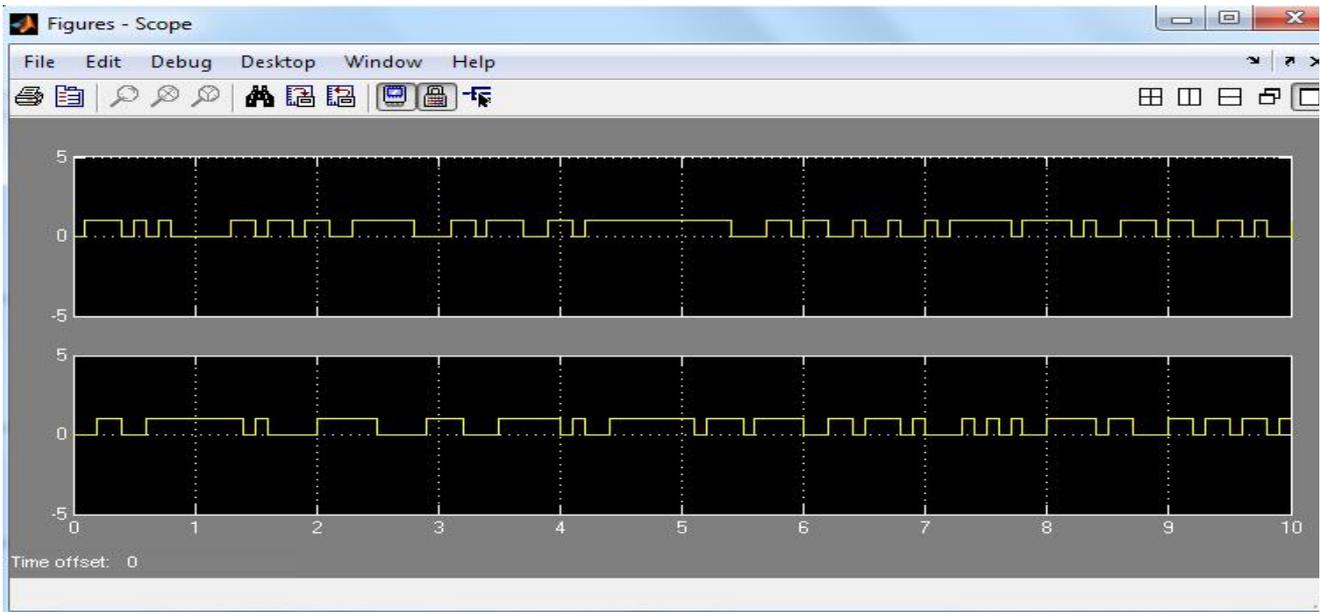


Figure 6. Comparison between linear and non linear output signal

Regarding the cryptanalysis of the methods used in A5/1 algorithm versus proposed methods, the linear Boolean function-based attacks are not possible because our proposed majority function is a non-linear function. While for the non-linear function-based attacks, it will be very difficult since our majority function is based on new rule that is unknown.

We note that there are still other possibilities to improve the A5/1 algorithm to render it more secure without decreasing its speed. These possibilities will be the subject of our future works.

#### REFERENCES

- [1] J. Golic, "Cryptanalysis of alleged A5 stream cipher," *Advances in Cryptology. EUROCRYPT97*. LNCS, vol. 1233, 1997, pp. 239-255, Springer-Verlag.
- [2] A. Biryukov, A. Shamir, and D. Wagner, "Real time cryptanalysis of A5/1 on a PC," *Advances in Cryptology. Fast Software Encryption*. LNCS, vol.1978, 2001, pp. 1-18, Springer-Verlag.
- [3] E. Biham and O. Dunkelman, "Cryptanalysis of the A5/1 GSM stream cipher," *Progress in Cryptology, proceedings of INDOCRYPT00*. LNCS, 2000, pp. 43-51, Springer-Verlag.
- [4] P. Ekdahl and T. Johansson, "Another attack on A5/1," *IEEE Transactions on Information Theory*, vol. 49, Jan. 2003, pp. 284-289.
- [5] A. Maximov, T. Johansson, and S. Babbage, "An improved correlation attack on A5/1," *SAC 2004*. LNCS, vol. 3357, Aug. 2004, pp. 1-18.
- [6] E. Barkan and E. Biham, "Conditional estimators: an effective attack on A5/1," *SAC 2005*. LNCS, vol. 3897, 2006, pp. 1-19, Springer-Verlag.
- [7] S. E. AlAschkar and M. T. El-Hadidi, "Known attacks for the A5/1 algorithm," *International Conference on Information and Communications Technology (ICICT03)*, 2003, pp. 229-251.
- [8] R. Alpesh Sankaliya, V. Mishra, and A. Mandloi. "Implementation of Cryptographic Algorithm for GSM and UMTS Systems," *International Journal of Network Security & Its Applications (IJNSA)*, Vol.3, No.6, November 2011, pp. 81-88.
- [9] G. Rose, "A précis of new attacks on GSM encryption," *Qualcomm, Australia*, Sept. 2003.
- [10] A. Canteaut. Available from: <http://www.springerreference.com/> [retrieved: December, 2013]
- [11] J. Fernando, "Attacks on A5/1 Cryptography Algorithm." *Makalah IF3058 Kriptografi Sem. II Tahun*, [retrieved: December, 2013].
- [12] M. Briceno, I. Goldberg, and D. Wagner, "A pedagogical implementation of A5/1". [Online]. Available from: <http://cryptome.org/jya/a51-pi.htm/> [retrieved: December, 2013].
- [13] R. Anderson and M. Roe, A5. Available from <http://cryptome.org/jya/crack-a5.htm/>[retrieved: April, 2014].
- [14] S. Babbage, "A Space/Time Tradeoff in Exhaustive Search Attacks on Stream Ciphers," *European Convention on Security and Detection*. IEEE Conference publication, No. 408, May 1995, pp. 216-224.
- [15] P. Sarkar and S. Maitra, "Construction of nonlinear Boolean functions with important Cryptographic properties," *Advances in Cryptology EU-ROCRYPT00*. LNCS, vol. 1807, 2000, pp. 485-506, Springer-Verlag.

# Enhancing Safety-Critical Message Dissemination in WAVE

Daniel Neville, James Brill, Joseph Juzl, Christopher Howell and Victor Sanchez

Department of Computer Science, The University of Warwick

Coventry, United Kingdom

{D.Neville, J.E.Brill, J.Juzl, C.K.Howell}@warwick.ac.uk, vsanchez@dcs.warwick.ac.uk

**Abstract**—The competing priorities of safety critical messages and infotainment messages present a significant challenge when designing effective IEEE 1609.4 protocol enhancements. In this paper, we investigate the latency reduction provided by an additional CCH check back midway through the Service Channel interval. Mathematical analysis and experimentation using simulations have shown that this method results in significant latency reduction. The maximum transmission delay for safety-critical messages can be reduced by approximately half whilst only reducing the Service Channel capacity by one quarter. As the work progresses, we will optimise the duration of the check back to find the best compromise between safety and infotainment.

**Keywords**—latency reduction; VANET; WAVE; 1609.4; safety messages

## I. INTRODUCTION

In recent years, the interest in Vehicular Ad Hoc Networks (VANETs) has increased considerably, particularly to improve the safety and efficiency of transportation networks through wireless communications. VANETs comprise vehicles equipped with transceivers capable of exchanging information either Vehicle-to-Vehicle (V2V) or Vehicle-to-Infrastructure (V2I). While V2I communication requires huge infrastructure investment, V2V communication is more viable. The Wireless Access in Vehicular Environment (WAVE) standardised communication technology has been recently designed specifically for VANETs. WAVE includes the 1609 protocol family, as well as the IEEE 802.11p standard, a redesigned version of IEEE 802.11 that reduces the use of control packets and authentication in order to ensure swift delivery of data [1]. The 1609.4 protocol sits on top of 802.11p and allows for multi-channel communication over a single radio. Infotainment and safety applications can then coexist as the 1609.4 protocol [2] alternates periodically between a Control Channel (CCH) and one of six Service Channels (SCHs).

Infotainment applications in VANETs are aimed at enhancing the driving experience by providing non-safety applications, such as entertainment services [3]. Such services are transmitted as infotainment messages over one of the SCHs. Safety applications, on the other hand, are aimed at assisting drivers with real-time information about road and traffic conditions in order to reduce the number of accidents caused by human error. These applications are transmitted as safety-critical messages over the CCH. Examples of safety-critical applications are Cooperative Collision Warning (CCW), Elec-

tronic Emergency Brake Light (EEBL) and Slow/Stopped Vehicle Alert (SVA). Low latency communication is essential for these applications to be effective.

In this work in progress, we are particularly interested in improving dissemination of safety-critical messages while allowing the coexistence of infotainment applications. A number of solutions to this problem have already been proposed. Ghandour et al. [4] propose an enhancement in which the SCHs are made available for safety-critical messages, with nodes informing their neighbours of which SCH they intend to use for safety-critical transmissions. They also propose a second enhancement with the aim of mitigating the issue of synchronous collisions that can occur between transmitted safety-critical messages. This involves measuring the level of contention at the start of the safety-critical transmission interval and increasing the probability of transmissions being deferred if contention levels are high, spreading safety-critical transmissions out more evenly to reduce collisions.

Mak et al. [5] attempt to find a balance between safety-critical and infotainment transmissions by assuming safety only communications between vehicles by default while adopting a roadside access point to handle the coordination of infotainment services.

Jiang et al. [6] propose a number of alterations to 1609.4, including an ECHO protocol in which nodes receiving safety-critical messages rebroadcast them if they have not heard them recently. This improves the probability of reception due to multiple transmissions of the same safety-critical message, a useful enhancement for a protocol that operates in a dynamic environment in which packets can easily be lost. Their paper also outlines a Piggybacked Acknowledgement (PACK) protocol, which 'piggybacks' acknowledgements on safety messages to allow safety-critical broadcasters to get feedback on the performance of their safety-critical broadcasts and retransmit their safety-critical message if the failure rate is too high.

Our work in progress involves a simple solution that increases the time in which vehicles access the CCH for transmission of safety-critical messages. This solution preserves the overall time structure of the multi-channel communication in the 1609.4 protocol, while reducing latency.

The rest of the paper is organised as follows. Section II briefly reviews the time division in the 1609.4 protocol and provides a mathematical analysis of transmission delay.

Section III details our proposed solution. Experimental results over a simulated VANET using NS-2 are presented in Section IV. Conclusions are presented in Section VI.

## II. THE 1609.4 PROTOCOL

In the 1609.4 protocol, vehicles switch frequencies between the CCH and one of six different SCHs, while maintaining synchronisation with the Coordinated Universal Time (UTC) via Global Positioning System (GPS) devices. The time period in which vehicles are tuned to the CCH and then one of the SCHs is called the SYNC interval, which lasts 100 ms. More specifically, each SYNC interval comprises a CCH interval, which lasts 46 ms, followed by a SCH interval, which also lasts 46 ms. There is a 4 ms guard period before switching to the CCH and the SCH interval. This is illustrated in Figure 1. The SYNC interval repeats indefinitely.

Under the existing 1609.4 protocol, vehicles spend as much time on the SCHs as they do on the CCH. Given the importance of safety applications, we argue that safety-critical messages should be given a higher priority and greater access to the medium than infotainment messages, which are currently given equal access within each cycle. This is a very reasonable assumption considering the recent advancements in 4G/LTE that permit very high data rates and could be used for transferring large files that do not have significant latency requirements.

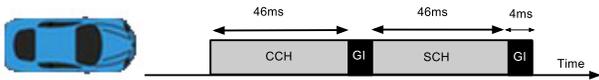


Figure 1. The SYNC interval in the 1609.4 protocol.

### A. Transmission delay of safety-critical messages

We can perform mathematical analysis on this protocol under the assumption that the time at which a packet is sent within the SYNC interval is chosen randomly from a uniform distribution  $\mathcal{U}(0\text{ms}, 100\text{ms})$ . This is realistic because in a real world scenario there is no reason that an emergency event should fall within a specific interval. Additionally, this analysis assumes that the channel is clear upon transmission, and the packet, if sent at any point within the CCH interval, is received successfully. Under these assumptions, the expected delay  $E[\text{Delay}]$  to receive a safety message sent on the channel may be calculated as follows:

$$E[\text{Delay}] = \delta + \sum_{i=1}^4 \left( \frac{\text{duration}_i}{\text{duration}_{\text{total}}} \times \text{CCH Delay}_i \right) \quad (1)$$

where  $\delta$  represents the average propagation delay for a packet sent using Nakagami Propagation [7],  $\text{CCH Delay}_i$  represents the average time between interval  $i$  and the next CCH interval. Note that (1) assumes four intervals, the CCH and SCH intervals plus two guard periods. Should a packet be sent within the CCH, the expected delay is simply  $\delta$ .

Using (1) and the channel durations shown in Figure 1 we can calculate the expected latency for a packet within the 1609.4 protocol. The results of this are shown in Table I.

TABLE I. EXPECTED DELAY OF A SAFETY-CRITICAL PACKET IN THE 1609.4 PROTOCOL.

Interval	Duration (s)	CCH Delay (s)	Product
(1) CCH	0.046	0	0
(2) GUARD	0.004	0.052	0.00208
(3) SCH	0.046	0.027	0.01242
(4) GUARD	0.004	0.002	0.00008
Total	0.100		0.01458
		$+ \delta = 0.00215$	0.01673

As shown in Table I, the expected delay for a safety-critical message sent under the assumptions listed previously is 0.01673 seconds.

## III. THE PROPOSED SOLUTION

Currently, the main issue with the SYNC interval in the 1609.4 protocol is the disproportionate amount of time spent by vehicles on one of the SCHs, which prevents detection of any safety-critical messages. Our proposed solution is therefore to introduce an additional CCH interval within the SCH interval to reduce the latency of safety-critical message delivery. In other words, vehicles now tune back to the CCH during the SCH to transmit/receive pending safety-critical messages. This second CCH interval, referred to as CCH\_CHECK, occurs halfway through the SCH interval such that each newly created half of the SCH interval, referred to as SCH\_1 and SCH\_2 respectively, are of equal length. A 4ms guard period still exists before intervals. This is illustrated in Figure 2.

The CCH\_CHECK is  $k$  ms long. The benefits gained by using different values of  $k$  are being investigated through ongoing research, with initial forays investigating the benefits of  $k = 4$  ms. This initial value has been chosen as it allows sufficient time to send safety-critical messages without significantly impeding on the SCH duration.

Under this initial proposal, the CCH\_CHECK is to be treated as a normal CCH interval for safety-critical messages, but is not used for low priority safety-critical messages. Should all nodes successfully transmit all queued safety-critical messages, they will wait until the full  $k$  ms have elapsed before returning to the SCH via a GUARD period. An alternative method is discussed in Section V.

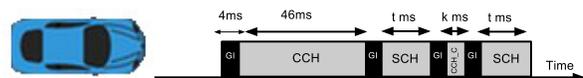


Figure 2. Our proposed schedule.

$$t = \frac{(46 - (2 \times \text{GUARD}) - k)}{2} = 19 - \frac{k}{2} \text{ ms} \quad (2)$$

Our proposed solution preserves the overall timing structure of the 1609.4 protocol, with a 100 ms SYNC duration split evenly between the CCH interval and the SCH interval. The SCH interval, however, now comprises a short CCH interval. The length of each SCH interval is half the remaining time period as shown in (2). The proposed intervals are summarised in Table II.

TABLE II. PROPOSED INTERVALS.

Interval	Duration (ms)
CCH	46
GUARD	4
SCH_1	$19 - \frac{k}{2}$
GUARD	4
CCH_CHECK	$k$
GUARD	4
SCH_2	$19 - \frac{k}{2}$
GUARD	4
Total	100ms

In the current 1609.4 protocol, when a safety-critical message is generated outside of the CCH interval, it is queued and sent in the following CCH interval. The worst-case scenario is when a safety-critical message is generated at the end of the current CCH interval and the time to transmit that message is longer than the time left in the CCH interval. In this case, the message is queued and scheduled for transmission at the beginning of the following CCH interval, which can cause a wait of 54 ms. In the proposed solution, there is at most a  $27 - \frac{k}{2}$  delay; with  $k = 4$  ms, this represents only a 25 ms delay, which is less than half the original delay.

It is important to note that at the beginning of a CCH interval in the existing 1609.4 protocol, there may be a high risk of packet collision because vehicles attempt to transmit all the safety-critical messages queued up since the last CCH interval, which may have accumulated over as long as 54 ms. The proposed solution reduces this to  $27 - \frac{k}{2}$ . An open question that requires investigation is how long  $k$  should be in order to successfully transmit all the safety-critical messages generated in previous SCH intervals, including the guard periods.

The proposed solution inevitably has a detrimental effect on the transmission of infotainment messages. However, this detrimental effect is expected to be minimal in comparison to the positive effects on the dissemination of safety-critical messages, which is arguably the most important aspect of WAVE, particularly during emergency events.

#### A. Expected transmission delay of safety-critical messages

Using the same assumptions and calculations as those used in Section II-A, Table III provides a mathematical analysis of the expected delay of a packet which is sent using the proposed solution to neighbours within broadcast range of the transmitting node, with a value  $k = 4$  ms. Note that the expected latency is 8.4 ms, which represents an improvement of  $\frac{0.0084}{0.01673} \approx 50\%$ . The minimum expected delay reduction is therefore  $0.01673s - 0.0084s = 0.00833s$ .

TABLE III. EXPECTED DELAY OF A SAFETY-CRITICAL PACKET IN OUR PROPOSAL FOR  $k = 4$  MS.

Interval	Dur. (s)	CCH Delay (s)	Product
(1) CCH	0.046	0	0
(2) GUARD	0.004	0.023	0.00092
(3) SCH_1	0.017	0.0125	0.00213
(4) GUARD	0.004	0.002	0.00008
(5) CCH_CHECK	0.004	0	0
(6) GUARD	0.004	0.023	0.00092
(7) SCH_2	0.017	0.0125	0.00213
(8) GUARD	0.004	0.002	0.00008
Total	0.100		0.00625
		$+ \delta = 0.00215$	0.0084s

The proposed reduction on the SCH interval reduces the amount of time available for transmission of infotainment messages. Our proposed solution reduces the SCH interval from 46 ms to  $2 \times (19 - \frac{k}{2}) = 34$ ms, when  $k = 4$  ms, which represents a SCH capacity reduction of  $\frac{46-34}{46} \approx 26\%$

Using the same method as shown in Table III after some simplification, (3) shows the expected delay with an arbitrary  $k$  value.

$$\frac{(k - 54)^2}{400} + \delta \text{ ms} \quad (3)$$

#### IV. EXPERIMENTAL RESULTS

NS-2 [8] was used to simulate the 1609.4 protocol with the proposed solution with a value of  $k = 4$  ms for the CCH\_CHECK interval.

The simulations were ran 2,000 times with each simulation lasting 100 seconds. Within each run, one safety-critical message was sent. This is to represent one emergency event occurring in each run. The time the packet is sent is determined by the same uniform random distribution as used in the mathematical analysis in Sections III-A and II-A. A total of 100 simulated vehicles were spread across a 1 km road with 4 lanes. Each vehicle was randomly located within the road and in this initial simulation, vehicles were static. The Nakagami Propagation model was used for all simulations with a minimum transmission range of 250 m and an event horizon of 1km. Other parameters in 802.11p were set to their default values. When nodes receive a safety-critical message, their distance from the initial transmitting node and the time difference between transmission and reception of the message is recorded within a database. This is later used for analysis.

Since generating the data used in this document, several mobility traces using the Simulation of Urban MObility (SUMO) [9] tool have been generated that will enable the simulation of more realistic roads and vehicle movements. SUMO is an open source, highly portable, microscopic and continuous road traffic simulation package designed to handle large road networks. Also, added is the ability to send arbitrarily large numbers of messages within the simulation, which is to be used in later simulations.

Table IV tabulates the average delay for a single safety-critical message for different distances. Note that for those

nodes within the broadcast range (0-250m), the improvement is between 8.4 – 8.8 ms. This is very close to, but slightly higher than, the predicted value of 8.33 ms, as computed in Section III-A. When nodes are significantly outside the broadcast range, multi-hop transmissions pass the message onto all nodes until the event horizon is reached. The event horizon ensures packets are not forwarded indefinitely.

Figure 3 plots the average delay discretised in Table IV.

TABLE IV. AVERAGE DELIVERY DELAY FOR A SINGLE SAFETY-CRITICAL MESSAGE.

Distance	1609.4 protocol	Proposed solution	Improvement
0-100m	16.989 ms	8.278 ms	8.711 ms
100-200m	16.637 ms	8.286 ms	8.351 ms
200-300m	16.835 ms	8.054 ms	8.781 ms
300-400m	21.864 ms	12.762 ms	9.102 ms
400-500m	28.679 ms	20.700 ms	7.979 ms
500-600m	34.156 ms	25.359 ms	8.797 ms
600-700m	38.809 ms	28.832 ms	9.977 ms
700-800m	43.418 ms	31.740 ms	11.677 ms
800-900m	46.950 ms	34.883 ms	12.068 ms

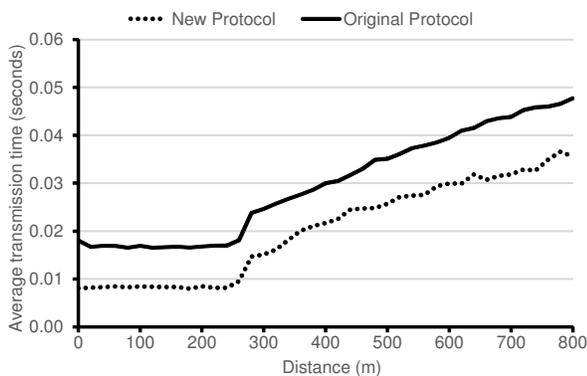


Figure 3. Average delivery delay for a safety-critical message sent at a random instant of time within the simulation.

An important metric to analyse is the latency incurred when safety-critical messages are generated at the very end of the CCH interval, i.e., generated at 46ms into the cycle, which represents a worst-case scenario. Let us recall that under the 1609.4 protocol, these messages would be queued to be sent in the next CCH interval. In the proposed solution, these messages can now be sent within the CCH\_CHECK interval. Table V shows the improvement attained by our proposal for nodes within the broadcast range of the transmitting node.

TABLE V. LATENCY TO ONE HOP NEIGHBOURS FOR THE TRANSMISSION OF A SINGLE SAFETY-CRITICAL MESSAGE GENERATED AT THE END OF THE CCH INTERVAL.

1609.4 protocol	Proposed solution	Improvement	Reduction
55.61 ms	26.60 ms	29.01 ms	52.2%

As can be seen in Table V, the reduction in latency in this worst case scenario is over 52%, a significant reduction.

## V. FURTHER RESEARCH

Our work will next explore the advantages gained by the modification of the  $k$  value. We intend to focus on the range  $k = \{2..16\}$  ms to ensure that sufficient time is available for the SCH interval. Additionally we will investigate benefits gained by immediately returning back to the SCH if there are no messages to be sent during the CCH\_CHECK. An example case we may investigate is that if no messages are sent within the first  $\frac{k}{2}$  milliseconds then those inactive nodes can re-tune to the SCH and continue offering infotainment services. An additional point of investigation is how our modified protocol handles heavy busy environments, with particular interest in how effectively collisions are handled.

We would like to examine the effects of our protocol on a real-world WAVE application such as EEBL, comparing the existing 1609.4 protocol with our proposed solution in an appropriate SUMO simulation. We will compare the performance of our solution with similar solutions proposed by others.

## VI. CONCLUSION AND FUTURE WORK

This paper presented our work in progress on enhanced dissemination of safety-critical messages in the 1609.4 protocol for VANETs. Specifically, we presented the idea of introducing CCH interval within each SCH interval. Our initial results are greatly encouraging and appear to satisfy our objective, which is to attain a significant reduction in safety-critical message latency with minimal reduction in SCH duration. We are continuing to investigate the benefits generated by the introduction of this short CCH interval and explore alternative modifications. As part of our future work, we are interested in finding the best trade-off between the enhancement of safety-critical message dissemination and the reduction in service channel capacity.

## REFERENCES

- [1] J. B. Kenney, "Dedicated short-range communications (DSRC) standards in the United States.", Proceedings of the IEEE, 99(7), July 2011, pp. 1162-1182.
- [2] IEEE 802.11 Working Group, "IEEE standard for wireless access in vehicular environments multi-channel operation", Feb 2011, pp. 1-89.
- [3] M. Amadeo, C. Campolo, and A. Molinaro, "Enhancing IEEE 802.11 p/WAVE to provide infotainment applications in VANETs.", Ad Hoc Networks 10(2), March 2012, pp. 253-269.
- [4] A. J. Ghandour, M. D. Felice, H. Artail and L. Bononi, "Modeling and Simulation of WAVE 1609.4-based Multi-channel Vehicular Ad Hoc Networks", 5th ACM International Conference on Simulations Tools and Techniques (SIMUTools 2012), Italy, March 2012, pp. 148-156.
- [5] T. K. Mak, K. P. Laberteaux, and R. Sengupta, "A multi-channel VANET providing concurrent safety and commercial services.", Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks, September 2005, pp. 1-9.
- [6] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, and R. G. Herrtwich, "Design of 5.9 GHz DSRC-based vehicular safety communication.", Wireless Communications, IEEE 13(5), October 2006, pp. 36-43.
- [7] K. W. Yip and T. S. Ng, "A Simulation Model for Nakagami-m Fading Channels,  $m < 1$ ", Transactions on Communications, IEEE 48(2), February 2000, pp. 214-221.
- [8] Network Simulator ns-2, [Online] <http://www.isi.edu/nsnam/ns/>, [Retrieved: April 2014]
- [9] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker. "Recent Development and Applications of SUMO - Simulation of Urban MOBility.", International Journal On Advances in Systems and Measurements. 5(3&4), December 2012, pp. 128-138.

# Investigating Bit Error Patterns for Radar Pulse Detection in IEEE 802.11

Claudio Pisa, Andres Garcia-Saavedra, and Douglas J. Leith

Hamilton Institute, NUI Maynooth  
Maynooth, Ireland

claudio.pisa@uniroma2.it, {andres.garciasaavedra, doug.leith}@nuim.ie

**Abstract**—The shift towards use of the 5 GHz band by WiFi networks comes with the requirement that these networks coexist safely with existing systems using that band, e.g., meteorology, aeronautics or military radars. Regulatory bodies are mandating the implementation of Dynamic Frequency Selection (DFS) in wireless communication protocols to (i) detect radar operations and (ii) move away from channels populated by these. Conventional approaches to implementing such mechanisms, however, can result in massive underutilization of the radio channel since the radio must be kept silent for a large amount of time in order to ensure sufficient detection accuracy. This inevitably impacts the throughput capacity of the wireless network. In this paper, we consider whether bit-error patterns at the receiver of a WiFi link can be used for radar detection. In our experimental study, we adopt a pair-packet transmission technique to mitigate the interference inherent to the 802.11 protocol due to, e.g., other contending stations. Our initial results show that the observation of bit-error patterns due to radar interferences is indeed possible, establishing that the potential exists to design unobtrusive detection mechanisms that work transparently with existing network protocols without loss of network capacity.

**Keywords**—Radar; 802.11; Interference management.

## I. INTRODUCTION

The growing density of radio frequency (RF) transceivers operating in the ISM 2.4 GHz band (Industrial, Scientific and Medical radio band) is placing increasing pressure on the limited spectral resources available (up to 13 channels spaced 5 MHz apart) for e.g., IEEE 802.11b/g [1], ZigBee [2] or Bluetooth [3] users. This is encouraging greater utilization of the wider 5 GHz band (up to 25 channels of 20 MHz) for newer WiFi standards, first with IEEE 802.11n and ultimately with the advent of IEEE 802.11ac [4] which only uses this band.

The trend towards utilization of the 5 GHz band in commodity wireless devices potentially impacts the functioning of pre-existing communication systems, e.g. meteorological radars [5]. Weather radars periodically emit unidirectional electromagnetic pulses and listen for echoes, e.g., reflected by raindrops. Fig. 1 illustrates the regular behavior of a radar that emits bursts of pulses separated by a time equal to  $T_{PR}$ . For each transmitted pulse, the radar processes the echoes for a measurement time  $T_{meas}$  and then remains idle for an Inter-Measurement Gap time  $T_{IMG}$  before the next pulse is transmitted. For example, Fig. 2 shows measured weather radar traces taken with a Rohde & Schwarz FSL-6 *spectrum analyzer* near to Dublin airport in Ireland. These measurements show bursts of electromagnetic pulses, with 3ms inter-pulse separation time. From Fig. 2(b) it can be seen that on a

larger time-scale the amplitude of transmissions is periodic in nature, with a period of approximately 20s. This is attributed to periodic rotation of the radar antenna, with high amplitudes corresponding to intervals when the radar antenna is directed towards the spectrum analyzer.

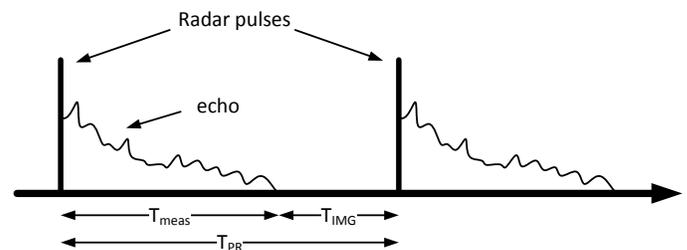
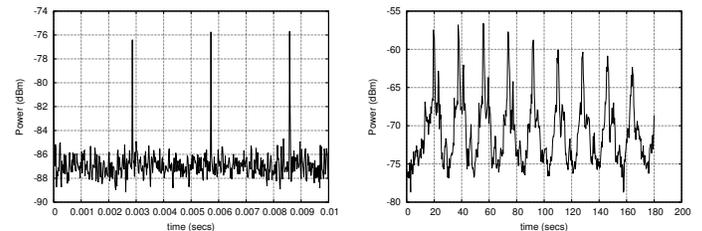


Figure 1: Schematic illustrating radar signals in the time domain.



(a) Small timescale.

(b) Large timescale.

Figure 2: Weather radar time history near Dublin airport, Ireland.

In order to facilitate coexistence, Dynamic Frequency Selection (DFS) is enforced for unlicensed devices, e.g., through FCC [6] regulations in North America and ETSI [7] regulations in Europe. To meet these regulatory requirements, the IEEE 802.11h [1] amendment introduces DFS to the IEEE 802.11 PHY/MAC standard. However, although the 802.11h specifies the mechanisms for supporting DFS, the ultimate responsibility for implementing efficient radar detection schemes lies with the device manufacturer.

DFS mandates that the master device (e.g., the Access Point in a WLAN) monitors the channel for potential radar interference for at least the *channel availability check time*. A major issue with this requirement is that no transmissions can occur during this check time, which may be up to 10 minutes duration for some channels [7]. A further issue is that, even after the check time has elapsed, when using half-duplex radios, radar detection cannot be carried out while data is being transmitted/received. Hence, radar detection in future dense (and heavily loaded) wireless networks is likely to be problematic. Moreover, weather radar technology keeps

changing, with more complex and faster scan patterns that inflate the ratio of false positives in legacy wireless devices, scaling up dramatically the amount of sensing time needed and so of channel unavailability.

In this paper, we present initial results of our experimental evaluation of radar detection through *in-service* bit-error pattern recognition. The ability to detect (and react to) the presence of radar signals while a transmission is ongoing not only offers the potential to increase the spectrum efficiency, but also to improve the reliability of radar detection in heavily loaded networks.

The rest of the paper is structured as follow. A brief survey of related literature is given in Section II, the description of the experimental methodology in Section III, and a set of representative results in Section IV. Finally, in Section V we discuss the open issues to resolve in future work, and Section VI concludes the paper.

## II. RELATED WORK

Most of the related literature focuses on the design of off-line detectors. For instance, in [8] the authors develop a detector based on Compressive Sensing (CS), while [9] designs a radar detector based on a Constant False Alarm Rate (CFAR) and a Complex Approximate Message Passing (CAMP) algorithm. [10] evaluates a few algorithms based on a matched filtering technique and propose a method based on power detection in the time domain. However, these detectors, if implemented in a regular wireless device with a half-duplex transceiver, could only work while communication is not ongoing, which limits its applicability as explained above.

The coexistence between IEEE 802.11 transceivers implementing DFS and radars has been studied in [5], [11]. These evaluate DFS in the presence of a Doppler weather radar system and show that the 802.11 radio introduces an additive and uncorrelated noise into the radar system.

A few works analyze the radar detection probabilities. Within the FCC and ETSI standards, the computation of the detection probability was carried out initially using a basic random transmission model [12]. The authors of [13] perform a theoretical evaluation of the detection of radar pulses in time division duplexed systems, and compute the expected number of pulses occurring during the receive period of a transmission.

In contrast to all previous works, this paper is, to the best of our knowledge, the first attempt to experimentally assess radar detection via bit-error recognition over received packets.

## III. METHODOLOGY

The key idea is to search for radar interference footprints, via inspection of bit error patterns, while packet processing is ongoing at the receiver. Fig. 3 illustrates this idea with a burst of eleven pulses (at the bottom) and three examples of data packet communication with different packet rates and lengths (at the top). Those packets, transmitted concurrently with one or more radar pulses, will be disrupted by a burst of bit errors (shown in black in the figure). The observation of this footprint, jointly with the knowledge of the underlying radar process, should make possible the detection of radar signals.

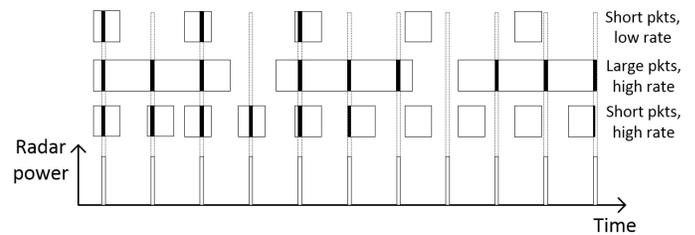
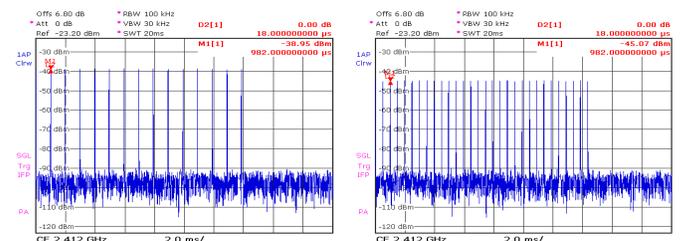


Figure 3: Radar interference footprint during packet reception.



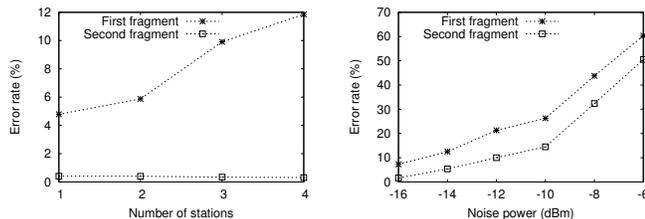
(a) Radar test signal type 2 [7] (b) Radar test signal with double speed

Figure 4: Measured traces while radar emulator is in operation.

The first step to study the feasibility of bit-error pattern recognition for radar detection requires a controlled environment. To this end, we programmed a Rohde&Schwarz SMBV100A *vector signal generator* to emulate the behavior of a radar. Unless otherwise stated, each pulse width is set to  $1 \mu s$ ,  $PRF = 1/T_{PR} = 1000 s^{-1}$  and we send bursts of 15 pulses, to mimic the behavior of a radar test signal type 2 [7]. Given the traces collected with our spectrum analyzer when our “radar emulator” is in operation, shown in Fig. 4, and their similarity with respect to those of Fig. 2, we validate this set-up for our experiments.

One of the main challenges in the detection of errors caused by the operation of radars is the ability to distinguish these from other sources of error, for instance those caused by another transmitter within the same WLAN, i.e., *collisions*. Collisions are part of the correct operation of the CSMA/CA MAC protocol in IEEE 802.11 and are caused by two or more stations selecting the same slot to transmit during the random backoff procedure. If a station is successful (i.e., its backoff counter reached zero before other senders), the other stations will defer their backoff while the channel is not *idle*. The physical layer (PHY) implements a Clear Channel Assessment (CCA) scheme based on a carrier-sense threshold for energy detection, and the MAC uses a Network Allocation Vector (NAV) parameter transported in each header to inform other stations about the duration of the ongoing transmission. A channel is idle only if the CCA fails to detect a carrier and if the NAV timer is zero. However, collisions are not the only cause of packet impairment. Nodes that are too far to be able to decode the NAV of concurrent transmissions whose energy level is below the CCA threshold (i.e., *hidden nodes*) could cause errors in theoretically successful transmissions as well. In order to narrow down the sources of interference to just those due to radar coexistence, we applied the technique proposed in [14]. To this end, we force each packet to be fragmented at the MAC layer to ensure that the second and subsequent fragments are protected from both collisions and

hidden nodes. The protection against collisions is granted by the NAV value used in the first fragment, which is set to the amount of time required to send all the fragments. Each fragment is sent back-to-back (separated by a SIFS interval) to the AP and individually acknowledged by the AP; given the fact that all stations should be able to hear the AP's transmissions, and therefore the NAV value carried within the acknowledgment frames, these fragments are protected against hidden nodes (similarly to the RTS/CTS scheme).



(a) Test link and other interfering links. (b) Test link and a source of noise.

Figure 5: Technique to narrow down sources of interferences.

We have run two simple experiments to validate this methodology; the results are shown in Fig. 5. In both experiments, we configure a laptop as a WiFi AP transmitting test traffic to another laptop, configured as a WiFi client. The AP sends 50 bursts of 3000 ICMP packets, with 1150 bytes of payload, generated with SING, an ICMP-packet generator which allows customization of the transmitted packets. The transmission power is set to 16dBm and we set a fragmentation threshold of 600 bytes to force the delivery of 2 fragments per packet (i.e., the first has to contend for channel access, the second is protected against collisions and hidden nodes). In the first experiment, the AP transmits the test traffic towards the client, while we increase the number of additional contending stations to the network. Fig. 5a depicts the ratio of bit errors for each of the pair of packets sent in each transmission opportunity. As we can see, the first fragment is severely affected by the growing number of collisions incurred by the additional contending stations. On the other hand, the second fragment is unaffected by this type of interference source. In

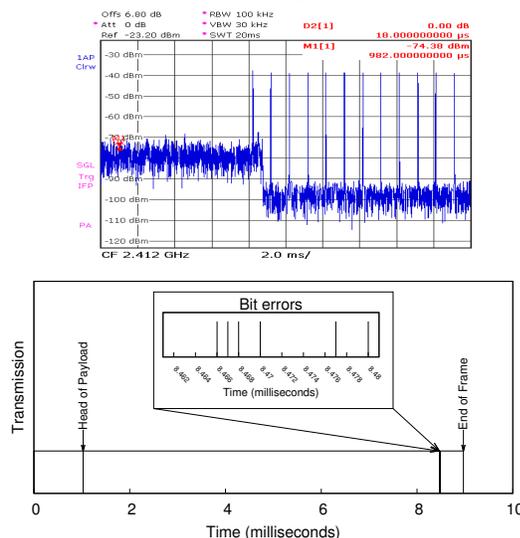


Figure 6: Radar traces and bit error patterns, PHY rate 1Mbps.

the second experiment we switch the additional contending stations for the signal generator, to generate synthetic RF noise. Fig. 5b shows the error rate as a function of the power of the noise generated. The results indicate that both the first and second fragment experience the same effects upon a noisy environment. These two experiments serve as validation for this technique to identify bit-error patterns due to radar signals while excluding those inherent to the IEEE 802.11 MAC protocol. In the following, we will apply this technique, along with our radar emulator, to study the bit-error patterns in received frames and its feasibility in the detection of radars in the environment.

#### IV. MEASUREMENTS

We set up two laptops with an IEEE 802.11b/g Atheros AR5008 wireless card each using the driver madwifi 0.9.6 and separated by 100cm. Note that we have disabled the IEEE 802.11h capabilities from the driver to avoid that the legacy DFS operation affecting channel selection upon the observation of our radar signals. One laptop serves as WiFi AP and the other as a regular client. Each experiment consists of the transmission of a stream of packets from the AP to the client. Each UDP packet transports a 1796-byte payload with known information, which allows us to assess the error patterns occurring at the receiver. In the following, we summarize our initial results where we compare the trace given by the spectrum analyzer and the error pattern found in selected packets.

In the first experiment, each packet is sent at 1 Mbps (so they are of relatively long duration). Fig. 6 top shows how the end of a packet transmission collides with one pulse from the beginning of a radar burst. In Fig. 6 bottom we illustrate the results of the packet inspection at the receiver. We mark the beginning of the payload (the part that we process), each bit error found after decoding (we also zoom in the burst of errors) and the end of the packet. This figure shows that the footprint of the colliding radar pulse is clearly observable. However, in this case, the observation of one burst of errors is not sufficient to infer the presence of a radar signal.

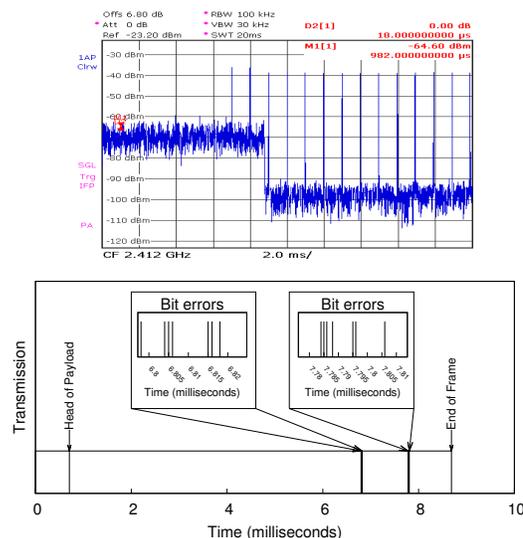


Figure 7: Radar traces and bit error patterns, PHY rate 1Mbps.

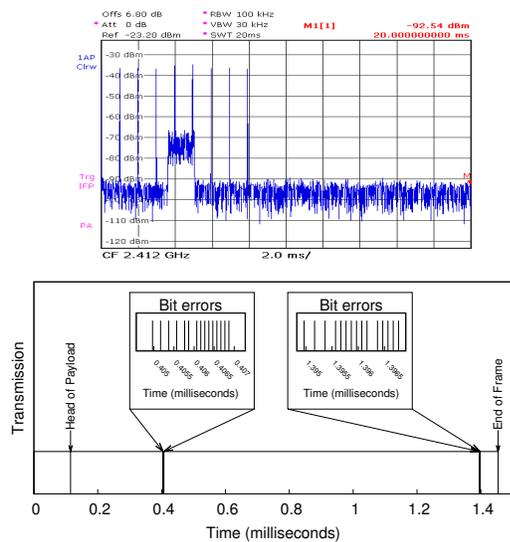


Figure 8: Radar traces and bit error patterns, PHY rate 6Mbps.

Fig. 7 shows the resulting packet processing of another packet sent at 1 Mbps where two pulses of a radar signal corrupts the payload of the data transmission. We can easily identify the radar footprint out of the packet inspection, i.e., two bursts of errors separated by  $1\text{ms}$  ( $PRF$  of our test radar signal) with pulses of approximately  $1\mu\text{s}$  duration.

We now configure the AP to transmit 6-Mbps packets. Similarly as before, Fig 8 depicts the results for a selected packet of this experiment which is corrupted by two radar pulses. This result allows a similar observation as in the previous case, that is, we found two burst of bit errors of approximately  $1\mu\text{s}$  of duration and separated by  $1\text{ms}$ , matching the characteristics of the underlying radar test signal.

## V. OPEN QUESTIONS

The experiments we have carried out so far unveil a clear footprint from radar interference which is observable while a packet reception is being processed. However, the heterogeneous (and random) nature of real-life data packet communications makes the actual design of a practical algorithm for radar detection through bit-error pattern recognition inherently challenging. The major challenge is indeed to discern the constant rate footprint of radar bursts from a decoupled data transmission, i.e., with independent rate/size distribution.

Fig. 3 illustrates this with three examples with different packet rates and packet lengths (although packet rate and length are constant in these examples, which will not be the case in general). First, a single large packet might be enough to observe a constant error burst rate that would trigger a radar presence alarm. However, the case with short packets requires a longer-term observation. Second, the analysis of consecutive packets from a high packet rate transmission could also be sufficient. However, slower rates (and/or heterogeneous distributions) might hide the presence of a radar given that most of its bursts could coincide with idle inter-frame spaces, not causing bit errors.

## VI. CONCLUSIONS

Dynamic Frequency Selection (DFS) is a mandatory scheme for wireless communication protocols operating in the 5 GHz band in order to mitigate the interference caused to coexistent systems such as radars. Unfortunately, conventional implementations may cause severe underutilization of resources and poor performance in heavy loaded networks. Motivated by this observation, we explore the feasibility of radar detection via duly observation of bit-error patterns in received packets. In this paper, we present the initial results of our experimental evaluation and discuss a series of open questions that need to be resolved to design an effective *in-service* detection mechanism.

## ACKNOWLEDGMENT

Work supported by Science Foundation Ireland grant 11/PI/11771.

## REFERENCES

- [1] IEEE Standard - Local and metropolitan area networks - Part 11: Wireless LAN Medium Access Control and Physical Layer specifications, IEEE Std. 802.11, 2012.
- [2] IEEE Standard - Local and metropolitan area networks - Part 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs), IEEE Std., 2013.
- [3] IEEE Standard - Local Area Networks - Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs), IEEE Std., 2005.
- [4] IEEE Std 802.11ac-2013 (Amendment to IEEE Std 802.11-2012, IEEE Std. 802.11, 2013.
- [5] A. L. Brandão, J. Sydor, W. Brett, J. Scott, P. Joe, and D. Hung, "5 GHz RLAN interference on active meteorological radars," in Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st, vol. 2. IEEE, June 2005, pp. 1328–1332.
- [6] FCC, "Federal Communications Commission Report and Order FCC 03 287 released November 18," 2003.
- [7] ETSI, "EN 301 893 v1.7.1: Broadband Radio Access Networks (BRAN); 5 GHz high performance RLAN; Harmonized EN covering essential requirements of article 3.2 of the R&TTE Directive," June 2012.
- [8] L. Anitori, M. Otten, and P. Hoogetboom, "Detection performance of compressive sensing applied to radar," in Radar Conference (RADAR), 2011 IEEE. IEEE, 2011, pp. 200–205.
- [9] L. Anitori, M. Otten, W. van Rossum, A. Maleki, and R. Baraniuk, "Compressive cfar radar detection," in Radar Conference (RADAR), 2012 IEEE. IEEE, 2012, pp. 0320–0325.
- [10] M. Wen and L. Hanwen, "Radar detection for 802.11 a systems in 5 ghz band," in Wireless Communications, Networking and Mobile Computing, 2005. Proceedings. 2005 International Conference on, vol. 1. IEEE, 2005, pp. 512–514.
- [11] J. Kruys, "'dfs compliance criteria, status and prospects'," in Oaktree Wireless, 2009.
- [12] ITU, "Recommendation ITU-R M.1652, International Telecommunications Union," 2003.
- [13] B. W. Zarikoff and D. J. Leith, "Analysis of radar detection probabilities in time division duplexed systems," in Communications (ICC), 2012 IEEE International Conference on. IEEE, 2012, pp. 1698–1702.
- [14] D. Giustiniano, D. Malone, D. J. Leith, and K. Papagiannaki, "Measuring transmission opportunities in 802.11 links," IEEE/ACM Transactions on Networking (TON), vol. 18, no. 5, 2010, pp. 1516–1529.

# Cross-Platform End-to-End Encryption of Contact Data for Mobile Platforms using the Example of Android

Markus Hofmarcher, Michael Strauß, and Wolfgang Narzt

Department of Business Informatics – Software Engineering

Johannes Kepler University Linz

markus.hofmarcher@gmail.com, michael@strauss.eu.com, wolfgang.narzt@jku.at

**Abstract**— Storing and synchronizing personalized or business-related data, especially contact data, is increasingly done by cloud services. Thereby, control over personal data is subject to the technical conditions and measures of the cloud service providers. However, abstaining from the utilization of cloud services is not an alternative as their amenities are indispensable both for private and business users. End-to-end encryption, as it is already maturely applied for various communication services, would enable users to still keep their contacts on remote storage nodes, but save them in encrypted form. Although, the principal concepts for this kind of security measure are well-studied, there is still no service for protecting cloud-based contact data by end-to-end encryption using mobile platforms. This paper presents the ideas and the architecture for end-to-end encryption of contact data using the example of Android.

**Keywords** – end-to-end encryption; cloud services; Android; contact data; cross-platform

## I. INTRODUCTION

The list of one's own contacts is the backbone for modern communication. Cloud services offer powerful interfaces for conveniently managing this data on remote servers, from where they can be synchronized to arbitrary clients and platforms. They have become indispensable for present communication systems, from which private as well as business users cannot escape.

However, utilizing cloud services is potentially unsafe and simultaneously means giving up control to the service providers due to varying international security laws and regulatory frameworks. Users are unaware of the location and the amount of physical server nodes where their data is stored on. As a consequence, legal positions and applicable laws are not transparent to the users, especially for the private user sector [1]. Deleting data can also be challenging, as the physical deletion is not executed in many cases even when users request it [1]. Moreover, recent press releases have pointed out the drawbacks of cloud services in respect to security and compliance [2][3]. Missing transparency and the lack of control and security mechanisms are mentioned as the major issues.

Nevertheless, cloud services provide a valuable method for managing and synchronizing data and are even obligatory nowadays for many business companies [4]. Hence, we propose a method for protecting cloud-based data by (well-proven) end-to-end encryption. In particular, we refer to contact data to be securely exchanged and synchronized with

mobile devices. The advantages of cloud services remain; only the data is stored encrypted. We present a generic cross-platform architecture, show an exemplary implementation for the Android platform, where users can select those apps, which are granted access to the decrypted contact data, and finally discuss re-utilization of the concept for other mobile platforms. Simplicity and usability are major aspects to be considered in our work, in order to raise acceptance among the users, whereby the main reasons people refrain from using encryption for email and file transfer are the lack of knowledge and the additional time and effort for installing and using security features [5].

The paper is organized as follows. Section II discusses related work and similar approaches. Section III introduces our architecture followed by its implementation in Section IV. Section V shows the results of our investigation. Sections VI and VII discuss possible solutions for other platforms and concepts for synchronization. Finally, in Section VIII we draw conclusions and formulate ideas for future work.

## II. RELATED WORK

Among the scientific investigations concerning privacy protection and information security in cloud systems, research is mainly focused on strategies and mechanisms offered by the cloud service provider [6][7][8] or is even restricted to the selection of trustworthy providers. Security issues concerning cloud services are thoroughly analyzed with respect to architecture, data delivery models and from the stakeholders' perspective [9].

In order to overcome the uncertainties of cloud services, Puttaswamy et al. [5] propose to insert a trustworthy (because self-managed) organizational node in between client and cloud in order to pre-encrypt data before transferring them to the cloud. The organizational node holds the keys and additionally identifies functionally encryptable data, i.e., data never interpreted by the cloud and not breaking application functionality. It automatically encrypts them and generates and provides appropriate access keys for its users. Thus, the authors' proposal represents an approach of minimizing the weak points of cloud services by an intercalated server instance offering end-to-end encryption. End-to-end encryption, in general, is the preferred method for storing data in cloud services [10][11].

Practically applied end-to-end encryption can be found, e.g., in PrivacyCrypt for Facebook [12], a special security feature for encrypting messages in the social network Facebook. The public key used for asymmetric encryption

has to be exchanged with the target users who are allowed to decrypt and read those messages. A Firefox extension contains the implementation for this security feature. Fig. 1 illustrates its principle scheme. The left part in Fig. 1 shows the message in plain language, which is encrypted by the Firefox plugin at the client side using the public key(s) of the receiver(s). The encrypted message is then transferred to and stored on the server ("Facebook cloud"). Facebook consequently receives encrypted messages, which can only be decrypted by those having the appropriate private key.

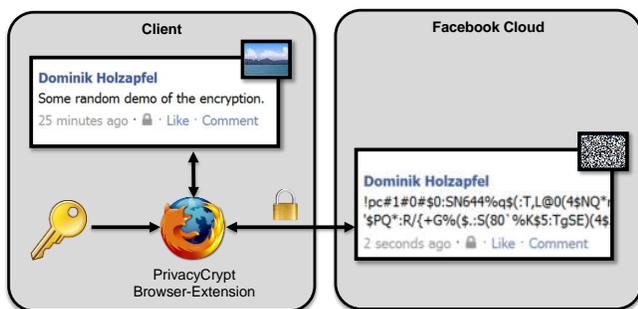


Figure 1. PrivacyCrypt for Facebook

Boxcryptor [13] uses a similar approach. This application supports users encrypting their data stored on cloud servers such as Dropbox, Google Drive or Microsoft SkyDrive. First, the files are encrypted locally on a user’s device and afterwards sent to and stored on a remote server. Boxcryptor is available for Microsoft Windows, Mac OSX, iOS and Android. Furthermore, there is an extension for Google Chrome which enables users to access and decrypt their data via a browser instance.

Fig. 2 briefly explains Boxcryptor’s support for two exemplary platforms (desktop with Google Chrome and Android) having a comparable working procedure to PrivacyCrypt as both approaches use end-to-end encryption.

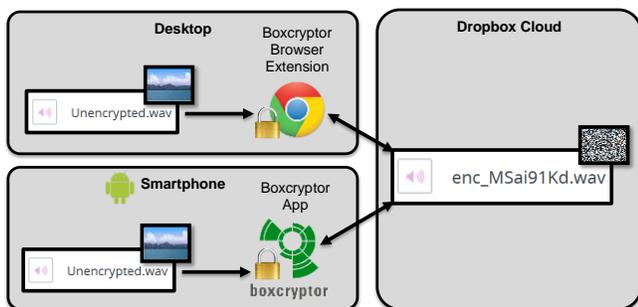


Figure 2. BoxCryptor

End-to-end encryption is also our proposed method to securely store (contact) data on cloud servers, allowing its users to keep control over their own data. Privacy and information security is therefore guaranteed by end-to-end encryption [11]. Even insufficient protection mechanisms of cloud service providers have no negative effects in respect to data security as the data can only be decrypted by the owners themselves as long as they use state-of-the-art keys [14].

There are numerous existing solutions for data encryption. For instance the first version of Pretty Good Privacy (PGP) was released in 1991 and is nowadays, beside S/MIME, an established standard of e-mail encryption. Although several implementations are available, PGP or S/MIME are not commonly used, despite the high need of e-mail encryption. A case study pointed out that it is hardly possible to set up or use PGP without specialized knowledge [15]. Lack of knowledge and inconvenience are the main reasons for not using encryption [16].

### III. APPROACH

The challenge in developing a suitable method for encryption is security and also usability. To achieve high acceptance it is necessary to design the application for encryption as transparent and as user friendly as possible. Furthermore, compatibility with third party applications must be ensured. In conclusion, we claim that there are three primary “quality” criteria which must be fulfilled by our solution:

- Comfort
- Transparency
- Security

Comfort means that the use of cloud services is still possible even when the data is stored in encrypted form. That must be possible without or rather with little migration effort. In addition to this, the user should be able to use encrypted data like unencrypted data. A password recovery concept is also important. Another aspect is multi-platform support. The user should not be limited, and therefore, the encryption solution should work on the most common platforms.

Third party applications must still be able to access contact data if this is allowed by the user. Hence, the encryption mechanisms must be implemented transparently so that there is no need to adapt these third party applications. They should be able to access data when needed in the same way even though the data is stored in encrypted and decrypted form. This is important for software components providing contact data synchronization.

Of course, data must be encrypted securely. That means an up-to-date algorithm for encryption must be used. Fig. 3 outlines the common data flow of contact data in combination with cloud services for synchronization.

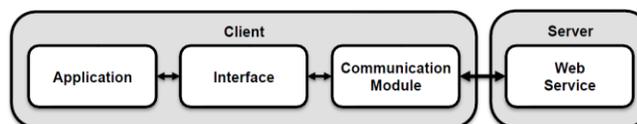


Figure 3. General contact data flow

Data is used and managed by the client application, which usually provides a GUI enabling the user to read or change contact data. These applications retrieve data from a (central) interface, where sometimes data persistence is also implemented. To synchronize data with cloud services a communication module is necessary.

Depending on the platform some of the described modules are possibly bundled in one component. However,

as a general rule data is processed according to the following steps:

- Representation and user access (application)
- Data management (interface and persistence layer)
- Server communication (through a communication module)

These steps are potential levels where en- or decryption can take place. According to the three mentioned quality criteria, it is inappropriate to implement encryption at the application level because that would require adaptations of every used application.

Fig. 4 shows the discussed data flow specifically for Android. There are three applications (People, Facebook and Email) that are able to access contact data. The central Android interface for accessing contacts is called “Contacts Provider”. All requests are handled by this interface, which also implements data persistence. “Sync Adapters” are responsible for synchronizing data with cloud services and resolving conflicts.

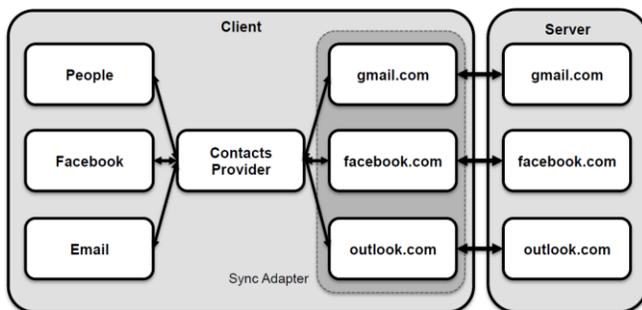


Figure 4. Android contact data flow

Further, it is also improper to implement encryption at the level of the communication module. This would also require adapting every used Sync Adapter in the case of Android. Therefore, we chose the central access interface (Contacts Provider) for our proof-of-concept to implement encryption. This ensures that all quality criteria are fulfilled, which is further discussed in Section V.

For this implementation to work seamlessly the internal data structure of the contacts has to remain unchanged and only the value of the individual fields can be encrypted so applications and backend services can still process the data. The central data interface (Contacts Provider) provides three methods with the following functionalities:

- “query”: Retrieve one or more contacts
- “update”: Modify or delete an existing contact
- “insert”: Create a new contact

A hook on every one of the three mentioned methods is an easy and effective way to intercept all requests according to contact data operations.

#### IV. IMPLEMENTATION

Before implementing the proposed architecture for Android, it was necessary to determine if it is possible to set hooks on method calls of other processes and applications. Our research shows that there is no official interface for this purpose due to Android’s security concepts, but as Android

is an open source platform, this functionality has been added by external developers. One such implementation is provided by the “Xposed Framework” which we used to implement a module according to the proposed architecture.

Installing this framework requires so-called “root access” as it needs privileged access rights. The process of attaining this privileged access varies from device to device and can lead to a partial or complete loss of warranty. We are aware that this can pose a significant obstacle, especially in regards to ease of use, but at the time of writing we are not aware of any alternative. This is a direct result of Android’s security concepts not providing access control mechanisms for third party applications. Conventional server and desktop platforms provide various ways of managing permissions, while on Android the user typically has no way of gaining privileged access rights. The idea behind this limited privileges concept is to prevent applications from damaging the system. This concept, however, neglects the need for privileged access for security applications like anti-virus software or firewalls.

Android applications are executed in a virtual machine, called Dalvik, which is a specialized implementation of the Java VM. The Xposed Framework loads additional classes to every instance of a Dalvik-VM allowing the implementation of hooks for every method call inside such an instance. A hook is a method that is called before or after a specified method and allows modifications to a method’s parameters or its return value.

The prototype that was developed to prove the viability of the proposed architecture was implemented as a module for the Xposed Framework. This module defines hooks on the query, insert and update methods to Android’s internal Contacts Manager to implement encryption and decryption of all contacts handled by this central manager.

Fig. 5 shows a schematic representation of a query method call to obtain contact details from the Contacts Provider. The highlighted modules have been implemented by us.

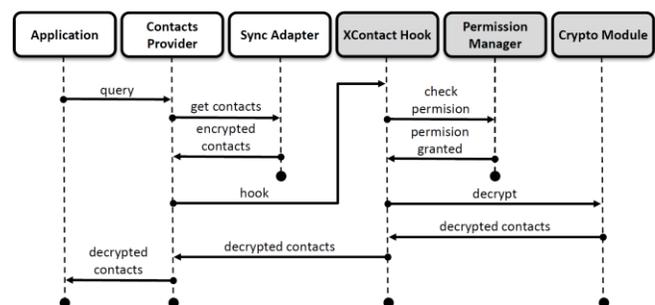


Figure 5. Sequence of a query action

##### A. XContact Hook

The module “XContact Hook” provides the necessary hooks on the three methods specified before to intercept calls of all applications to the Contacts Provider. It defines hooks that are executed before calls to the insert and update methods of the Contacts Provider. The new or updated contact data is encrypted in these hooks and then passed to

the insert or update methods. It also defines a hook that is executed after the query method collected the contacts specified by its parameters. These are then decrypted and passed to the calling methods if the application has the permissions to get the decrypted contacts. In both cases, the data structure of the contact data is kept intact, and internal ID fields are not encrypted to keep compatibility with applications and sync adapters.

### B. Permission Manager

This module provides an interface for managing the permission on an application level. It enables the user to enable or disable the cryptographic module for specific applications, therefore granting only trusted applications access to unencrypted contact data.

### C. Crypto Module

The cryptographic module provides routines for encrypting and decrypting contact data while maintaining interface compatibility by preserving the structure of the encrypted data. For encryption the Advanced Encryption Standard (AES) [17] algorithm is used as it is implemented on Android and at the time of writing is sufficiently secure [18]. Initially, a key length of 256 bit was used but with a large number of contacts the additional processing required for encryption and decryption was perceptible by the user.

Since only field data is encrypted, a new random initialization vector is used for each field to mitigate frequency analysis and known plaintext attacks.

## V. RESULTS

The major result of our investigations is an architecture concept for end-to-end encryption and a proof-of-concept for Android. Furthermore, we analyzed the performance impact of encryption key lengths and discussed how the proposed architecture could be applied on other platforms. The defined three quality criteria are fulfilled by the proof-of-concept.

**Comfort:** Although the contact data is encrypted the user is able to use it as usual. The only difference is that users are able to manage permissions with the “XContact Permission Manager” (see Fig. 6). We also discussed potential ways for recovering the password. Furthermore, the user is not forced to use a specific platform because the concept is multi-platform capable.

**Transparency:** All cloud services for remote data storage can be used with our proof-of-concept because the structure of the data is not changed. Only the data fields are encrypted. Also third party applications (apps) are able to access contact data as usual. There are no specific changes for them.

**Security:** AES, which was published by the National Institute of Standards and Technology as a standard for data encryption in 2000, is a secure encryption algorithm. There are no known relevant attacks and even a successful brute-force attack is hardly possible [19]. The private key for encryption is only stored on the user’s device. So, the cloud service provider is unable to decrypt the users’ data. In case of any errors, the calling action is not executed. So, there is no chance of an unencrypted data leak.

### A. Proof-of-concept (prototype)

In this section, we present our proof-of-concept for end-to-end contact data encryption on Android. The main components of “XContact” are:

- Xposed Hook
- XContact Permission Manager

Xposed Hook works completely in background and is used to hook every relevant method call on the Android Contacts Provider. XContact Permission Manager is an app that allows the user to control which applications are able to access unencrypted contact data.

For installation it is necessary to install the Xposed Framework. This is done in four steps, and, as already mentioned, requires root-permissions. After that our prototype can be installed with an APK file. To use the described features it is necessary to active the XContact module (see Fig. 6). So, users can select those applications they would like to grant access to unencrypted contact data.

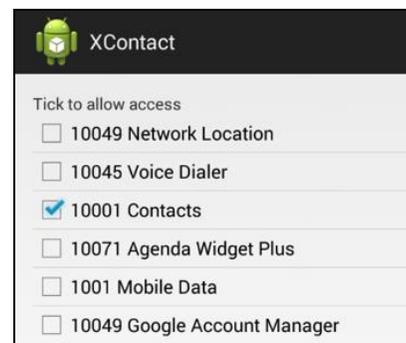


Figure 6. XContact Permission Manager GUI

The XContact Permission Manager shows all installed applications that have requested permission to access contact data at installation time. If the checkbox is not ticked, the application will receive an empty list when it calls the “query” method, and new contacts will not be encrypted. Since XContact encrypts data transparently, the procedures to retrieve, modify or delete contact data are not affected. Therefore, encrypted data can be automatically synchronized with cloud services like Google Contacts.

### B. Performance impact measurements

In this section we discuss the performance impact of encryption. Our test device is a Nexus 7 (2013) tablet running Android 4.3.1 and Xposed 2.4.1. For our measurements the time at the beginning and the end of an en- or decryption hook have been monitored. The difference of these values is the additional amount of time consumed by the encryption algorithms.

For every test we used the same contact with three fields (name, phone number and e-mail). Fig. 7 describes the arithmetic mean of 30 measurements.

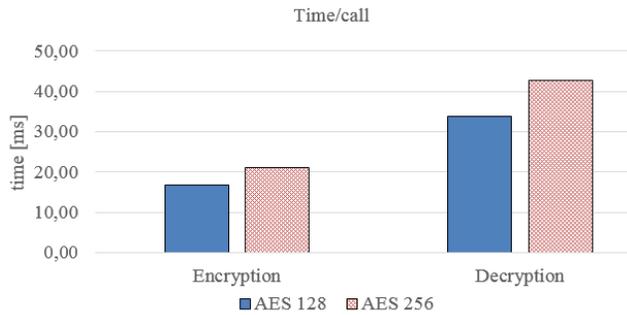


Figure 7. Performance Measurements

The comparison of AES-128 and AES-256 shows that the latter encryption algorithm is about 25% slower. AES-128 is state-of-the-art and generally considered as secure [19]. A key length of 128 bit for symmetric encryption is recommended for long-term data protection [18].

The general performance impact depends on many different factors (e.g., used CPU, other running processes, number of different address books, number of contact groups and single contacts). With our test device and one contacts application we were not able to visually perceive or recognize any delays for different use cases.

## VI. PASSWORD AND SYNCHRONIZATION

The nature of the proposed architecture requires a different approach to certain basic features. These features are typically implemented server-side, however, our approach is located only on the client. It, therefore, is essential to minimize the risk for the users losing access to their data due to a forgotten password. Typically, services provide mechanisms for resetting passwords or asking secret questions, but this would require storing the users' password on the device or a server-side implementation.

An alternative method for solving the password problem is to securely store passwords with a password manager, so the user only has to remember one password to access all other passwords. The passwords for multiple services are stored in an encrypted container that can be backed up using cloud services or other backup solutions as it follows the same end-to-end encryption principles as our approach.

While this method cannot strictly be labeled, we believe that password recovery offers the best trade-off between comfort and security. Furthermore, it can be implemented without the need for a server or service provider and, therefore, is best suited for the proposed architecture.

A stated goal of this paper is to utilize data encryption as user-friendly as possible. The platform independent aspect suggests that users may use the application on multiple devices and, therefore, have to enter the password more than once. Also, in the event of changing the password, the new password has to be propagated to all devices using the encryption module requiring a synchronization mechanism. The client-side implementation requires alternatives to usual cloud-based synchronization methods.

We propose the use of a peer-to-peer approach for synchronizing the password across multiple devices and

platforms. A discovery protocol identifies relevant nodes on a local network, while a specification of the transmission protocol is needed for transmitting the data.

Based on such a protocol it is possible to synchronize passwords across multiple devices on a local network. If the user installs the encryption module on a new device, it can detect already configured devices on the same network and request the password from them. The user is notified about this request and can either permit or deny it. It is also possible to propagate password changes to all relevant devices on a local network.

## VII. CONCEPTS FOR ALTERNATIVE PLATFORMS

Enabling transparent end-to-end encryption for platforms other than Android-based operating systems means implementing an application for each targeted platform. While the proposed architecture was designed with minimal implementation overhead in mind, the nature of end-to-end encryption makes this impossible to avoid.

Platforms can be divided into two main categories with different requirements for implementation.

### A. Desktop Platforms

A variety of operating systems and applications exist for desktop computers and workstations that require contact data. Therefore, we focus our research on web-browsers as they are available on multiple platforms and are a common way to access and manage contact data, largely because service providers often provide web applications for doing so. The authors of [12] show that there are no technical restrictions for implementing the cryptographic module as an extension for common web-browsers.

Contact data is an essential part of e-mail applications. At the time of writing the most commonly used e-mail application on desktop platforms is Microsoft Outlook with a market share of 20.14% [20]. This application can be extended with so-called Add-Ins similar to browser extensions, thus facilitating the implementation of an XContact extension for this application.

### B. Mobile Platforms

The mobile operating system iOS by Apple is similar to Android as it has a UNIX core and therefore similar (unofficial) methods to acquire privileged access. A framework similar to the Xposed Framework on Android is called "Cydia Substrate" and provides the APIs for defining hooks on methods of other applications. Installing Cydia Substrate on an Apple iOS device requires superuser privileges that can be acquired by "jailbreaking" the device. On Apple iOS contacts are managed by a central address book, providing the basic prerequisites for implementing the XContact module for iOS.

For "Windows Phone" there is no official way of acquiring privileged access and at the time of writing also no unofficial way similar to iOS or Android. It is possible to register a Windows Phone device as a developer device but this is not feasible for end-users and the acquired privileges are insufficient for implementing applications with the proposed functionality.

## VIII. CONCLUSION

Privacy and information security in respect to distributed data management and mobile data availability are on the rise and are not solely discussed in scientific communities but have also drawn public interest.

One effective measure for protecting user data is end-to-end encryption, a well-established security paradigm already applied to a series of scenarios concerning remote data management, but especially for exchanging sensitive data between two or more participants. The major drawback of this technique is the administrative effort for creating and securely exchanging keys, thus preventing it from being widely used, e.g., for secure mail transfer.

In terms of synchronizing one's own contact data through cloud services, exchanging keys is irrelevant, for one user is administering one key to be used for various clients. A prototype implementation for end-to-end encryption of contact data for the Android platform has proven the applicability of such a security feature for mobile devices with minimal impact in terms of convenient, transparent and seamless usage. However, (still) off-the-record techniques had to be applied (the device had to be rooted) in order to practically implement this encryption model.

Nevertheless, we are convinced that developers will be able to access critical system functions of mobile platforms needed for special security operations in the near future, as there are upcoming issues concerning malware that need to be counteracted [21]. Appropriate third party partner programs, for example, might be a way to enable only licensed vendors in developing security-related software.

Independent from this vision, security issues for mobile platforms will increasingly become a topic of research [22]. The security concept presented in this paper is an attempt to protect personal contact data stored on cloud servers, the architecture of which is adequately applicable for encrypting calendar dates or private photos, and all this for various platforms.

## REFERENCES

- [1] V. Tchifilionova, "Security and Privacy Implications of Cloud Computing – Lost in the Cloud," in *Open Research Problems in Network Security SE - 14*, vol. 6555, J. Camenisch, V. Kisimov, and M. Dubovitskaya, Eds. Springer Berlin Heidelberg, 2011, pp. 149–158.
- [2] A. Greenberg, "Cloud Computing's Stormy Side - Forbes," *Forbes*, 2008. [Online]. Available: [http://www.forbes.com/2008/02/17/web-application-cloud-tech-intel-cx\\_ag\\_0219cloud.html](http://www.forbes.com/2008/02/17/web-application-cloud-tech-intel-cx_ag_0219cloud.html). [Accessed: 09-May-2014].
- [3] V. Strauss, "Student privacy concerns grow over 'data in a cloud,'" *The Washington Post*, 2014. [Online]. Available: <http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/01/03/student-privacy-concerns-grow-over-data-in-a-cloud/>. [Accessed: 09-May-2014].
- [4] ISACA, "Cloud Computing: Business Benefits With Security, Governance and Assurance Perspectives," 2009.
- [5] K. P. N. Puttaswamy, C. Kruegel, and B. Y. Zhao, "Silverline: Toward Data Confidentiality in Storage-intensive Cloud Applications," in *Proceedings of the 2Nd ACM Symposium on Cloud Computing*, 2011, pp. 10:1–10:13.
- [6] V. P. Lijo and S. Kalady, "Cloud Computing Privacy Issues and User-Centric Solution," in *Computer Information Systems – Analysis and Technologies*, vol. 245, N. Chaki and A. Cortesi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 449–451.
- [7] J. C. Muñoz, G. Tamura, N. M. Villegas, and H. A. Müller, "Surprise: User-controlled Granular Privacy and Security for Personal Data in SmarterContext," in *Proceedings of the 2012 Conference of the Center for Advanced Studies on Collaborative Research*, 2012, pp. 131–145.
- [8] S. Ruj, M. Stojmenovic, and A. Nayak, "Privacy Preserving Access Control with Authentication for Securing Data in Clouds," in *Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (Ccgird 2012)*, 2012, pp. 556–563.
- [9] A. Behl and K. Behl, "An analysis of cloud computing security issues," in *Information and Communication Technologies (WICT), 2012 World Congress on*, 2012, pp. 109–114.
- [10] S. Pearson, M. C. Mont, L. Chen, and A. Reed, "End-to-End Policy-Based Encryption and Management of Data in the Cloud," in *Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on*, 2011, pp. 764–771.
- [11] R. Karim, C. Ding, and A. Miri, "An End-to-End QoS Mapping Approach for Cloud Service Selection," in *Services (SERVICES), 2013 IEEE Ninth World Congress on*, 2013, pp. 341–348.
- [12] R. Koch, D. Holzapfel, and G. Dreo Rodosek, "Data control in social networks," *2011 5th Int. Conf. Netw. Syst. Secur.*, pp. 274–279, Sep. 2011.
- [13] Secomba GmbH, "Technical overview - How Boxcryptor works," 2014. [Online]. Available: <https://www.boxcryptor.com/en/technical-overview>. [Accessed: 09-May-2014].
- [14] J. Xu, E.-C. Chang, and J. Zhou, "Weak Leakage-Resilient Client-side Deduplication of Encrypted Data in Cloud Storage Categories and Subject Descriptors," in *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security - ASIA CCS '13*, 2013, no. 2, pp. 195–206.
- [15] A. Kapadia, "A Case (Study) For Usability in Secure Email Communication," *IEEE Secur. Priv. Mag.*, vol. 5, no. 2, pp. 80–84, Mar. 2007.
- [16] BITKOM, "Mehr Sicherheit durch Verschlüsselung," *BITKOM Presseinfo Verschlüsselung 19 12 2013 (German)*, 2013. [Online]. Available: <http://www.bitkom.org/files/dohttp://www.bitkom.org/files/documents/Verschlsselung.jpg>. [Accessed: 09-May-2014].
- [17] H. C. A. van Tilborg and S. Jajodia, *Encyclopedia of Cryptography and Security*. Boston, MA: Springer US, 2011.
- [18] N. P. (BRIS) Smart, "ECRYPT II Yearly Report on Algorithms and Keysizes (2011-2012)," 2012.
- [19] A. Biryukov and J. Großschädl, "Cryptanalysis of the Full AES Using GPU-Like Special-Purpose Hardware," *Fundam. Informaticae*, vol. 114, no. 3–4, pp. 221–237, Aug. 2012.
- [20] Campaign Monitor, "Email Client Popularity." [Online]. Available: <http://www.campaignmonitor.com/resources/will-it-work/email-clients/>. [Accessed: 09-May-2014].
- [21] A. Gupta, S. Dutta, and V. Mangla, "Malware Attacks on Smartphones and Their Classification Based Detection," in *Contemporary Computing*, S. Aluru, S. Bandyopadhyay, U. V. Catalyurek, D. P. Dubhashi, P. H. Jones, M. Parashar, and B. Schmidt, Eds. Springer Berlin Heidelberg, 2011, pp. 242–253.
- [22] M. La Polla, F. Martinelli, and D. Sgandurra, "A Survey on Security for Mobile Devices," *Commun. Surv. Tutorials, IEEE*, vol. 15, no. 1, pp. 446–471, 2013.

# Problematic of QoS-based On-demand Routing Protocols to Improve Communication in Mobility Context

Tiguiane Yélémou  
Polytechnic University of Bobo Dioulasso,  
Bobo Dioulasso, Burkina Faso  
Email: tyelemou@gmail.com

Toundé Mesmin Dandjinou  
Polytechnic University of Bobo Dioulasso,  
Bobo Dioulasso, Burkina Faso  
Email: dandjimes@yahoo.fr

**Abstract**—In mobile ad hoc networks, achieving good QoS is a critical issue and is very difficult to guarantee due to rapidly changing of network topology and the lossy nature of wireless links. In this context, routing protocols must be smart enough to select better paths for data transmissions. In this paper, we focus on the well-known Ad-hoc On-demand Distance Vector (AODV) protocol. Standard route discovery approach used in AODV is expected to obtain the best path in term of delay. However, in lossy-links context, multimedia data packet transmission success, on path established thanks to control packets, may require several attempts. These retransmissions increase delay and overhead. Many QoS-based methods failed to make a meaningful improvement due to added complexity and additional delay and overhead. In this paper, we use a metric based on number of Packet Retransmissions to show that improving performance of on-demand routing protocols, in the mobility context, lies on effective control of node neighborhood.

**Keywords**—mobility; reliability; wireless networks; quality of service; on-demand routing.

## I. INTRODUCTION

Mobile Ad hoc NETWORKS (MANET) are characterized by instability of their topology. This is mainly due to node mobility and the lossy nature of wireless links. Selecting reliable paths for data transmission in this context is a challenge. In order to guarantee Quality of Service (QoS), routing protocols should be smart enough to choose a reliable route in order to avoid packet loss. To deal with the problem, QoS-based routing protocols are proposed. Route selection process should take into account link quality. However, most methods proposed for link quality estimation and best path selection are not appropriate for this rapid topology change. They require a long period to find QoS path and the obtained path is, very often, longer than the shortest path (in terms of number of hops). Long paths are more vulnerable to breakage than shortest paths.

The main contributions of this paper include:

- Use of a convenient and practical way to evaluate quality of links in mobile context,
- Design of QoS-based Ad-hoc On-demand Distance Vector (AODV) protocol [1]. The QoS metric used

(called PR-metric) is based on the number of retransmissions. It takes into account accurately the proportion of retransmission time with respect to time of first issue. We use this metric to compare effectiveness of different QoS-based methods used to improve on-demand routing protocols performance,

- A detailed analysis of different QoS-based AODV protocol performance. For our tests, we used realistic wave propagation model and realistic mobility model.

The remainder of the paper is organized as follows: In Section II, we present and analyze related work. In Section III, we present our QoS-based routing protocols. Performance evaluation and discussions are made in Section IV. We conclude in Section V.

## II. RELATED WORK

In recent years, much effort has been made to improve the standard AODV protocol [1]. In this section, after presenting the critical behaviors of the protocol, we review various proposed improvements.

### A. AODV protocol

On-demand routing approaches are source-initiated reactive mechanisms. When a node desires to send a packet to an other node and does not have a valid route, it initiates a path discovery process in order to locate the destination node [1]. Then, a route request (RREQ) packet is issued and flooded in the network. Once the first RREQ packet reaches the destination node or an intermediate node with a fresh route toward the destination, a route reply (RREP) packet is sent back to the source node. The source node rebroadcasts the RREQ if it does not receive a RREP during a Route Reply Wait Time (RREP\_WAIT\_TIME). It tries discovery of path up to a given maximum number of attempts and aborts the session if it fails. As the RREP packet is routed back along the reverse path, the intermediate nodes along the path record a tuple for the destination in their routing tables which point to the node from which the RREP is received. This tuple indicates the active forward route.

AODV uses a timer-based technique to remove stale routes promptly. Each routing entry is associated with a route expiration timeout. This timer is refreshed whenever a route is used. Periodically, newly expired routes are invalidated.

Route maintenance is done using route error (RERR) packets. When a link breakage is detected, routes to destinations that become unreachable are invalidated. RERR propagation mechanism ensures that all sources using the failed link receive the RERR packet. RERR packet is also generated when a node is unable to forward a data packet for route unavailability.

The *first RREQ* consideration approach means the selected path is the one with the better Round Trip Time (RTT) and the shortest path in term of hops count if all links are considered as similar. Contrary to proactive routing approaches, in on-demand routing methods, nodes maintain information only for active routes. But, route request and route error broadcasted may be important if established routes are much bits error-prone. This can be demonstrated by simulation with the use of a realistic physical layer and a realistic wave propagation model.

### B. Enhanced AODV

One of the well-known problem of AODV protocol is the long end-to-end delay due to overtime induced by route discovery process. Also, when the frequency of link failures is high, routing load and jitter become important. Since the publication of standardized version of AODV, many efforts have been made to improve it. The major challenge is to limit the frequency of route discovery process. Thus, several optimizations have been proposed in the literature. Among them, we note taking into account link quality in the route selection process and adapting timers to the network dynamics.

1) *Tacking into account link quality*: To take into account link quality in the route selection process, several methods are proposed with different QoS metrics including bandwidth, delay, packet delivery ratio, Bit Error Rate (BER).

Khaled et al. [2] propose a path robustness-based quality of service routing for MANET. They proposed that before processing RREQ packet, an intermediate node must assure that its lifetime and the delay toward the neighbor from which it receives the RREQ packet are above given delay-threshold and lifetime-threshold. At each hop, at least five checks are made and RREQ packet size increased with a node address. Destination node and source node must wait for copies (that have followed different paths) of RREQ and RREP packets until a timeout. The overhead (additionnal delay and routing load) and the complexity of this approach hypothecate protocol effectiveness.

Some works, such as [3], use optimal link metric value in the path choice. Path selection choice based on optimal link metric value may not allow to get the best path. For example, for number of hops or retransmissions count-based metric, a path with minimum link metric value  $m$  (the minimal metric value among other feasible paths), is preferred than anyone with just one link with metric value upper than  $m$  even if the other links are better.

Some authors use additive and multiplicative metric to enhance AODV route discovery process. To find the optimal path in wireless mesh networks, Kim et al. [4] modify the standard AODV RREQ process. They propose that duplicate RREQs with better cumulative link metric value be forwarded, so that all the possible routes are considered. As link quality

metric, they use an improved Expected Transmission Time (ETT) [5]. Their RREQ packet carries the cumulative link ETT value. They estimate the achievable throughput of their approach more than twice compared to standard AODV. We presume it is not necessary to re-broadcast duplicate RREQ packets. The intermediate node may note all possible reverse paths and retain as active reverse path to the source the better one according to the considered QoS metric. Their approach needs to be tested in MANET context with realistic simulation assumptions.

2) *Taking into account network dynamics*: Mobility of nodes is one of the essential issue of MANET. Taking into account the mobility of nodes is countered, first, by difficulties to adequately measure the mobility degree of a node. Many papers [6][7] propose to privilege nodes with low speed but network topology change is not local problem. A node may be fixed but if its neighborhood moves a lot, integrating this node into transmission path will not allow efficient communication.

Some authors propose to use link breakage prediction for packet loss avoidance. In fact, when intermediate node detects degradation of neighbor link quality on active route, it may anticipate route maintenance process. Then, source node is advertized to the probable path failure and anticipates route recovery process. This avoids transmission interruption. QoS metrics used in this method include received signal strength [8], packet delivery ratio of control packets [9]. Very often, the power of modeled signal depends only on the distance to the concerned neighbor node. It is known that obstacles in wave propagation environment has an impact on signal strength [10][11]. Even if these metrics are accurately measured, the approach only anticipate the break of the link. The source must initiate a new route recovery process. The impact on delay improvement is not significant.

Amruta et al. [8] and Naif et al. [12] focused on accessibility prediction to restrict route discovery for future communications. Indeed, during the usual routing operations, a node can collect significant information enabling it to predict the accessibility and the relative mobility of the other nodes in the network. However, due to rapid change of network topology and since they are not actively maintained, these routes become obsolete.

## III. QoS-BASED ON-DEMAND ROUTING PROTOCOLS

In this section, we present the PR-metric and three variants of AODV based on this metric. However, comprehensive presentation of this metric is beyond the scope of this paper.

### A. QoS metric

To quantify link quality, we focus on metrics based on link reliability. Very often, criteria like as BER, Packet Delivery Ratio (PDR, e.g., Expected Number of Transmissions, ETX) are used. For this study, we use a new metric based on the expected number of retransmissions required to communicate successful data packet on this link. Let us call it PR-metric. With PR-metric, distance between a node and its neighbor will not be 1 but  $1 + a * (n - 1)$ , where  $n$  represents the average number of transmissions required to make a data transmission successful and  $a$  is a parameter to weigh retransmission cost.

For retransmission, we want to design a transmission made after the first issue (after the first transmission attempt). The coefficient  $a$  is the ratio between the average time required for a retransmission over the time necessary for an initial successful transmission. Statistical analysis and results permit us to estimate  $a$  to 0.65 with 0.03 as standard deviation.

The number of retransmissions can be obtained from network interface statistics (MAC level). This metric has a direct impact on delay and throughput. Contrary to the well-known metrics like BER or Expected Transmission Count (ETX) [13], it takes into account real time network load. Its estimation is local. It does not induce a significant routing load or a large computation time. It is a good compromise between the number of hops criterion and the BER or ETX criterion which induces selection of long route [14].

### B. AODV-BL-PR

AODV-BL-PR picks out AODV where we apply black-listing approach to route recovery process. With AODV-BL-PR, when an intermediate node receives a RREQ packet, it compares the PR-metric value of link on which this packet is received to a predetermined threshold. If this PR-metric value is higher than this threshold, the packet is discarded, otherwise it is managed as in standard AODV. We set this threshold to 2. We estimate that, in mobility context, after 2 attempts to transmit data, the path used is no longer valid. Note that maximum number of retransmissions at MAC layer is 4 for our test. We note that a control message (usually lighter) can be successfully transmitted on a poor quality link when a normal payload message can not be transmitted. With this route selection approach, paths containing bad links are disregarded. This will also limit the dissemination of RREQ messages and then reduces routing overhead.

### C. AODV-sum-PR

To design AODV-sum-PR, two main modifications are made to standard AODV, namely QoS-information dissemination and duplicate RREQ packets process by intermediate node.

- QoS-information dissemination: for AODV-sum-PR, RREQ and RREP packets are extended with the cumulative PR-metric (C-PR-metric) field. Source node initializes this metric to 0.0. An intermediate node increases the value of C-PR-metric by the PR-metric of the link on which it received the packet. The intermediate node also integrates reverse path into its routing tables. Each entry is improved with the C-PR-metric as QoS-metric. The RREP packet also carries the C-PR-metric. The field is, this time, initialized to 0.0 by the destination node or to the current value of entry related to this destination by intermediate node which initiates the RREP packet.
- Duplicate RREQ packet process: contrary to standard AODV, an intermediate node manages duplicate RREQ packet. Indeed, if the C-PR-metric of a duplicated RREQ packet is lower than the recorded one, the entry for source node (reverse path) is updated: the previous hop to the source node will be the new

transmitter. Finally, the source node obtains a path to the destination with the lowest C-PR-metric value.

Note that intermediate node does not need to re-broadcast the duplicate RREQ packet and does not need to integrate the PR-metric value of all its neighbors as control packets header information, as widely done.

In Table I, we summarized the duplicate packet processing.

TABLE I. SAMPLE OF DUPLICATED PACKET PROCESSING ALGORITHM

for the concerned reverse path
if new C-PR-metric < current C-PR-metric
update next-hop
update C-PR-metric
else
drop the packet

### D. AODV-Timer

In AODV-Timer, we reduce the timers associated to the various recorded routes, established links with neighbors and waiting for a response (hello timer, route validity timer, waiting RREP packet timer, etc.). These timers are used to manage routes and links validation or recovery processes. The new parameters are presented in Table II. This coordinated reduction globally means that a node more frequently inventories its links and routes.

With this approach, we want to know the determining factor between taking into account link quality or a convenient control of neighborhood information for better performance in mobility context.

TABLE II. DEFAULT (AT LEFT) AND MODIFIED (AT RIGHT) AODV PARAMETERS FOR OUR TESTS

Timer Parameter	AODV-st	AODV-Timer
MY_ROUTE_TIMEOUT	10s	5s
ACTIVE_ROUTE_TIMEOUT	10s	5s
REV_ROUTE_LIFE	6s	3s
BCAST_ID_SAVE	6s	3s
MAX_RREQ_TIMEOUT	10s	5s
NETWORK_DIAMETER	30 hops	10hops
RREP_WAIT_TIME	1.0s	0.7s
HELLO_INTERVAL	1s	0.5s
BAD_LINK_LIFETIME	3s	1.5s

In summary, the reduction of route timeout value to 5s means that a path that is not used 5s ago is considered obsolete. The default value in standard AODV is 10s. The source waits less time (0.7 instead of 1.0) to restart a new request if it receives no response to a previous query. The network diameter is reduced to 10 instead of 30. We estimate that over 10 hops it is impossible to communicate in node mobility context. A HELLO\_INTERVAL timer set to 0.5s instead of 1.0s, means that nodes should test their neighborhood more frequently.

## IV. PERFORMANCE EVALUATION

In this section, we first present our simulation environment, we then present the results of simulation tests and analyze the performance of different protocols.

A. Experimental setup

To compute more real simulations, we use a realistic wave propagation model taking into account environment characteristics. Therefore, we enhanced NS2 [15] with a ray-tracer simulator, Communication Ray Tracer (CRT) [16], that has been developed at the XLIM-SIC laboratory. CRT simulator provides a 3D ray-tracer wave propagation model. It takes into account the geographical data, electrical properties of materials, the polarization of the antennas, the position of the transmitters and receivers, the carrier frequency and the maximum number of interactions with the surrounding obstacles.

To realistically model the movement of the node, we use the VANET-Mobisim [17] software. Node speed is computed by this software. The mobility model implemented is more realistic than widely used ones [18][19][20]. Paths are defined in correlation and consistency with our environment model. VANET-Mobisim is also easily interfaced with NS2. Specifically, VANET-Mobisim uses a mobility file in XML format, which contains all the detailed informations of the microscopic and macroscopic models that govern mobility of nodes. The mobility model used in this software takes into account the environmental parameters of the mobile nodes (traffic lights, speed limits, etc.) and possible interactions between mobile nodes. A node may thereby accelerate, decelerate according to environment constraints.

The global parameters for the simulations are given in Table III.

TABLE III. SIMULATION PARAMETERS.

Parameters	Values
Network simulator	ns-2
Simulation time	180s
Simulation area	1000m*1000m
Maximum number of transmissions	4
Transmission power	0.1w
Data types	CBR
Data packet size	512 bytes
MAC layer	IEEE 802.11a

We also use a realistic model of the Munich town (urban outdoor environment, see Figure 1), obstacles (building, etc.) are printed red. Dots represent nodes. Other real environments could be used in a more comprehensive study.

As routing protocols, we compare AODV-st, the standard AODV protocol [1], to the three PR-metric based ones presented in Section III.

B. Simulation results

In this section, we study the impact of mobility on performance of the four protocols. 60 mobile nodes move in the Munich town environment (Figure 1). Their average speeds range from 4m/s to 20m/s. 10 simultaneous end-to-end transmissions are initiated during 165s. As performance parameters we rely primarily on average end-to-end delay of data packets, PDR and Routing Overhead (RO). End-to-End Delay concerns only successfully delivered packets. PDR is the ratio of the number of successfully delivered data packets over the number of sent data packets. Routing overhead is the number of routing protocol control packets. It permits



Figure 1. Simulation environment when number of nodes=60. Obstacles are printed red.

to evaluate the effective use of the wireless medium by data traffic.

Figure 2 and Figure 3 show that AODV-new-timer is better in delay and PDR. It reduces unnecessary waiting time and the knowledge of neighborhood, in real time, avoids node to process obsolete paths. The node implementing AODV-new-Timer detects links breakage quickly. For QoS-based AODV (AODV-BL-PR and AODV-sum-PR), determining QoS routes requires substantial time and with node mobility, established routes become obsolete quickly. These show that better neighborhood information control is more important than taking into account link quality for AODV efficiency.

A thorough analysis of the simulation shows that the majority of communications where source and destination are far apart from each other have failed. Established routes become obsolete even before the first data packets arrive at the destination.

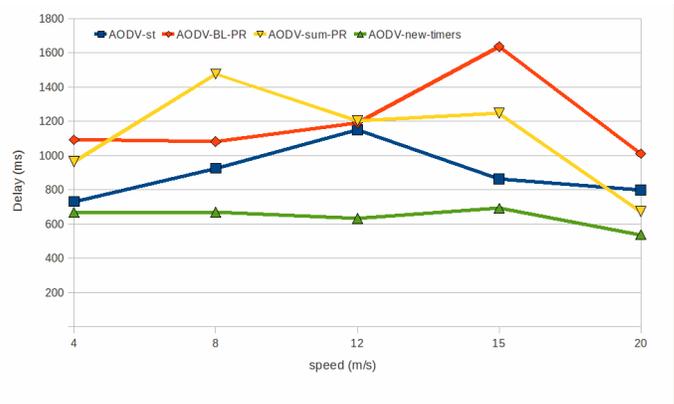


Figure 2. Delay evolution when speed increases.

Protocol's performance in RO parameter is presented in

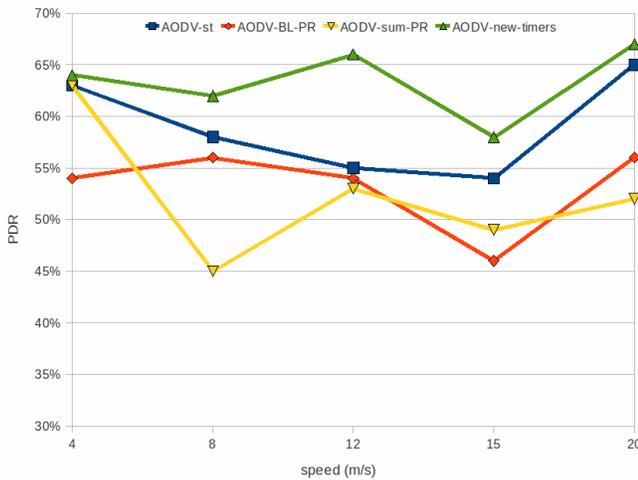


Figure 3. PDR evolution when speed increases.

Figure 4. The high cost of AODV-new-Timer is expected since Hello and RREQ messages emitting frequency increased. The better performance of QoS-based AODV compared to standard one can be explained by better paths selection. In addition, blacklisting approach of AODV-BL-PR limits the dissemination of RREQ messages.



Figure 4. RO evolution when speed increases.

## V. CONCLUSION AND FUTURE WORK

We tested the effectiveness of different QoS-based methods under realistic wave propagation model and realistic mobility model. For QoS metric, we use number of retransmissions count-based metric. Although we used a simple and effective method for link quality estimation, the results show that taking into account the quality of links is not effective for the MANET performances improvement. The additional complexity, induced by QoS management, increases delay and precipitated the obsolescence of the links.

To achieve better performance in high speed MANET context, the real challenge is the effective control of node neighborhood and accurate established routes lifetime and waiting RREP packet timeout.

A solution where the inventory frequency of the neighborhood depends on the network dynamics might improve the performance of on-demand routing approach in mobility contexts.

A more comprehensive study of the problematic of on-demand routing protocol performance could concern other real environments (than Munich town one) and a refinement of the penalty coefficient due to retransmissions.

## ACKNOWLEDGEMENT

I would like to thank my colleagues of XLIM-SIC laboratory of University of Poitiers for simulation tools developed including Communication Ray Tracer (CRT) simulator. I especially thank Professor Philippe Meseure for everything he taught me as knowledge, especially for the writing of scientific articles and metrics design.

I also thank the Polytechnic University of Bobo Dioulasso for facilitating the financing of my participation in the conference where I presented this paper.

## REFERENCES

- [1] C. Perkins, E. Belding-Royer, and S. Das, "Ad hoc on-demand distance vector (AODV) routing protocol," Tech. Rep. RFC3561, IETF Network Working Group, July 2003.
- [2] K. A. Al-Soufy and A. M. Abbas, "A path robustness-based quality of service routing for mobile ad hoc networks," in *2010 IEEE 4th International Conference on Internet Multimedia Services Architecture and Application (IMSAA)*, pp. 1–6, IEEE, Dec. 2010.
- [3] Y. Chen, Z. Xiang, W. Jian, and W. Jiang, "A Cross-Layer AOMDV routing protocol for V2V communication in urban VANET," in *5th International Conference on Mobile Ad-hoc and Sensor Networks, 2009. MSN '09*, pp. 353–359, IEEE, Dec. 2009.
- [4] S. Kim, O. Lee, S. Choi, and S. Lee, "Comparative analysis of link quality metrics and routing protocols for optimal route construction in wireless mesh networks," *Ad Hoc Networks*, vol. 9, pp. 1343–1358, Sept. 2011.
- [5] P. Esposito et al., "Implementing the expected transmission time metric for OLSR wireless mesh networks," in *Wireless Days 2008 (WD08)*, pp. 1–5, Nov. 2008.
- [6] N. Enneya, K. Oudidi, and M. Elkoutbi, "Enhancing delay in manet using olsr protocol," *International Journal of Computer Science and Network Security IJCSNS*, vol. 9, no. 1, 2009.
- [7] M. Benzaid, P. Minet, and K. Al Agha, "Integrating fast mobility in the OLSR routing protocol," in *4th International Workshop on Mobile and Wireless Communications Network, 2002*, pp. 217 – 221, 2002.
- [8] A. Chintawar, M. Chatterjee, and A. Vidhate, "Performance analysis of ad-hoc on demand multipath distance vector routing protocol with accessibility and link breakage prediction," in *IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET)*, pp. 1–6, 2011.
- [9] A. Khosrozadeh, A. Akbari, M. Bagheri, and N. Beikmahdavi, "New AODV routing protocol with break avoidance," in *2011 International Symposium on Computer Science and Society (ISCCS)*, pp. 303–306, IEEE, July 2011.
- [10] W. Hamidouche, R. Vauzelle, C. Olivier, Y. Pousset, and C. Perrine, "Impact of realistic mimo physical layer on video transmission over mobile ad hoc network," in *20th Personal, Indoor and Mobile Radio Communications Symposium, 2009*.

- [11] M. Guenes, M. Wenig, and A. Zimmermann, "Realistic mobility and propagation framework for manet simulations," in *NETWORKING 2007: Ad Hoc and Sensor Networks, Wireless Networks, Next Generation Internet, Proceedings*, vol. 4479 of *Lecture Notes in Computer Science*, pp. 97–107, Springer Verlag, 2007.
- [12] N. Alsharabi, L. Ping, and W. Rajeh, "Avoid link breakage in on-demand ad-hoc network using packet's received time prediction," in *Proceedings 19th European Conference on Modelling and Simulation*, 2005.
- [13] D. S. J. D. Couto, D. Aguayo, J. Bicket, and R. Morris, "a high-throughput path metric for multi-hop wireless routing," *Wireless Networks*, vol. 11, pp. 419–434, July 2005.
- [14] T. Yélémou, P. Meseure, and A.-M. Poussard, "Improving zrp performance by taking into account quality of links," in *WCNC 2012 Paris, France*, April 2012.
- [15] L. Breslau et al., "Advances in network simulation," *COMPUTER*, vol. 33, May 2000.
- [16] R. Delahaye, *Efficient and Realistic Simulation of the Physical Layer to Aid Routing in Ad hoc Networks*. PhD thesis, University of Poitiers, 2007.
- [17] J. Haerri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks : a survey and taxonomy," in *IEEE communications surveys and tutorials*, vol. 11, pp. 19–41, IEEE-INST Electrical Electronics Engineers INC, 2009.
- [18] Q. Zheng, X. Hong, and S. Ray, "Recent advances in mobility modeling for mobile ad hoc network research," in *Proceedings of the 42nd annual Southeast regional conference*, pp. 70–75, ACM, 2004.
- [19] X. Hong, M. Gerla, G. Pei, and C.-C. Chiang, "A group mobility model for ad hoc wireless networks," in *Proceedings of the 2nd ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems*, MSWiM '99, (New York, NY, USA), p. 5360, ACM, 1999.
- [20] F. Bai, N. Sadagopan, and A. Helmy, "Important: a framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *22th Annual Joint Conference of the IEEE Computer and Communications*, vol. 2, pp. 825–835, 2003.

# A Channel-Aware Uplink Scheduling Algorithm for Mobile 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

Mohamed N. El-Derini

Computer & Systems Engineering Dept.  
Alexandria 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

Mostafa A. Youssef

Computer Science and Engineering  
E-JUST  
Alexandria, Egypt  
moustafa@cs.umd.edu

**Abstract-** There has been a rapid growth of new Internet services, such as video conferences, online video games, and multimedia applications, offered to end users. These services need to satisfy their quality requirements, and thus an efficient scheduling algorithm is needed. In the literature, the interest is focused on throughput and delay as inputs to the scheduler in its bandwidth allocation decision. Jitter, though of great significance, did not receive considerable attention, yet. Researchers, in the area of WiMAX networks, often recommend weighted scheduling algorithms with dynamic weight functions. The channel quality is particularly important as well in scheduler decision in wireless networks for the determination of channel strength. In this paper, we develop an uplink channel-aware scheduling algorithm for mobile WiMAX networks. Use is made of a weight function with four terms: throughput, delay, jitter, and channel quality. A comparison is made between the proposed algorithm and two famous channel-aware algorithms, namely, proportional fair scheme (PFS) and maximum carrier-to-interference ratio (Max C/I). Simulation results, obtained by an OPNET simulator, reveal that our algorithm outperforms both PFS and Max C/I with respect to WiMAX delay and jitter, as functions of the number of mobile stations. However, the WiMAX throughput takes on a slightly lower value. For real-time applications, the algorithm is applied to a video conference and high quality video applications, and better values for both delay and jitter are attained in both application types.

**Keywords:** *WiMAX Networks; IEEE 802.16e; channel-aware algorithms; scheduling schemes; QoS*

## I. INTRODUCTION

Worldwide interoperability for microwave access (WiMAX) [1] is one of the emerging broadband wireless access networks. It is designed according to IEEE 802.16 standard [2]. The rapid growth of new services, such as online video games, video conferences, and multimedia services demands a reliable and efficient Internet access. WiMAX is an efficient solution to provide last-mile access to the Internet. It is suitable for real-time applications since WiMAX has a multitude of advantages features in this respect [1][3]; these are (1) use of orthogonal frequency division multiple access (OFDMA) in the physical layer, which allows WiMAX to operate in Non-Line of sight (NLoS) by using multiaccess scheme for broadband wireless access; (2) having a high data rate, capable of using 74Mbps on 20MHz wide spectrum; (3) supporting adaptive modulation and coding rate technology (AMC); (4) being suitable for both time division duplexing (TDD) and

frequency division duplexing (FDD); (5) enjoying strong and robust security; and (6) allowing Quality of Service (QoS), which is responsible for serving different applications with dissimilar requirements.

QoS is the main feature of WiMAX networks [3][4], used to manage the available resources in such a way as to enhance the performance of the network. It has three main parameters, namely, throughput, delay, and jitter. QoS in broadband wireless access is highly important. Yet its achievement is sophisticated since the performance of the radio link channel is unpredictable. The channel status indicating parameter is the carrier-to-interface-plus-noise ratio (CINR) [1][2].

A scheduling algorithm is part of the QoS architecture. Its function is the division of bandwidth among subscriber stations (SSs) in order to maximize throughput and minimize both delay and jitter. The scheduler should be simple, fair, and efficient [3][4]. The wireless channel is influenced by many factors such as: signal attenuation, fading, interference, and noise ratio. So, it is preferred to use a scheduling algorithm which can take the channel status in its bandwidth allocation decision [1][3].

A comprehensive survey on scheduling algorithms in WiMAX networks is given in [3]. In [5], a detailed performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks is made, where simulation analysis is carried out using average delay, average throughput, fairness and frame utilization and the simulation results indicate that none of the algorithms considered is capable of effectively supporting all WiMAX classes of service. Recently, in [6], a comparative descriptive analysis for various scheduling algorithms in WiMAX networks is presented.

The work in [7] is focused on the importance of the scheduling algorithm in WiMAX networks to ensure the usefulness of delay, jitter and throughput. Also, it explains deficit round robin (DRR) and weighted deficit round robin (WDRR), but no analysis study is given. In [8], a new packet scheduling and bandwidth allocation algorithm is developed. This scheduling algorithm is divided into two tiers. In tier 1, four queues are used by the service classes and each service class uses an intra-class scheduling algorithm to decide which packet will be served. In tier 2, the chosen packets are arranged in a matrix and an inter-class scheduling algorithm is used to move packets to the frame. Two inter-class scheduling algorithms, dynamically allocating priority queue (DAPQ) and dynamically allocating weighted fair queue

(DAWFQ) are used. A simulation analysis using OPNET is presented. A new strategy about scheduling algorithms based on intelligent systems is suggested in [9]. The suggested algorithm is based on fuzzy systems and neural networks. This algorithm is divided into two stages: Priority stage using fuzzy systems and the allocation of bandwidth stage using neural networks. A performance comparison between the suggested algorithm and some scheduling algorithms are presented using OPNET and Matlab. All these algorithms belong to the channel-unaware category.

Weighted scheduling algorithms are preferred for the satisfaction of QoS requirements [10]. 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 [7] [10], in which floating point numbers are used. Further, the resulting algorithm will be much less sophisticated. In [11], a study of the performance of four scheduling algorithms: round robin (RR), MAX CINR, fair throughput (FT), and proportional fair scheme (PFS) is presented. The results reveal that MC has the highest throughput but with lowest fairness, but the converse happens for FT scheduling. PFS has the ability of adjusting the throughput and fairness with application requirements. In [12], three channel-aware scheduling algorithms in mobile WiMAX networks: PFS, modified longest weighted delay first (MLEDF), and exponential rule are studied and compared with a suggested algorithm using a queue length and waiting time of the packet in a weight equation. The results indicate that the suggested algorithm outperforms the other algorithms in throughput and delay.

It is known that channel-aware schedulers are more suitable than channel-unaware schedulers in ensuring QoS requirements for wireless networks [1][3]. Also, the weighted scheduling algorithms are preferred to use in WiMAX networks [10]. The present paper develops a channel-aware uplink scheduling algorithm for mobile WiMAX networks. Its bandwidth allocation decision is taken based on throughput, delay, jitter, and channel quality. The proposed algorithm is compared with PFS, and MAX CINR with performance metrics: WiMAX throughput, delay, jitter, real-time application delay, and real-time application jitter. The results reveal that the proposed algorithm outperforms the other algorithms, with respect to WiMAX delay, and jitter as functions of the number of mobile stations, but the proposed algorithm gives a lower throughput than both algorithms PFS and Max C/I. Results of the proposed algorithm demonstrate that delay and jitter in real-time applications evidence that our algorithm transcends the other algorithms.

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. WiMAX FRAMEWORK

WiMAX is the most efficient technique of broadband wireless access networks [1][3]. It is used as a last-mile

network to introduce the Internet for end users in an efficient and reliable way. WiMAX networks based on IEEE 802.16 standard are divided into two main layers: the physical layer (PHY) and the medium access control layer (MAC).

The PHY layer is defined on IEEE 802.16 standard [2] using four physical layer types: wireless MAN-OFDM (orthogonal frequency division multiplexing), wireless MAN-SC (single carrier), wireless MAN-SCa, and wireless MAN-OFDMA (orthogonal frequency division multiple access).

WiMAX uses two types of duplexing [1][2][3]: frequency division duplexing (FDD) and time division duplexing (TDD) in the frame structure. In our work, we use TDD. In TDD, the frame structure consists of two subframes: downlink subframe and uplink subframe. Downlink subframe sends data from the base station (BS) to subscriber stations (SSs) together with some control information such as: preamble, downlink and uplink maps. Uplink subframe implies uplink bursts in addition to control information, such as channel quality information, which is sent from SSs to BS. Downlink and uplink subframes are separated by a transmit-receive transition gap (TTG) and a receive-transmit transition gap (RTG).

The MAC layer [1][3][4] is the intermediate layer between the PHY layer and higher layers. It is responsible for many important jobs such as header suppression, packet scheduling, bandwidth allocation, QoS management, and security and authentication issues.

To facilitate MAC layer work, the MAC layer is divided into three sub-layers. Each sub-layer is responsible for doing some of MAC functions. The three sub-layers are: convergence sub-layer, common-part sub-layer, and security sub-layer. First, a convergence sub-layer is designed as a link between 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 is responsible for bandwidth allocation, connection establishment and maintenance for all QoS requirements. Third, a security sub-layer is developed for authentication, security key exchange, and encryption.

WiMAX supports two types of operation modes [5]: Point-to-multipoint (PMP) mode and mesh mode. In PMP mode, the communications between all SSs are organized and passed through the BS. But in mesh mode, the communications can be achieved directly between SSs.

WiMAX supports QoS by defining five different service classes for constant and variable bit rate applications. These service classes are [3][5]: unsolicited grant service (UGS), used to support constant data rate real-time applications such as VoIP without silence suppression; real-time polling service (rtPS), defined to support real-time applications with variable data rate such as a MPEG compressed video; extended real-time polling service (ertPS), used to support real-time applications with variable data rate such as VoIP with silence suppression; non-real-time polling service (nrtPS), defined for variable bit rate non-real-time applications; and the best effort (BE) service class, responsible for non-real-time applications with no need of any special requirements.

To ensure a good performance of WiMAX networks for the different requirements of QoS in real-time applications, a suitable bandwidth allocation algorithm is needed [3][4]. In the beginning of each WiMAX frame, the scheduling

algorithm computes the allocated bandwidth for each SS to send this information in UL-MAP to SSs.

### III. EARLIER SCHEDULERS

A powerful scheduling algorithm is essential in WiMAX networks to satisfy the growth of end user requirements for different applications [4]. There is no specific scheduling algorithm stated in IEEE 802.16 standard to use. The selection of the algorithm is left for service providers to pick a suitable one, which satisfies network application requirements.

Scheduling algorithms are classified into two main classes: channel-aware and channel-unaware scheduling algorithms, outlined in what follows.

#### A. Channel-unaware algorithms

In channel-unaware algorithms, the bandwidth allocation is done without any use of information about the channel status. These algorithms include: the round robin (RR) algorithm [3] is simple and fair in assigning one allocation for each connection in each serving cycle. The weighted round robin (WRR) [3][5][6] assigns a weight to each connection; then the connections are served according to their weights. The main problem of WRR is that it provides incorrect percentage of bandwidth allocation when the traffic has a variable packet size. The deficit round robin (DRR) [7][10] solves this problem of WRR. DRR defines two variables for each queue, deficit counter (DC) and quantum (Q). Q is set to a constant value equal to the maximum traffic packet of the queue, and DC is initialized by a zero value when the queue is created. When the queue is visited to serve, the value of Q is added to DC and the queue is still served until the head packet size is greater than DC. For each packet served, the value of DC decreases by the value of packet size. When the queue is empty, DC returns to zero. The deficit weighted round robin (DWRR) [13] is the same as DRR; it adds a weight variable for each queue and the Q value depends on the weight value. Another modification on DRR, named modified deficit round robin (MDRR) [13], works in the same way as DRR but a queue priority parameter is added to each queue to contribute to queue selection.

#### B. Channel-aware algorithms

These algorithms use channel state information from the channel quality indicator (CQI) to make the bandwidth allocation decision. Channel-aware algorithms include modified largest weighted delay first (MLWDF), proportional fairness schema (PFS), and maximum carrier to interference ratio (MAX C/I). MLWDF [14] is one of QoS-guaranteed algorithms which support minimum throughput and minimum delay. In this algorithm, for each queue  $j$  the scheduler computes a function " $\rho_i * W_j(t) * r_j(t)$ ", where  $\rho_i$  is a constant which should take a different value for each service class,  $W_j(t)$  can be either the delay of the head of line packet or the queue length, and  $r_j(t)$  is the channel capacity for traffic  $i$ . The queue selection occurs on the basis of the function value starting from the largest value. There are many modifications of MLWDF. PFS [15] belongs to a fairness scheduler family which works based on maximizing

the long-term fairness. PFS uses a ratio of channel capacity  $W_i(t)$  to the long-term throughput  $R_i(t)$  to select the queue which will be served. The queue selection occurs based on the ratio value starting from the largest value. The main disadvantage of PFS is that delay is not taken into account when defining the weight function. MAX C/I [11] is used to maximize the throughput. In MAX C/I, the queue is selected based on the best channel conditions. In WiMAX, the most used channel quality indicator is CINR. This algorithm checks the value of CINR for each queue and the queue with the largest CINR is served first. The movement between the queues occurs based on the CINR value in descending order.

All the above-mentioned scheduling algorithms, channel-aware and channel-unaware, have the following drawbacks. In weighted scheduling algorithms, the bandwidths are assigned statically and do not vary with the burst changes; not enough attention is given to jitter, causing problems in real-time applications; priority scheduling algorithms caused starvation in low priority classes; and finally, the use of channel-aware scheduling algorithms are preferred because of the nature of wireless communications. According to [3], and to the best of our knowledge, no channel-aware scheduling algorithms using jitter delay in its weight function are available in the literature.

### IV. THE PROPOSED ALGORITHM

In WiMAX networks, the BS is responsible for scheduling service classes in both uplink and downlink directions. Any scheduling algorithm works on the basis of the bandwidth requests of SSs in the uplink direction [2][3][4].

The proposed approach is a type of channel-aware weighted scheduling algorithms with a weight equation defined in terms of: throughput, delay, jitter, and channel quality. These parameters characterize the QoS and the wireless communication 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 channel quality is important to be considered for both real- and non-real-time applications.

The problem under consideration is concerned with the development of a dynamic uplink scheduling algorithm for mobile 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 weight function in terms of the parameters: throughput, delay, jitter, and channel quality. To this end, a weight  $W_i(t)$  is assigned to queue  $i$  as a positive factor of the form:

$$W_i(t) = \frac{S_i(t)}{\sum_{j=1}^n S_j(t)} \quad 1 \leq i \leq n \quad (1)$$

In (2),  $S_i(t)$  is expressed as the sum of four terms corresponding to contributions of throughput, delay, jitter, and channel quality, respectively. Specifically, we propose the following formula for a weight function  $S_i(t)$ :

$$S_i(t) = TH_i + Dly_i(t) + JTR_i(t) + CH_i(t) \quad (2)$$

The first term  $TH_i$ , in (3), is the fractional throughput contribution to  $S_i(t)$ , defined as:

$$TH_i = \frac{X_i}{\sum_{j=1}^n X_j} \quad (3)$$

where  $X_i$  is the minimum reserved traffic rate for queue  $i$ . The second term  $Dly_i(t)$  is the fractional delay contribution

$$Dly_i(t) = \frac{\alpha_i Y_i(t) / L_i}{\sum_{j=1}^n \alpha_j Y_j(t) / L_j} \quad (4)$$

where  $Y_i(t)$  is a time-varying average delay,  $L_i$  is the given maximum latency, and  $\alpha_i$  is a positive delay weighting factor. In (4), the ratio  $Y_i(t)/L_i$  (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  $Y_i(t)/L_i$  is weighted by a factor  $\alpha_i$ , 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  $JTR_i(t)$  is the fractional jitter contribution,

$$JTR_i(t) = \frac{\beta_i Z_i(t) / K_i}{\sum_{j=1}^n \beta_j Z_j(t) / K_j} \quad (5)$$

where  $Z_i(t)$  is a time-varying average jitter,  $K_i$  is the given maximum jitter and  $\beta_i$  is a positive jitter weighting factor. The terms in (5) can be interpreted in the same way as in (4). The fourth term  $CH_i(t)$  is the fraction channel quality indicator

$$CH_i(t) = \Omega_i * \frac{CINR_i(t)}{\sum_{j=1}^n CINR_j(t)} \quad (6)$$

where  $CH_i(t)$  is the carrier-to-interface-plus-noise ratio of the channel between BS and the MS,  $\Omega_i$  is the CINR status which indicates that CINR increases or decreases.  $\Omega_i$  takes on a value of +1 when CINR increases and the value -1 when CINR decreases. Then the bandwidth is divided among the  $n$  queues using the form given in (7):

$$BW_i(t) = W_i(t) * UL_{BW} \quad (7)$$

where  $BW_i(t)$  is the bandwidth reserved to queue  $i$  and  $UL_{BW}$  is the total bandwidth of the uplink subframe.

Equation 2 is valid for both real- and non-real time applications; this implies that the weighting factors  $\alpha_i$  and  $\beta_i$  should take on different values of the two types of applications. The values of  $\alpha_i$  and  $\beta_i$  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 sensitive to delay and jitter. The criterion for the choice of the values of  $\alpha_i$  and  $\beta_i$  depends on a developed algorithm which is introduced in [16] on the basis of the importance of delay for real- and non-real-time applications. We use the ratio 1:10 for  $\alpha_i$  to  $\beta_i$ .

The computational scheme of the proposed algorithm is itemized in the following consecutive steps:

1) For each queue, get the values of  $Y_i(t)$  and  $Z_i(t)$ .

- 2) Calculate the values of  $TH_i$  (in (3)),  $DLY_i$  (in (4)),  $JTR_i$  (in (5)), and  $CH_i(t)$  (in (6)).
- 3) Calculate the four-term weight function  $S_i(t)$  according to (2).
- 4) Calculate the weight  $W_i(t)$  by virtue of (1)
- 5) Divide the bandwidth of the uplink subframe among the  $n$  queues based on (7).
- 6) The value of the bandwidth of each queue is sent to MS.
- 7) The service for the queue is continued until the assigned division of the bandwidth is used up.
- 8) The service is moved between the queues using round robin mechanism.

## V. EXPERIMENTAL SCENARIOS AND SIMULATION ANALYSIS

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 taken as those used in [5], listed in table I. Simulation in this paper is performed by the OPNET simulator [17].

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 experimental scenarios by varying the number of mobile stations (MSs). Each scenario consists of one BS serving a number of MSs in PMP mode of operation. The frame duration is 5 msec, with the uplink and downlink subframes having 50% of this duration each. A random topology in 1000 x 1000 m square space is used. The number of MSs varies from 6 to 36 with ratio 1:2:2:1 MSs for service classes ERTPS:RTPS:NRTPS:BE, respectively. The proposed weighted scheduling algorithm is compared with PFS and Max C/I. The WiMAX throughput, delay, jitter, real-time application delay, and real-time application jitter are considered as performance metrics. Simulation time is 10 minutes.

The results of the overall WiMAX throughput, delay, and jitter are shown in Figures 1, 2, and 3, respectively. The performance of the real time applications delay and jitter are shown in Figures 4, 5, 6, and 7, respectively.

As shown in Figure 1, the proposed algorithm has a lower throughput than PFS and Max C/I, and a higher throughput varying between PFS and MAX C/I. As mentioned in [14][15], PFS and Max C/I are designed to maximize throughput with no delay guarantee in PFS, but Max C/I supports delay minimization.

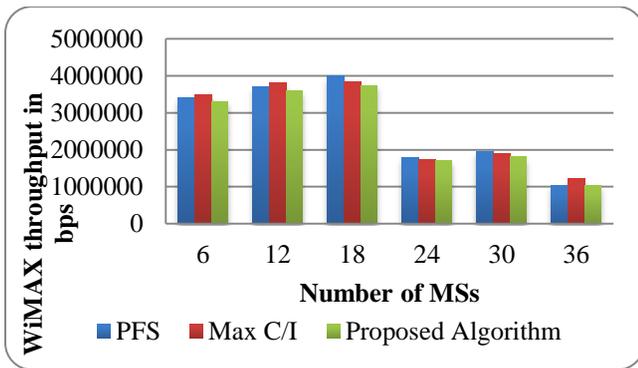


Figure 1: Average WiMAX Throughput in bps vs. number of mobile stations

As shown in Figure 2, it is clear that the proposed algorithm has a lower delay than PFS and MAX C/I for increasing the number of SSs. PFS has a higher delay since it does not support any delay guarantee [14]. Max C/I has a delay less than PFS since Max C/I supports delay minimization [15].

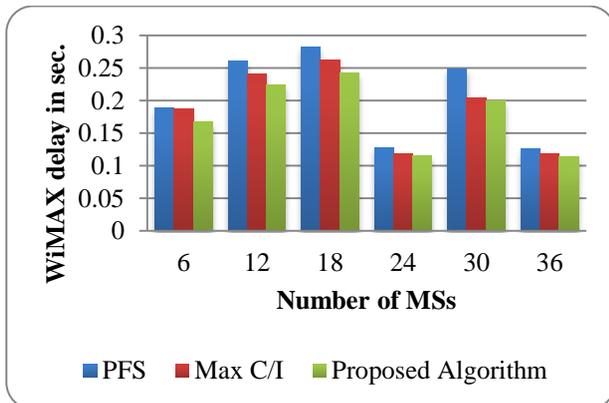


Figure 2: Average WiMAX delay in sec. vs. number of mobile stations

From Figure 3, we can conclude that the proposed algorithm has a lower jitter than PFS and MAX C/I for increasing the number of SSs.

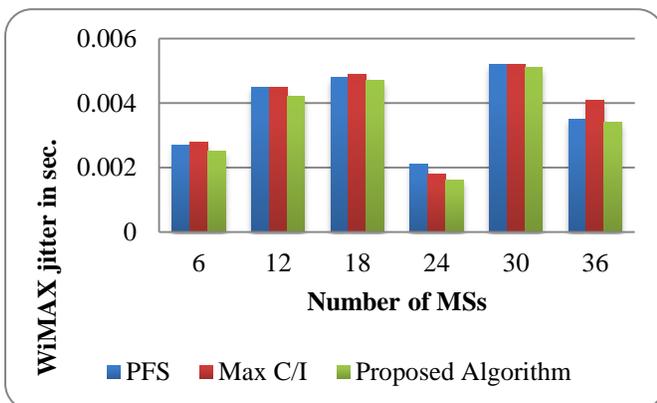


Figure 3: Average WiMAX jitter in sec. vs. number of mobile stations

In Figure 4, the results of the video conference delay are presented, and we can see that the proposed algorithm

presents the best video conference delay value when the number of MSs changed.

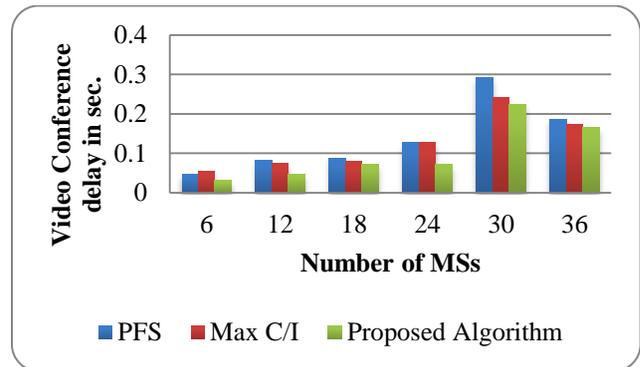


Figure 4: Video Conference delay in sec. vs. number of mobile stations

In Figure 5, the results of the video conference jitter are displayed, and we can see that the proposed algorithm has a lowest jitter value than PFS and Max C/I when the number of MSs changed.

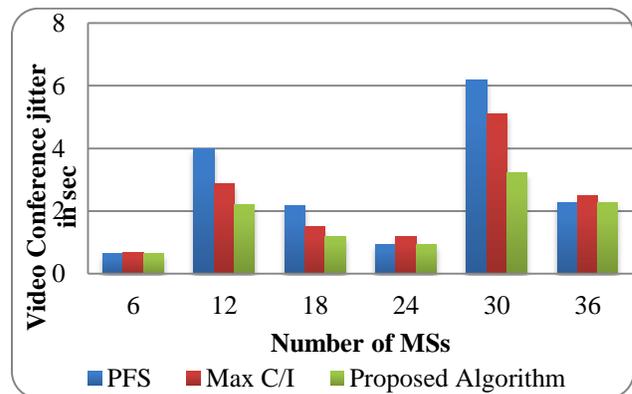


Figure 5: Video Conference jitter in sec. vs. number of mobile stations

High quality video performance delay and jitter are shown in Figures 6 and 7. From Figures, we conclude that the proposed algorithm has better values for both delay and jitter than the others algorithms.

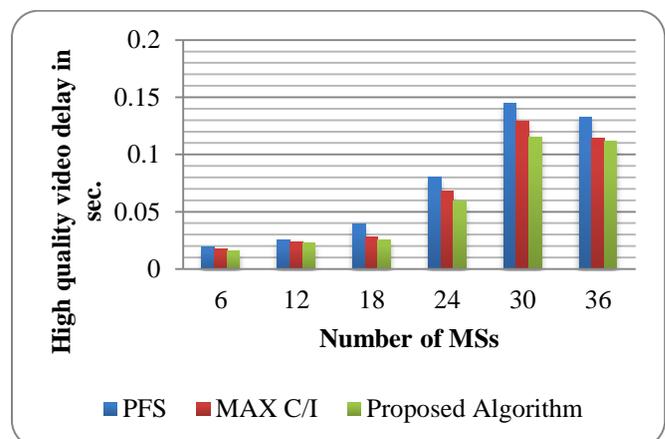


Figure 6: High Quality Video delay in sec. vs. number of mobile stations

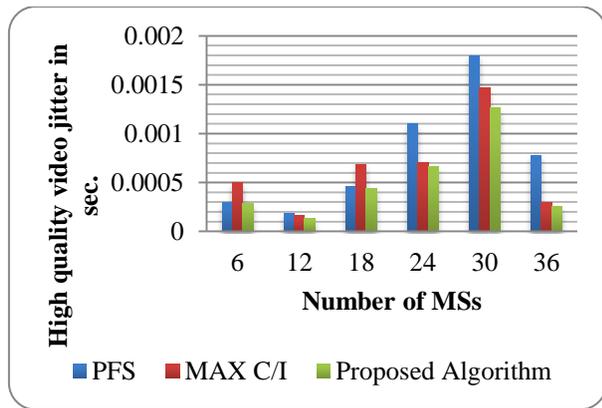


Figure 7: High Quality Video jitter in sec. vs. number of mobile stations

The best performance of our algorithm is caused by the using of delay and jitter contribution terms in the weight function with high importance for real-time applications. But the other algorithms (PFS and MAX C/I weight functions focused on the throughput contribution only.

## VI. CONCLUSIONS

An uplink channel-aware weighted scheduling algorithm for mobile WiMAX networks has been proposed. This algorithm has advantages: it is weighted scheduling algorithm with the use of jitter and channel quality as parameters in its weight function. A comparison is made with two powerful algorithms: PFS and Max C/I. The proposed algorithm is simulated using OPNET.

The results demonstrate that the proposed algorithm outperforms the other algorithms with respect to WiMAX delay, and jitter as functions of the number of mobile stations; also, the proposed algorithm gives a lower throughput and higher load than both algorithms PFS and Max C/I. The algorithm is applied to a video conference and high quality video applications. The results of the real-time applications delay and jitter show that our algorithm transcends the others. The performance evaluation of the proposed algorithm in Long-Term Evolution (LTE) networks [18] will be considered as a future work.

## VII. 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. 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, pp.156-171.
- [4] C. Cicconetti, A. Erta, L. Lenzi, and E. Mingozzi, "Performance evaluation of the IEEE 802.16 MAC for QoS support", IEEE Trans. Mob. Comput., 6, (1), 2007, pp. 26-38.
- [5] N. Abu Ali, P. Dhrona, and H. Hassanein, "A performance study of uplink scheduling algorithms in point-to-multipoint WiMAX networks", journal of Computer Communications, no. 32, 2009, pp. 511-521.
- [6] B. Kumar and P. Gupta, "Scheduling Algorithms in WiMAX Network", Second International Conference on Advanced Computing and Communication Technologies, 2012, pp. 457-462.
- [7] H. Javier and I. Patricia, "Evaluation of Scheduling Algorithms in WiMAX Networks", IEEE ANDESCON, 2010, pp. 1-4.
- [8] A. Esmailpour and N. Nassar, "A Novel Scheme for Packet Scheduling and Bandwidth Allocation in WiMAX Networks", IEEE International Conference on Communications, 2011, pp. 1-5.
- [9] D. David and K. Murugessan, "Performance Analysis of Neural Networks Based Priority Scheduler for WiMAX under Bursty Traffic Conditions", European Journal of Scientific Research, vol. 76, no. 3, 2012, pp. 351-365.
- [10] 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.
- [11] W. Wang and et. al., "Implementation and performance evaluation of QoS scheduling algorithms in mobile WiMAX NS-2 simulator", 4th international conference on Signal Processing and Communication Systems (ICSPCS), 2010, pp.1-6.
- [12] P. Chandur, R. Karthik, and K. Sivalingam, "Performance evaluation of scheduling algorithms for mobile WiMAX networks", Eighth IEEE PerCom Workshop on Pervasive Wireless Networking, 2012, pp. 764 -769.
- [13] 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.
- [14] H. Kim and Y. Han, "A Proportional Fair Scheduling for Multicarrier Transmission Systems", IEEE Commun. Lett., vol. 9, 2005, pp. 210-212.
- [15] V. Singh and V. Sharma, "Efficient and Fair Scheduling for Uplink and Downlink in IEEE 802.16 OFDMA Networks", IEEE Wireless Communication and Networking Conference, Lass Vegas, vol. 2, 2006, pp. 984-990.
- [16] D. Ali and K. Dimyati, "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.
- [17] OPNET Technologies, Opnet modeler Software Package available: www.opnet.com.
- [18] S. Sesia, I. Toufik, and M. Baker, LTE - The UMTS Long Term Evolution: From Theory to Practice, John Wiley & Sons Ltd., 2011.

# Novel Pico-cell Range Expansion with Adaptive RACH Resource Allocation for Random Access of M2M Devices

Woo-Jong Jo, Chang-Yeong Oh,

Samsung Electronics  
Suwon, Korea

Email: {wjong.jo, changyong.oo}@samsung.com

Yunmin Kim, and Tae-Jin Lee

School of Electronic and Electrical Engineering  
Sungkyunkwan University  
Suwon, Korea

Email: {kym0413, tjlee}@skku.edu

**Abstract**—In the next-generation cellular networks, Machine-to-Machine (M2M) communications, in which machine-type devices communicate with each other without user's control, is considered to create new services. In M2M communications, a large number of M2M devices distributed in a cell area may exist. So, the random access congestion problem caused by massive and simultaneous random access trials of user equipments (UEs) is one of the most important issues to be addressed. For next-generation cellular networks, an efficient random access control scheme is required to decentralize random access trials from a macro-cell to small-cells in a heterogeneous network (HetNet). So, we consider a new random access control scheme by adjusting pico-cell's coverage dynamically to improve the performance of the random access procedure in HetNets. In this paper, we propose a new Pico-Cell Range Expansion (PCRE) scheme with dynamic Random Access CHannel (RACH) resource allocation. We show that the random access efficiency and random access delay of UEs can be improved.

**Keywords**—HetNets; M2M communications; random access control; cell range expansion.

## I. INTRODUCTION

In the next-generation cellular networks, Machine-to-Machine (M2M) communications, which is also known as Machine-Type Communications (MTC) in the Third Generation Partnership Project (3GPP) specifications [1], is considered as a new key technology to create new services. In M2M communications, MTC Devices (MTCs), i.e., different type of User Equipments (UEs) compared to legacy UEs for Human-to-Human (H2H) communications, communicate with each other anytime and anywhere without user's control. Because of various applications for M2M communications, e.g., metering the usage of water, electricity, and gas and road security for announcing the emergency message, a very large number of MTCs can spread in a cell area and work simultaneously. So, one of the main issues for M2M communications is a random access congestion problem caused by the concurrent and massive random access trials of MTCs.

In order to reduce the congestion problem, some approaches on random access controls in a Long Term Evolution (LTE) system have been discussed [2]. In [3], a preamble separation scheme which splits a set of available preambles and allocates them to MTCs separately is introduced. A part of preambles for MTCs are shared by M2M and H2H UEs, but the dedicated preambles are always provided to H2H UEs. And a dynamic Random Access CHannel (RACH) resource allocation scheme for dynamic traffic is proposed by

Yilmaz et al. [4]. Amirjoo et al. [5] present an adaptive power control mechanism for preamble transmission according to the variation of random access loads.

However, conventional random access control schemes tend to focus on dealing with random access trials in a single cell. For the next-generation cellular network, a new random access control scheme considering the features of Heterogeneous Networks (HetNets) is required. HetNet is introduced by 3GPP to handle the growth of mobile data traffic load due to the wide spread of smart phones and tablet PCs. In a HetNet, various small-cells formed by low power network entities, such as pico-eNodeB, femto-eNodeB, and relay-eNodeB, can coexist and operate along with the conventional macro-eNodeB [6]. So, the random access loads can be spread from a macro-cell to small-cells. In particular, pico-cells can take over the random access loads of a macro-cell efficiently since they can be deployed by the operator's cell planning [7]. For an efficient random access congestion control in a HetNet, balancing random access loads between a macro-cell and small-cells is important.

As an approach for balancing random access loads in HetNets, we consider controlling the coverage area of pico-cells dynamically by utilizing Cell Range Expansion (CRE), which has been introduced initially and discussed by 3GPP [8]. If the CRE scheme is properly applied to pico-cells, the random access congestion problem can be relieved by decentralizing the random access load of a macro-cell to pico-cells. In [9], they introduce a simple CRE scheme and show simulation results when the random access loads are offloaded from macro-cells to pico-cells depending on different bias values. Guvenc [10] analyzes the impact of capacity and fairness by applying the CRE scheme to HetNets. The CRE scheme can be realized by the power control of eNodeB. Morimoto et al. [11] proposed a power control mechanism for the CRE. The simulation results show that the throughput of the edge cell area is improved although the entire cell throughput slightly decreases. Liu et al. [12] proposed an adaptive uplink power control of pico-eNodeB to decrease the inter-cell interference where power control parameters of each pico-eNodeB can be configured according to the interference level from adjacent eNodeBs. However, their study mainly focuses on improving the cell area capacity rather than addressing the random access and the congestion problem.

In this paper, we propose a new Pico-CRE (PCRE) scheme with dynamic RACH resource allocation to improve the average random access efficiency and the average access delay of

M2M and H2H UEs. Each pico-eNodeB measures the number of accessed UEs, and compares it with the estimated number of UEs. After the comparison, the pico-eNodeB adjusts its cell range to reduce the gap between the measured number of UEs trying to access and the estimated number of UEs trying to access, and allocates the RACH resource to cover the expected number of UEs trying to access in the expanded cell range appropriately. By the proposed scheme, the random access congestion problem can be addressed in HetNets. We show the performance improvement by our proposed scheme through simulations.

This paper is organized as follows. We describe the proposed PCRE with dynamic RACH resource allocation in Section II. In Section III, we evaluate the performance of the proposed scheme via simulations. Finally, conclusion is presented in Section IV.

## II. PROPOSED PICO-CELL RANGE EXPANSION WITH DYNAMIC RACH RESOURCE ALLOCATION

The congestion problem may occur when massive UEs try random access simultaneously in a network. Especially, in HetNets, the macro-eNodeB and pico-eNodeBs may utilize the same Physical Random Access Channel (PRACH) in the same time since the macro-eNodeB and pico-eNodeBs work independently for communication services. When macro UEs are near a pico-cell and they transmit their preambles to the macro-eNodeB on the same PRACH with pico UEs, the pico-eNodeB can overhear the preambles. Although the preambles are not for the pico-eNodeB, the pico-eNodeB misunderstands that the preambles are transmitted by their pico UEs, and thus allocates unnecessary RACH resource, i.e., Random Access Opportunities (RAOs). Therefore, the performance of random access is degraded in pico-cells. So, we aim to solve the random access congestion problem by adjusting the pico-cell's range and diverting random access trials of macro UEs to pico-cells.

In a HetNet, UEs generally try to establish a Radio Resource Control (RRC) connection with the eNodeB supporting the largest Reference Signal Received Power (RSRP). The number of UEs served by a pico-cell can be limited by the relatively low transmission power of a pico-eNodeB. The adoption of an efficient CRE scheme can increase the number of UEs served by pico-cells. Fig. 1 represents an example of our proposed scheme. Since pico-eNodeBs PeNB1, PeNB3, and PeNB4 suffer from interference by macro UEs near their cells due to the shared preamble set in Fig. 1 (a), the proposed CRE scheme is applied according to the number of UEs. In order to accommodate more UEs to a pico-eNodeB, the pico-eNodeB expands the cell range by controlling the transmission power. Then pico-eNodeBs also allocate RACH resource dynamically. As shown in Fig. 1 (b), PeNB1, PeNB3, and PeNB4 allocate more RACH resource because the number of UEs, which migrate from the macro-cell to the pico-cell, is increased. For the macro-cell, less RACH resource is needed due to the reduced number of macro UEs by the proposed scheme. The proposed scheme can efficiently distribute more macro UEs near a pico-cell to the pico-cell and allocate RACH resource appropriately. The dispersion of UEs from a macro-cell to a pico-cell can solve the congestion problem in a macro-cell due to the reduction of the number of access UEs.

Now, we describe the proposed PCRE with the dynamic

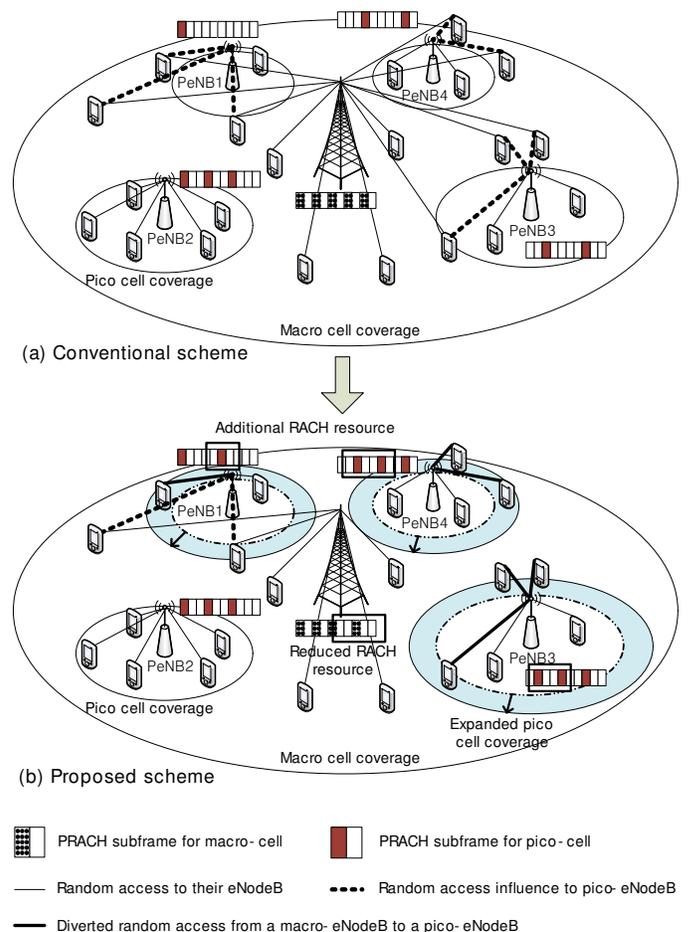


Figure 1. Random access in macro- and pico-cell networks (a) conventional scheme (b) proposed scheme.

RACH resource allocation algorithm. In our algorithm, a pico-eNodeB measures the idle RAOs in the current frame to estimate the number of UEs trying to access in the pico-cell area. Then, the pico-eNodeB calculates the estimated number of successfully accessed UEs and compares it with the measured number of successfully accessed UEs in the frame. If its gap is greater than a threshold, i.e., there are a number of macro UEs around the cell boundary, the pico-eNodeB expands its cell range to cover the macro UEs and to allocate an appropriate number of RAOs for the UEs in the expanded region.

At the  $i$ th frame, to estimate the number of UEs trying to access, the pico-eNodeB measures the number of idle RAOs ( $\tilde{L}_{idle\_pico,i}$ ). The probability of idle RAOs ( $\hat{P}_{idle\_pico,i}$ ) in the  $i$ th frame can be estimated as

$$\hat{P}_{idle\_pico,i} = \frac{\tilde{L}_{idle\_pico,i}}{L_{pico,i}}, \quad (1)$$

where  $L_{pico,i}$  is the number of allocated RAOs in the  $i$ th frame. Since the probability of idle RAOs can be computed as

$$P_{idle\_pico,i} = \left(1 - \frac{1}{L_{pico,i}}\right)^{N_{PUE,i}}, \quad (2)$$

where  $N_{PUE,i}$  is the number of UEs, we can get the estimated number of UEs  $\hat{N}_{PUE,i}$  by letting (1) to be equal to (2) such

that

$$\hat{N}_{PUE,i} = \frac{\log \hat{P}_{idle\_pico,i}}{\log \left( 1 - \frac{1}{L_{pico,i}} \right)}. \quad (3)$$

From  $\hat{N}_{PUE,i}$ , the probability of successful RAOs  $\hat{P}_{succ,i}$  in the  $i$ th frame is estimated as

$$\hat{P}_{succ,i} = \hat{N}_{PUE,i} \frac{1}{L_{pico,i}} \left( 1 - \frac{1}{L_{pico,i}} \right)^{\hat{N}_{PUE,i}-1}. \quad (4)$$

Then, the pico-eNodeB estimates the number of successfully accessed UEs  $\hat{N}_{succ\_pico,i}$  in the  $i$ th frame as

$$\hat{N}_{succ\_pico,i} = \hat{P}_{succ,i} \times L_{pico,i}. \quad (5)$$

In fact, the pico-eNodeB can know the actual number of successfully accessed UEs  $\tilde{N}_{succ\_pico,i}$  in the  $i$ th frame. So, by comparing  $\hat{N}_{succ\_pico,i}$  with  $\tilde{N}_{succ\_pico,i}$ , the pico-eNodeB recognizes the existence of macro UEs adjacent to its service coverage.

If the gap of the estimated number of successfully accessed UEs and that of the actual UEs is larger than the threshold  $N_{thr}$ , it conducts its cell range expansion to include the neighboring macro UEs. To expand its cell coverage, the pico-eNodeB should decide the amount of transmission power increment. First, the path-loss in the macro-cell  $PL_{macro,i}$  and the path-loss in the pico-cell  $PL_{pico,i}$  can be defined by the case model 1 in [13].

$$PL_{macro,i} = 128.1 + 37.6 \log (d_{macro,i}[km]), \quad (6)$$

$$PL_{pico,i} = 140.7 + 36.7 \log (d_{pico,i}[km]), \quad (7)$$

where  $d_{macro,i}$  and  $d_{pico,i}$  are the distance between the macro-eNodeB and a virtual UE, and the distance between the pico-eNodeB and a virtual UE, respectively. We define the RSRP (from the macro-eNodeB) at a UE as  $RSRP_{macro,i}$  and the RSRP (from the pico-eNodeB) at a UE as  $RSRP_{pico,i}$ .

$$RSRP_{macro,i} = 10 \log (P_{tx\_macro,i}) - PL_{macro,i}, \quad (8)$$

$$RSRP_{pico,i} = 10 \log (P_{tx\_pico,i}) - PL_{pico,i}, \quad (9)$$

where  $P_{tx\_macro,i}$  and  $P_{tx\_pico,i}$  are the transmission powers of the macro- and of the pico-eNodeB, respectively.

To expand its cell range, the pico-eNodeB needs to find its current cell coverage distance  $D_{pico,i}^*$ . When  $RSRP_{macro,i}$  and  $RSRP_{pico,i}$  are equal, the pico-cell coverage distance can be obtained.

$$RSRP_{macro,i} = RSRP_{pico,i}. \quad (10)$$

However, there may be many solutions for the coverage distance. To find the shortest one, the distance  $D_{mtop,i}$  between the macro-eNodeB and the pico-eNodeB is used.

$$D_{mtop,i} = \sqrt{x_{macro,i}^2 + y_{macro,i}^2}, \quad (11)$$

where  $(x_{macro,i}, y_{macro,i})$  is the position of the macro-eNodeB assuming the pico-eNodeB is at  $(0, 0)$ . When a virtual UE is in a straight line from the macro-eNodeB to the pico-eNodeB,  $D_{mtop,i}$  can be

$$D_{mtop,i} = d_{macro,i} + d_{pico,i}. \quad (12)$$

From (10) and (12), the solution  $D_{pico,i}^*$  for  $d_{pico,i}$  can be obtained.

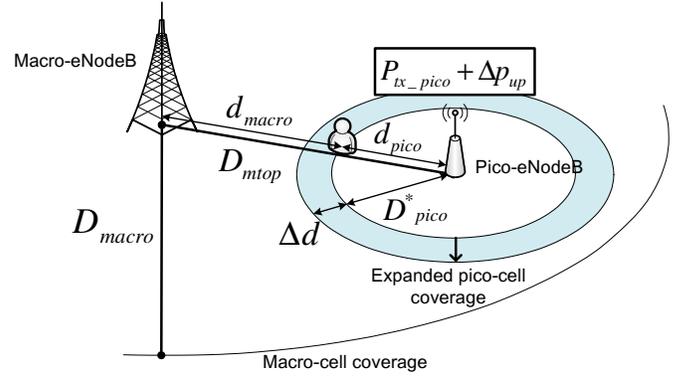


Figure 2. An example of pico-cell range expansion.

Once the pico-eNodeB obtains its current radius  $D_{pico,i}^*$ , it calculates the amount of the transmission power increment  $\Delta p_{up,i}$  to expand the cell range. Let  $N_{TOTAL,i}$  be the maximum number of UEs served in a cell and  $D_{macro,i}$  be the radius of the macro-cell. From  $\hat{N}_{PUE,i}$ ,  $N_{TOTAL,i}$ ,  $D_{macro,i}$ ,  $D_{pico,i}^*$ , and the assumption that UEs are uniformly distributed in a cell, the additionally required pico-cell range expansion  $\Delta d_i$  can be calculated as

$$\hat{N}_{PUE,i} : (D_{pico,i}^* + \Delta d_i)^2 = N_{TOTAL,i} : D_{macro,i}^2. \quad (13)$$

After the cell range expansion, the new RSRP of a virtual pico UE at the edge area of the pico-cell on a straight line from the macro-eNodeB to the pico-eNodeB is

$$RSRP'_{pico,i} = 10 \log (P_{tx\_pico,i} + \Delta p_{up,i}) - PL'_{pico,i}, \quad (14)$$

where

$$PL'_{pico,i} = 140.7 + 36.7 \log (D_{pico,i}^* + \Delta d_i). \quad (15)$$

The new RSRP of the virtual pico UE (from the macro-eNodeB) is

$$RSRP'_{macro,i} = 10 \log (P_{tx\_macro,i}) - PL'_{macro,i}, \quad (16)$$

where

$$PL'_{macro,i} = 128.1 + 37.6 \log (D_{mtop,i} - (D_{pico,i}^* + \Delta d_i)). \quad (17)$$

For the new  $RSRP'_{pico,i}$  to be the same as the new  $RSRP'_{macro,i}$ , the amount of transmission power increment  $\Delta p_{up,i}$  can be obtained. Fig. 2 shows an example of pico-cell range expansion.

$$RSRP'_{pico,i} = RSRP'_{macro,i}. \quad (18)$$

The pico-eNodeB determines the transmission power no more than the maximum transmission power  $P_{max}$  and no less than the minimum transmission power  $P_{min}$ . So, the transmission power in the next frame for the pico-cell is updated as

$$P_{tx\_pico,i+1} = \max (\min (P_{tx\_pico,i} + \Delta p_{up,i}, P_{max}), P_{min}). \quad (19)$$

On the other hand, if the difference between the estimated number of accessed UEs and that of the measured UEs is lower than the threshold  $N_{thr}$ , it indicates that there are not

many macro UEs contending in the pico-eNodeB. Thus, the transmission power for the next frame in the pico-cell is not updated.

In order to get the optimal number of RACH resource for dynamic RACH resource allocation of pico-cells, we use the random access efficiency of the system. We define the random access efficiency  $RA_{eff,i}$  as

$$RA_{eff,i} = \hat{N}_{PUE,i} \frac{1}{L_{pico,i}} \left( 1 - \frac{1}{L_{pico,i}} \right)^{\hat{N}_{PUE,i}-1}. \quad (20)$$

The random access efficiency indicates the average success rate of UEs trials in a frame. The maximum random access efficiency is obtained by differentiation of (20). Then, we can get

$$L_{pico,i+1} = \hat{N}_{PUE,i}. \quad (21)$$

The dynamic RACH resource allocation phase is performed by updating the number of RAOs for the pico-cell as the estimated number of pico UEs  $\hat{N}_{PUE,i}$ . The number of RAOs is limited to the maximum RAOs  $L_{max}$ . Then the number of RAOs in the next frame for the pico-cell is

$$L_{pico,i+1} = \min \left( \hat{N}_{PUE,i}, L_{max} \right). \quad (22)$$

### III. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our proposed PCRE with dynamic RACH resource allocation scheme in HetNets. The parameters used in the simulations are shown in Table I. We use the simulation parameters in [13]. We consider four pico-eNodeBs and UEs are uniformly distributed in a macro-cell. In the conventional scheme, UEs randomly access to their eNodeBs with fixed RAOs. For the path-loss model, all macro- and pico-eNodeBs are outdoor, and 3GPP TR 36.814 urban macro and pico path-loss model is employed [13].

Fig. 3 and Fig. 4 show the average random access efficiency of a macro-cell and a pico-cell. In Fig. 3, as the arrival rate of UEs trying to access becomes larger, the average random access efficiency of the conventional scheme with fixed RAOs in the macro-cell slowly increases and decreases after the certain arrival rate. When the arrival rate is small, the number of RAOs is more than the number of access trials since the arrival rate is small. Then the random access efficiency increases as the arrival rate increases. After that, the random access efficiency decreases since users have insufficient number of RAOs at high arrival rate. However, the average random access efficiency of the proposed scheme in the macro-cell is almost 36%, i.e., the

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Radius of a macro cell ( $D_{macro}$ )	288 m
Max. Macro-eNodeB Tx Power ( $P_{tx\_macro}$ )	20 W
Max. Pico-eNodeB Tx Power ( $P_{tx\_pico}$ )	1 W
Max. UE Tx. Power	200 mW
Arrival rate of UEs	1000 to 35000
Max. no. of preamble transmissions	10
Max. no. of HARQ transmissions	5
Max. no. of preambles	54

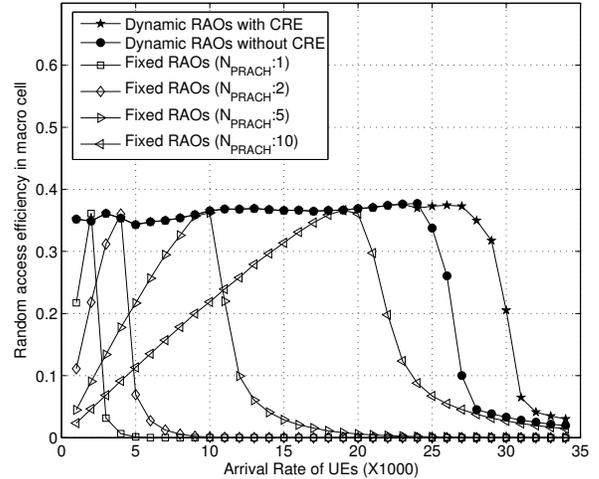


Figure 3. Average random access efficiency in a macro-cell.

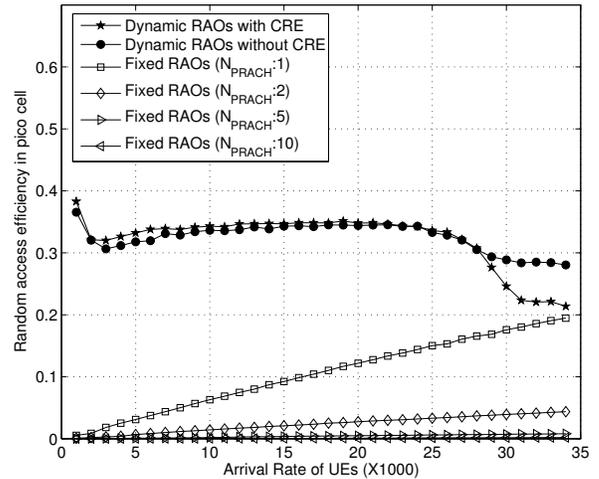


Figure 4. Average random access efficiency in a pico-cell.

maximum efficiency, since the RACH resource in the cell is allocated dynamically according to the number of UEs. After a large arrival rate, collisions may occur and the average random access efficiency decreases. When each pico-eNodeB employs the CRE scheme with dynamic RAOs, some of macro UEs near the pico-cell can access to the pico-eNodeB. Thus, when the arrival rate of UEs trying to access is large, the macro-eNodeB achieves better performance in terms of the average random access efficiency. The simulation result shows that the average random access efficiency of the proposed scheme with CRE in the macro cell is improved by 50% over that of the conventional scheme with fixed RAOs.

Fig. 4 shows the average random access efficiency of a pico-cell. In the conventional scheme without CRE, the number of UEs serviced by a pico-cell is relatively small. When many PRACH subframes for the pico-cell are assigned in a frame, there may be unused RAOs. So, the random access efficiency increases as the arrival rate of UEs increases. In the proposed

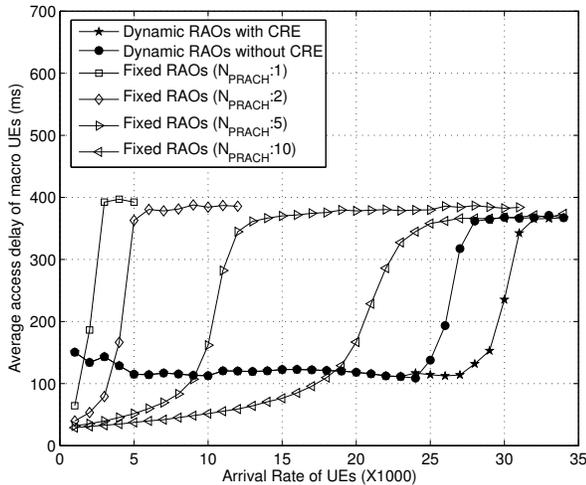


Figure 5. Average random access delay in a macro-cell.

scheme, the pico-eNodeB expands the cell range and allocates the RACH resource dynamically depending on the number of pico UEs. The average random access efficiency of the proposed scheme in the pico-cell reaches to 36%, which is similar to the average random access efficiency of a macro-cell. However, for a large arrival rate, performance may degrade since the number of macro UEs, which are allocated the same preambles with pico UEs in the near pico-cell, increases for massive UEs.

Fig. 5 and Fig. 6 show the average random access delay of UEs in a macro-cell and a pico-cell, respectively. In Fig. 5, at low arrival rate, the number of available RAOs for users may be sufficient in fixed RAOs. So, the random access delay of the conventional scheme is lower than that of the proposed scheme. As the arrival rate of UEs trying to access becomes larger, the average random access delay of the conventional scheme in the macro-cell dramatically increases. However, when the proposed scheme is applied, the average random access delay remains small for a wide range of arrival rates and starts to increase at a large arrival rate of UEs compared to the average random access delay of the conventional scheme. Collisions may occur frequently due to insufficient RAOs in the conventional scheme. However, the macro-eNodeB applying the proposed scheme allocates RAOs appropriately according to the number of macro UEs, and it results in high random access efficiency. Thus, the proposed scheme achieves better performance.

Fig. 6 shows the average random access delay of UEs in a pico-cell. For fixed RAOs, there may be unused RAOs at the low arrival rate. Then, the random access delay of the conventional scheme is lower than that of the proposed scheme. The simulation result shows that the proposed scheme with CRE can tolerate the congestion problem of UEs compared to the proposed scheme without CRE. When the arrival rate of UEs trying to access is extremely high, the average random access delay for the pico-cell may be degraded compared to that of the conventional scheme with fixed RAOs. When there are a number of UEs, PRACH subframes for the macro-cell are allocated in many of the subframes. Then pico UEs, which

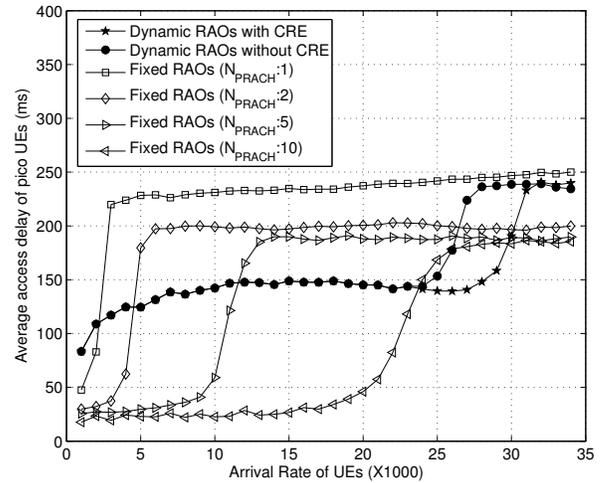


Figure 6. Average random access delay in a pico-cell.

transmit preambles on a PRACH, can be influenced by adjacent macro UEs, which transmit the same preambles on the same PRACH as the pico UEs. Nevertheless, the average random access delay of the proposed scheme is shown to be improved in general.

#### IV. CONCLUSION

In this paper, we aimed at studying the performance improvement of random access in HetNets. Especially, we consider the feature of M2M communications in which a very large number of MTCs can exist in a cell and massive MTCs contend simultaneously. Then, the random access congestion problem can occur due to the concurrent and massive random access trials of MTCs. To solve the problem, we have proposed the PCRE with dynamic RACH resource allocation. The proposed scheme can attract more MTCs from a macro-cell to pico-cells by controlling the pico-cell's coverage area. The pico-eNodeBs then allocate appropriate RACH resource to cover the attracted MTCs in the expanded cell range. We have shown that the proposed scheme can improve the average random access efficiency and the average random access delay of UEs compared to the conventional scheme with fixed random access resource via simulations.

#### ACKNOWLEDGMENT

This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning (20100020729), and was supported by Basic Science Research Program through the NRF funded by the Ministry of Education(NRF-2010-0020210).

#### REFERENCES

- [1] 3GPP TR 23.888 V11.0.0, "System Improvements for Machine-Type Communications," Sep. 2012.
- [2] 3GPP TR 37.868 V11.0.0, "Study on RAN Improvements for Machine-Type Communications," Sep. 2011.

- [3] K.-D. Lee, S. Kim, and B. Yi, "Throughput Comparison of Random Access Methods for M2M Service over LTE Networks,," in Proc. of IEEE Global Communications Conference (GLOBECOM), Dec. 2011, pp. 373-377.
- [4] O. Yilmaz, J. Hamalainen, and S. Hamalainen, "Self-optimization of Random Access Channel in 3GPP LTE," in Proc. of Wireless Communications and Mobile Computing Conference (IWCMC), Jul. 2011, pp. 1397-1401.
- [5] M. Amirijoo, P. Frenger, F. Gunnarsson, J. Moe, and K. Zetterberg, "On Self-optimization of the Random Access Procedure in 3G Long Term Evolution,," in Proc. of International Symposium on Integrated Network Management-Workshops (IFIP), Jun. 2009, pp. 177-184.
- [6] A. Damnjanovic et al. "A Survey on 3GPP Heterogeneous Networks,," IEEE Wireless Communications, vol. 18, no. 3, pp. 10-21, Jun. 2011.
- [7] A. Ghosh et al. "Heterogeneous Cellular Networks: From Theory to Practice,," IEEE Communications Magazine, vol. 50, no. 6, pp. 54-64, Jun. 2012.
- [8] Motorola, "On Range Extension in Open-Access Heterogeneous Networks (R1-103181),," 3GPP TSG RAN WG1 Meeting-61, May 2010.
- [9] D. Luo, B. Li, and D. Yang, "Performance Evaluation with Range Expansion for Heterogeneous Networks,," in Proc. of Vehicular Technology Conference (VTC), Sep. 2011, pp. 1-5.
- [10] I. Guvenc, "Capacity and Fairness Analysis of Heterogeneous Networks with Range Expansion and Interference Coordination,," IEEE Communications Letters, vol. 15, no. 10, pp. 1084-1087, Oct. 2011.
- [11] A. Morimoto, N. Miki, H. Ishii, and D. Nishikawa, "Investigation on Transmission Power Control in Heterogeneous Network Employing Cell Range Expansion for LTE-Advanced Uplink,," in Proc. of Conference on European Wireless (EW), Apr. 2012, pp. 1-6.
- [12] J. Liu et al. "Uplink power control and interference coordination for heterogeneous network,," in Proc. of International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Sep. 2012, pp. 519-523.
- [13] 3GPP TR 36.814 V9.0.0, "Further Advancements for E-UTRA Physical Layer Aspects,," Mar. 2010.

# Insights into Cellular Networks: Anatomy of Traffic Profiles

Wieslawa Wajda  
Bell Labs  
Alcatel-Lucent Deutschland AG  
Stuttgart, Germany  
Wieslawa.Wajda@alcatel-lucent.com

**Abstract**—Analysis of real telecommunication data can provide a fine grasp and good understanding of user behavior dynamics and can be used to improve network management strategies. Following this approach, we analyze actual wireless traffic and provide basis for a suitable network model with a given network traffic profile. Unlike previous approaches, we introduce criteria for intelligent network resources utilization and define metrics to discover optimization potential of wireless system architectures. Our traffic fluctuation analysis is of interest for traffic load prediction and can boost BS switching time strategies. Thus, this study provides hints into hardware realization of a future base station and basis for development of efficient network architectures.

**Keywords**—base station; traffic profile; traffic fluctuation; cell categorization; efficiency index.

## I. INTRODUCTION

In the existing literature, traffic observation and resource utilization in real networks were analyzed in different context. Investigation how efficiently radio resources are used by different subscribers and applications in macro Base Stations (BS) were done by Paul et al. [1]. The observations are related to traffic spread, mobility and efficiency in connection to subscriber pricing, protocol design, spectrum allocation and energy savings. Gender-based traffic characterizations based on campus traffic data were done by Kumar and Helmy [2]. Much research has been devoted to traffic classification models and methods in context of network security and management. In this context, the analysis concerned different kind of traffic generated by single application flow and the real-time traffic data were used for verification of proposed methods by Khalife et al. and by Won et al. [3][4].

In this paper, we agree with Jiang et al. [7] and are convinced that only the analysis of real traffic measurements give a deeper understanding into the network and thus provides a basis for a suitable network model with a given network traffic profile. Unlike the mentioned approaches, we introduce criteria for intelligent network resources utilization and define metrics to discover optimization potential of wireless system architectures.

### A. Contributions

Focusing on actual network data, we:

- Introduce criteria of BSs capacity utilization in a capital city centre, so called *Dense Urban Centre* (DUC) and in an extended circle of the capital city centre, so called *Greater Dense Urban* (GDU) area.
- Define a metric to assess optimization potential for energy saving in the specific areas.
- Define BS categories with regard to their traffic profiles.
- Provide details of traffic fluctuation.
- Contribute to better understanding of resource sharing.

The results of this study give hints for hardware (HW) realization of future BS, contribute to resource optimization of the BS and provide basis for development of future efficient network architectures and adequate methods for resource management.

### B. Analysed data

Data at our disposal have been provided by a single European operator in an European city. The analyzed traffic data cover BS of different power and capacity for dense urban centre and greater dense urban areas. The traffic data has been sampled for down- (High Speed Downlink Packet Access) and for uplink (High Speed Uplink Packet Access) in 900 sec (15 min.) intervals over a period of a week.

Given intervals of 15 min. make traffic fluctuation of shorter time invisible. Due to the fact that our analysis is oriented toward system utilization, this represents no real disadvantage and does not reduce the quality of the data. Furthermore, short fluctuations are typically resolved with the help of store and forward technology and are not relevant for analyzed mechanisms.

The analyzed system covers about 400 pico, micro and macro UMTS-cells in dense urban area, as well as over 900 pico, micro and macro UMTS-cells in greater dense urban area. In the analyzed area, macro-cells were deployed with high mounted antenna at ca. 24 – 44 m and with 15 – 20 W transmission power. Micro-cells were deployed with ca. 0.85W and with low mounted antennas at ca. 5 - 6 m. In addition, some pico-cells were deployed.

C. Structure of the paper

This paper is organized as follows. Section II presents load dynamics in analyzed network in opposite to load dynamics in single cells presented in Section III. Section IV gives a general look at cell load, especially in DUC. Section V shows cells utilization in relation to their capacity and proposes metric discovering potential for energy efficiency in the cells. Based on the study, cells are characterized regarding their traffic profiles in Section VI. Section VII deals with traffic fluctuation and contributes to understanding of its basic structural elements. Hints for BS architecture and its possible advantage for energy saving in the network are presented in Section VIII. Finally, the paper is concluded in Section IX.

II. GLOBAL TRAFFIC PROFILE

Global traffic of the analyzed cell population was calculated as shown in (1).

$$F(t) = \sum_i f_i(t) \tag{1}$$

with  $f_i(t)$  being the cells individual relative traffic data. Diagrams of  $F(t)$  for DUC and GDU areas show sinusoid like shape and periodicity over a week (see Figure 1 and Figure 2). The fluctuation dynamic is within the range of about 13 dB. In the case of a DUC area, highest traffic load appears between 10:00 a.m. to 05:00 p.m., while in the GDU area the high traffic is by about 02:00 p.m. and decreases up to 10:00 p.m.

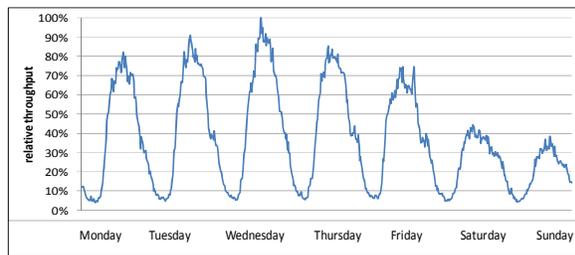


Figure 1. DUC weekly traffic profile.

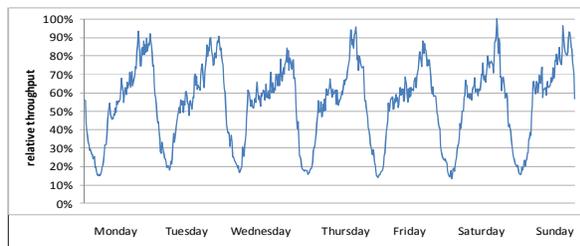


Figure 2. GDU traffic profile over a week.

Also, there is a difference on weekends between traffic density in the DUC and GDU areas. The hypothesis can be made that this difference results from population movement. Where the population in the GDU is dominated by inhabitants and therefore steady, the DUC has more office buildings and therefore its population drops on evenings and on weekends.

III. LOCAL TRAFFIC PROFILE

In some cases, an aggregated traffic profile is used by Gonzales et al. and Ambrosy et al. [5][6], but the load profiles of individual cells turn out to behave very differently from shown global profiles, as presented in the Figure 3.

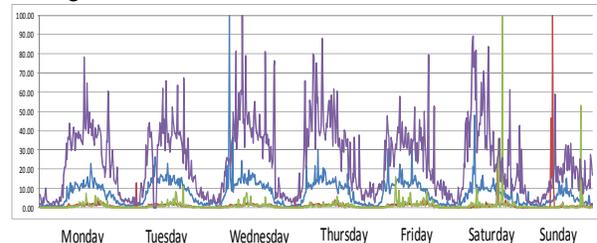


Figure 3. Traffic examples of cells in different areas.

Cell load profiles are different, even in cells of the same BS. They are extremely volatile and aperiodic. Depending on the BS location and served cell, load profiles can have very different utilization. Also, they can show some periodicity or be unpredictable. Summing up, elaboration and assessment of optimization measures for cells must be based on well understanding of a cell behavior and its traffic profile.

The considerable load dynamic justifies an optimization effort of BSs in analyzed areas. Some optimization examples can be: switching connections or hardware modules off, changing the transmission scheme, adapting the transmission speed of the backbone by slowing down clock frequency, selling released transmission resources to third party, while tuning off reserved resources can be controlled by actual time and date. The algorithms can be reinforced by self-learning procedures, to facilitate correct system reaction in case of nonstandard events and events not covered by a core algorithm.

IV. CELL LOAD

A general look at the traffic profiles of individual cells lets expect low cell utilizations because there are a few traffic peaks and the traffic load is low (see Figure 3). The Cumulated Density Function (CDF) calculated for all cells in DUC over the analyzed time period confirms this thought. Figure 4 shows high probability for low traffic in all analyzed DUC cells.

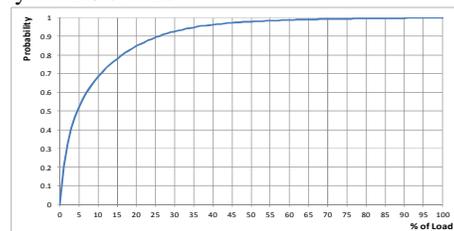


Figure 4. CDF of cells load in DUC.

According to this and to the fact that cell load profiles are greatly different, a closer look at their distribution, utilization, traffic oscillation and dynamics is essential for

support decisions on design, deployment and operation of single BS, as well as of a whole wireless network.

V. UTILIZATION OF BS CAPACITY

First, we define what the BS capacity and the BS utilization is. The BS capacity is defined as the max. data traffic the BS can handle at once, which is, based on data we have, the highest absolute data rate of a particular BS in a week. Therefore, if we speak of 100% of BS capacity, it means its highest data rate in the measured time period. Furthermore, the resource utilization is the throughput the BS processes at a given time. For example, following the green solid line in Figure 5, about 90% of analyzed BSs (on the Y axis) in the GDU area have 90% traffic load by only up to 25% of its capacity (X axis). Similar, according to the violet solid line 90% of BSs in the GDU and according to the violet dotted line 60% of BSs in the DUC area are utilized by less than 20% for 85% measured traffic load. Finally, it can be recognized that the processed traffic of BSs in both regions does not exceed 50% of their capacity.

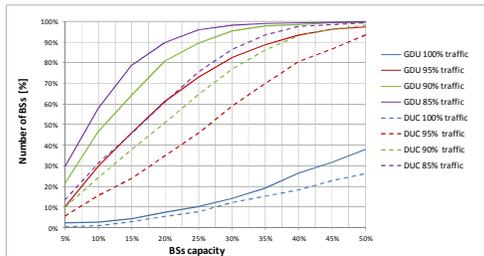


Figure 5. BSs utilization in DUC and GDU areas.

According Figure 5, it could be assert that the overall utilization of BS in the GDU is about 25% lower as those in the DUC area, and in both areas there is a big potential for improvements of the utilization factor.

A. Load distribution

Even though the load distribution in the time scale of a whole system is predictable, and has a shape with 24 h oscillation period influenced by a day of week as shown in Figure 1 and Figure 2, the load shapes of individual BSs are different and seem not to be bound to a periodic event. Figure 6 shows 5 different shapes for various cells in the GDU area between 07:00 a.m. and 06:00 p.m. on one day. We can see no similarity between them.

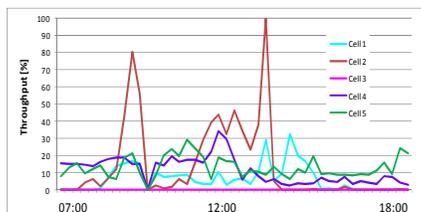


Figure 6. Load profiles of different BS between 07:00 a.m. - 06:00 p.m.

The CDF function of the traffic variance for all BSs in DUC area shows that more than 90% of evaluated BS population has a variance less than a half of the calculated

variance max value (see Figure 7). It lets us assume that the spreading of values, i.e., the most oscillation of a load shape is low to middle. This in turn has an impact on BS construction and deployment planning.

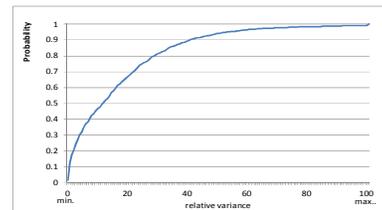


Figure 7. CDF function of load variance in DUC.

The above confirms also the Figure 8. It shows the view on the traffic load of a number of cells in the DUC area, where areas with low traffic at night and differences in the traffic of particular cells can be well recognized. Very interesting is also to see, that the traffic of a single cell is not steady but shaped by a short rising peaks which seems to be independent one from another.

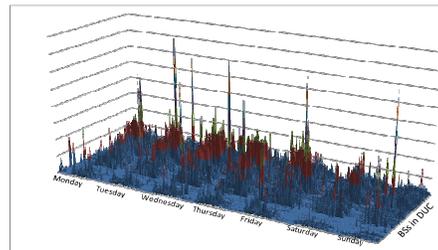


Figure 8. Traffic load over a week of exemplary BSs in DUC.

The discrete cross correlation function allows a closer look at the traffic divergence. The cross correlation function for two cells is defined as follow and has been used correspondingly for all cell pairs:

$$R = \sum_{-\tau}^{\tau} \frac{\frac{1}{T} \int_{-\infty}^{\infty} x(t) y(t + \tau) dt}{\sum_{-\tau}^{\tau} \frac{1}{T} \int_{-\infty}^{\infty} x(t) x(t + \tau) dt} \quad (2)$$

The correlation value of compared traffic profiles of cell x and cell y is shown in relation to the autocorrelation function of the source cell x where  $\tau$  is the correlation window time and T is the measured time period (one week).

Figure 9 shows the R values for all BSs in the DUC area for  $\tau$  in the range of  $\pm 1$  hour.

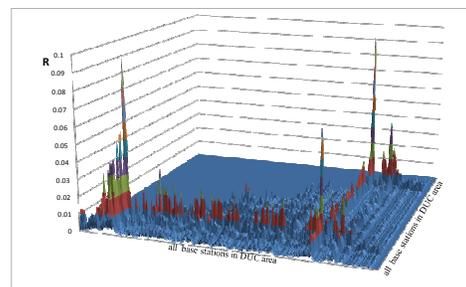


Figure 9. Cross correlation of cells in DUC area.

It can be seen that the  $R$  values are low, so there is no significant correlation of BSs traffic in the region. However, the high peaks show that there is some correlation between BSs traffic in the neighborhood.

### B. Metric to assess the potential for energy saving

Observed low BSs utilization and low correlation value of their traffic profiles let expect reasonable saving potential for energy and for HW costs. Therefore, some calculations have been made to identify the energy saving potential more exactly. For this reason, a metric was defined which express an Energy Saving Potential Index (ESPI) based on a histogram of load values between 0% - 100% load with 5% steps (bins).

The value of the ESPI is *zero* if there is no potential for energy saving and *one* if the energy saving potential is high.

$$ESPI = \frac{\sum_i n_i (1 - \alpha_i)}{\sum_i n_i} \quad (3)$$

Here,  $n$  is the number of measures for a given load bin,  $\alpha$  is the load fraction, e.g., 0% - 100%, and  $i$  the number of defined load bins.

Applying the ESPI metric to the DUC area and by consideration of 20 bins, each every 5% traffic load (see Figure 10) it can be identified, that the lowest ESPI value is 0.70. In total, 92% cells have saving potential between 0.81 - 0.95 and 60% cells had saving potential between 0.90 - 0.95.

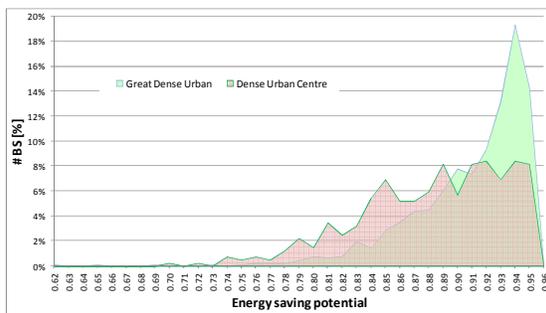


Figure 10. ESPI for DUC and GDU.

For the GDU area, most cells have saving potential index between 0.76 - 0.95 whereby 71% cells have saving potential over 0.9.

## VI. BASE STATION CATEGORIES WITH REGARD TO THEIR TRAFFIC PROFILES

It was expected that the best way to save energy is to design a base station optimized to its traffic profile and further to adopt the network architecture to recognized traffic categories. Therefore, we analyzed the individual BS traffic to categorize traffic profiles of individual cells. As shown later in this section recognized categories of traffic profiles have corresponding energy saving potential. The BSs were analyzed regarding their traffic load during working days, emerged peaks, traffic load on

the weekend and ESPI. At last, the studied traffic profiles were categorized into three generic traffic profiles in the DUC and two generic traffic profiles in the GDU area what is presented in Table I.

In the analysis, it was assumed that the highest throughput peak in the observed time embodies the maximum load or maximum resource utilization of the given BS, whereby that maximum resource utilization is typically in the range of 60% for macro BSs in dense urban areas as found in EARTH [10].

TABLE I. BS CATEGORIES

Dense Urban Centre				
BS Type	Traffic load	Peaks	Traffic on weekend	ESPI
1	Very low, Mostly < 5% of BS capacity	Random, sparse	Almost never	Typically 0.95
2	Medium to high load between 09:00a.m. - 05:00 p.m. 60% of the traffic load $\approx$ 5% of the BS capacity	Less peaks up to max. load	Lower than during the week	Typically 0.85 - 0.89
3	High load, during most of the 24 h; 95% of the traffic load $\approx$ 60% of BS capacity	Many	Lower or similar to traffic during the working days	Typically 0.74
Greater Dense Urban				
BS Type	Traffic load	Peaks	Traffic on weekend	ESPI
1	Low; up to 10% of BS capacity	Random, few high peaks	Yes	Typically 0.92
2	Mostly low traffic during a day and higher traffic in the night. The most traffic is up to 35% - 40% of the BS capacity	Many	Similar to traffic during the working days	Typically 0.70 - 0.80

As shown in Table I, recognized traffic profiles categories have corresponding energy saving potential.

## VII. FLUX

The analysis of traffic profile requires understanding of its basic structural elements. These elements can be described by means of the term *flux* that corresponds to envelope of the traffic function with different levels of aggregation. The envelope of traffic function, i.e., a single

flux, is calculated as a linear approximation of consecutive growing and falling traffic profile sections with symmetry defined by constant  $\rho$ .

In the calculations, the  $\rho$  value was set to the level at which the fluctuation integrated down to the general shape can be simply compensated by hardware measures and would not influence traffic control algorithms.

The formulas for flux calculation are defined below.

a) *Rising edge*

$$t_i < t < t_k, \quad f'(t) > 0 \quad (4)$$

b) *Falling edge*

$$t_{k+1} < t < t_m, \quad f'(t) < 0 \quad (5)$$

c) *Flux criterion*

$$abs \left| \frac{f(t_i) - f(t_m)}{\max_i f(t) - \min_i f(t)} \right| < \rho \quad (6)$$

d) *Flux definition:*

$$F(x) = f(t_i) + n \frac{(f(t_k) - f(t_i))}{t_k - t_i} \quad (7)$$

for  $t_i \leq n \leq t_k$

and

$$F(x) = f(t_k) - n \frac{(f(t_k) - f(t_i))}{t_m - t_k} \quad (8)$$

where  $t_{k+1} \leq n \leq t_m$

where the  $t_i$ ,  $t_k$  and  $t_m$  are described as in Figure 11.

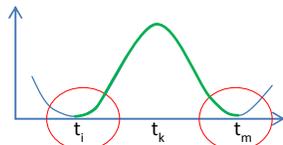


Figure 11. Parameters for flux definition.

An example for a traffic load fluctuation and the envelope for a given  $\rho$  is shown in Figure 12.

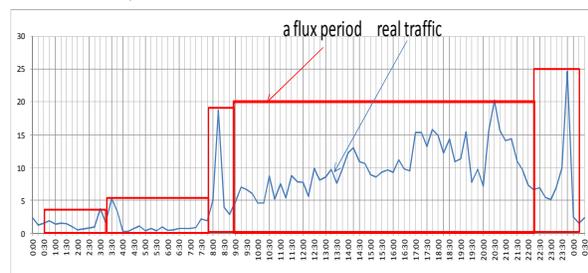


Figure 12. Traffic example and Fluxes.

The following flux parameters were calculated: flux duration, flux delta, i.e., the difference between the lowest and highest value of the flux, and slew rate.

The CDF functions for calculated parameters for the GDU area are shown in Figures 14, 15 and 16. Please note that traffic data has been sampled with 900 sec intervals. The results of flux duration calculation (see Figure 13) shows that traffic changes are relatively slow. About 50% of

analyzed fluxes last up to 1 hour and 90% last up to 2, 5 hours.

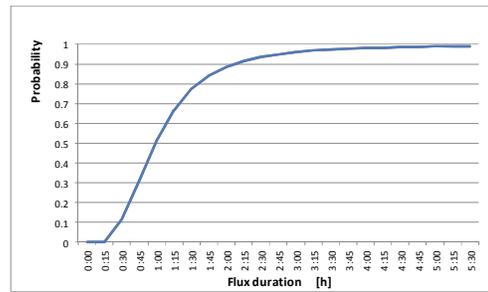


Figure 13. CDF of flux duration.

Figure 14 shows that the amplitude of 50% fluxes is up to 5% of max. load and 90% fluxes have the amplitude lower than 28% of max. load. It means that fluctuations with high amplitude are relatively seldom.

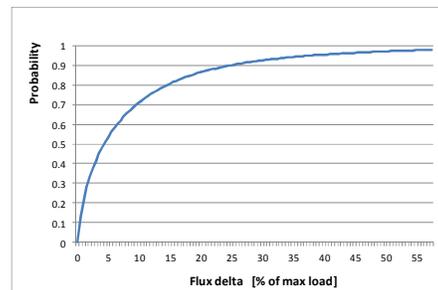


Figure 14. Flux delta.

Slew rate of rising and falling edges for the almost all fluxes is nearly the same (see Figure 15).

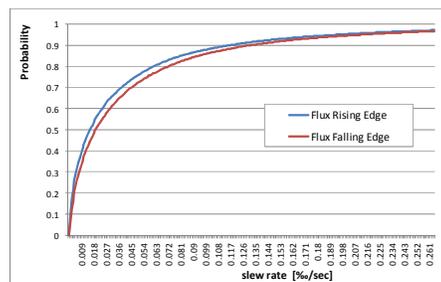


Figure 15. Flux edges.

The slew rate of low dynamics fluxes (with delta up to 50 % of max load) has the value of 0.5% of their peak value per second. This means that the slew time of the fluxes lasts about 30 minutes.

The 90% high dynamic fluxes (with delta higher than 50 % of the max load) have the slew rate of rising and falling edge by about 0.65% per second of the peak value.

Analysis of findings concerning traffic fluctuation can be used to estimate a number of design parameters for improved BS's local or network wide resource management, as, for example, required answer times, thresholds, safety margins for changes of system state and for short time traffic prediction.

The above results show that network optimization measures by hardware, as well as by software are feasible.

### VIII. POSSIBLE SOLUTION ON BS ARCHITECTURE BASED ON THIS SURVEY

The above analysis of load distribution and traffic profile characteristics provides basics for architecture of scalable BS hardware as proposed by Wajda [8]. The analysis shown that there are a number of traffic load ranges it can be optimally performed by specialized hardware modules with working characteristics adapted to supported traffic range. For example, assume BS consisting of two modules, where the first one supports the range up to 10% of max. traffic load and consumes 1 kW per hour and the second module supports the traffic peaks up to max. load and consumes 10 kW per hour. If the second module, due to the traffic load characteristics, can be switched off during 80% of the BS working time than the energy consumption can be greatly reduced. To assess energy savings we estimated traffic ranges and corresponding processing modules as shown in Table II and in Table III.

TABLE II. RESULTS FOR THE ANALYZED DUC AREA

Traffic profile	Number of BSs in the analysed population (DUC) [%]	Energy Saving factor	Proposition for HW modules [% load]	Power consumption savings by modular HW
1	ca. 40 %	0.90 – 0.95	5%, 100%	70%
2	ca. 53 %	0.89 – 0.85	5%, 25%, 100%	58%
3	ca. 7 %	0.84 – 0.70	5%, 25%, 50%, 100%	55%

TABLE III. RESULTS FOR THE ANALYZED GDU AREA

Traffic profile	Number of BSs in the analysed population (GDU) [%]	Energy Saving factor	Proposition for HW modules [% load]	Power consumption savings by modular HW
1	ca. 90 %	0.81–0.95	5%, 25%, 100%	56%
2	ca. 10 %	0.70 – 0.80	25%, 50%, 100%	45%

The calculations are based on power model and simulator used in EARTH project and presented by Imran et al. and Desset et al. [9][10][11].

### IX. CONCLUSION AND FUTURE WORK

Against well predictable global traffic profiles for DUC and GDU, traffic profiles of single cells are extremely volatile and not predictable. Therefore, optimization measures addressed to single cell must be well balanced and done separately from system wide optimization algorithms.

Traffic analysis of individual cells in DUC area had shown low utilization of cell resources and mostly low to middle oscillation of a load shape.

The developed metric, the Energy Saving Potential Index, confirmed the low utilization and, in consequence, a huge energy saving potential (over 0.9). Due to the estimated saving potential, it is worth to exploit low traffic utilization by BS architectural design.

Sophisticated network management strategies, for example those concerning resource sharing can benefit from this research, for example, common traffic

management can help to reduce traffic dynamic of particular BS and increase its resource utilization.

Furthermore, the defined BSs categories provide basis for a suitable network model with a given network traffic profile and can be used for development of energy effective HW by means of configurable set of procedural and physical building blocks that can be assembled manually, as well as in an automated way.

Flux details will help to boost BS switching time strategies and provide hints for hardware realization, as well as improve quality of traffic load prediction.

In the next research work, based on above findings, different modular HW solutions, transceivers based on power amplifiers with optimized dynamics, and new network management strategies will be proposed.

### ACKNOWLEDGMENT

I would like to give special thanks to my colleagues Andreas Wich and Ulrich Barth for the outstanding discussions. This work was supported in part by the German Federal Ministry of Economics and Technology funding program IT2Green (IntelliSpektrum).

### REFERENCES

- [1] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding Traffic Dynamics in Cellular Data Networks", The 30th IEEE International Conference on Computer Communications (IEEE INFOCOM 2011), April 2011, pp. 882 – 890, ISSN: 0743-166X, ISBN: 978-1-4244-9919-9
- [2] U. Kumar and A. Helmy, "Extract: Mining Social Features from WLAN Traces: A gender-based study", The 13-th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM'10), ACM SIGSIM, Oct. 2010, pp. 240-247, ACM 978-1-4503-0274-6/10/10.
- [3] J. Khalife, A. Hajjar, and J. Diaz-Verdejo, "A multilevel taxonomy and requirements for an optimal traffic-classification model", International Journal of Network Management, vol. 24, issue 2, March/Apr. 2014, pp. 101-120, doi: 10.1002/nem.1855.
- [4] Y. Won, B. Park, M. and Won, J. Chung, Kim, "Fine-grained traffic classification based on functional separation", International Journal of Network Management, vol. 23, Sep./Oct. 2013, pp. 350-381.
- [5] M.J. Gonzalez, D. Ferling, W. Wajda, A. Erdem, and P. Maugars, "Concepts for Energy Efficient LTE Transceiver Systems in Macro Base Stations", Future Networks and Mobile Summit 2011, Conference Proceedings, June 2011, pp. 33-41, IEEE Catalog Number: CFP1134J-PRT, ISBN: 978-1-4577-0928-9.
- [6] A. Ambrosy, O. Blume, D. Ferling, P. Jueschke, M. Wilhelm, X. Yu, "Energy Savings in LTE Macro Base Stations", 7th IFIP Wireless and Mobile Networking Conference, Vilamoura, Algarve, Portugal, May 2014, in press.
- [7] M. Jiang, M. Nikolic, S. Hardy, and L. Trajkovic, "Impact Of Self-Similarity On Wireless Data Network Performance", IEEE International Conference on Communications (ICC 2001), Juni 2001, pp. 477 - 481 vol. 2, ISBN: 0-7803-7097-1
- [8] W. Wajda "New concept of a modular wireless base station and benefits for network management and energy efficiency", IET International Conference on Wireless Communications and Applications (ICWCA 2012), Conference Proceedings, Oct. 2012, p. 34, ISBN:978-1-84919-550
- [9] EARTH (Energy Aware Radio and neTwork tecHnologies), EU funded research project FP7-ICT-2009-4-247733-EARTH, Jan. 2010 to June 2012, <https://www.ict-earth.eu/>
- [10] Ed. M.A. Imran et al., Energy efficiency analysis of the reference systems, areas of improvements and target breakdown, EARTH EC deliverable D2.3, December 2010, Available from <https://bscw.ict-earth.eu/pub/bscw.cgi/d71252>
- [11] C. Desset et al., "Flexible power modeling of LTE base stations", IEEE Wireless Communications and Networking Conference (WCNC 2012), Apr. 2012, pp. 28858-2862, doi: 10.1109/WCNC.2012.6214289.

# Classifying Users Based on Their Mobility Behaviour in LTE Networks

Bart Sas, Kathleen Spaey, Chris Blondia

iMinds/University of Antwerp

Antwerp, Belgium

Emails: {bart.sas, kathleen.spaey, chris.blondia}@uantwerpen.be

**Abstract**—When, in cellular networks like Long Term Evolution (LTE), there is a dense deployment of cells and/or users move at high velocities, handovers will occur frequently. This will have a severe impact on the Quality-of-Service (QoS) experienced by the users as there will be frequent call drops and a low throughput due to a service interruption time that is relatively long compared to the cell stay time. In order to mitigate these problems, users that experience a so called high mobility should be steered appropriately (e.g., to a cell on which they can be camped for a longer time). The first prerequisite for steering users that experience high mobility is identifying these users and predicting their future behaviour. In this paper, we present an algorithm that identifies users that follow similar trajectories through a cell. To perform this identification, a modified version of the Dynamic Time Warping (DTW) algorithm is used. Results show that the developed algorithm is able to detect users which follow similar trajectories adequately and is also able to distinguish between users that follow different trajectories through a cell.

**Keywords**—High Mobility; Dynamic Time Warping; Traffic Steering; Handover; LTE

## I. INTRODUCTION

In this paper, we focus on high mobility users in LTE networks, i.e., users that tend to make frequent handovers and for which the average amount of time they stay in a cell (time-of-stay) is low (order of magnitude of 10 seconds). For such users there will be a noticeable impact on user and network performance. This impact may be seen in a reduced QoS experienced by the users with high mobility due to a long data outage period relative to the cell stay time, an increased number of call drops and an increased signalling overhead in the core network due to handover signalling. The velocity of a user and the path it follows through a cell are the key factors that determine whether a user is subject to high mobility. Short time-of-stay can occur in two real-life situations: (1) when cell sizes are so small that even users with a low velocity perform frequent handovers and (2) when users move at a high velocity. In both situations, handovers will occur frequently and the time between entering a cell and leaving it will be small. Situation 1 might occur when there is a dense deployment of small cells, for instance in a shopping street/mall where users move at a low pace (for instance pedestrians). The cell inter-site distances in this case will be rather low (10-30 m) as will be the speed at which the users are travelling (2-3 km/h). Situation 2 might occur in macro cells along a busy highway or high-speed railroad: although the cells themselves are relatively large, the high pace of the users will cause cell stay times to be low. There do not necessarily have to be many cells involved, the aforementioned situations

might also occur in isolated cases with only a few cells that have users with high mobility.

In this paper, a Self-Organising Network (SON) function [1] that aims at optimising the handover behaviour of high mobility users is presented, called the High Mobility SON function. The goal of this SON function is to classify users according to their mobility behaviour and, based on this classification, steer them in an appropriate way such that the handover rate is reduced while the QoS is maintained or possibly improved. The focus of this paper is on the design and evaluation of a method to classify users according to the trajectory they follow through a cell. By assuming that users that follow similar trajectories (i.e., similar geographical paths at similar velocities) through a cell will have similar behaviour regarding mobility and handovers, the future behaviour of a user can be predicted based on observations made from past users that had similar trajectories. These predictions can then at a later stage be used to decide which handovers are useful and which are not and to decide what the best destination cell for a handover is such that the amount of handovers can be reduced by avoiding unnecessary and suboptimal handovers.

The remainder of this paper is structured as follows: Section II gives an overview of work that is related to the topic of this paper. Section III describes the general architecture of the SON function that we designed to find users that have high mobility and to steer them such that the negative effects of the frequent handovers are mitigated. The specific component of the SON function that is the focus of this paper, the Trajectory Classifier, will use standardised measurements; these are explained in Section IV. Section V describes the classical DTW algorithm and the modifications that were made to it in order for it to be able to identify users that follow similar trajectory through a cell. Section VI describes the simulation studies that were performed in order to assess the ability of the algorithm to classify users based on their trajectories. Finally, Section VII concludes this paper and lists the future work.

## II. RELATED WORK

Dynamic Time Warping is a well-known technique for finding similarities in time-series. It finds the most optimal alignment between two time series and is able to deal with slight variations in time and speeds. It is often used in automatic speech recognition for finding patterns, furthermore it is also used in speaker recognition and signature recognition [2][3]. DTW is more robust than more simple techniques like the Euclidean distance as DTW is able to deal with slight variations in the input like accelerations and decelerations.

Traffic steering is a technique whereby mobile users are intelligently steered towards certain base stations. It can be applied for various reasons like improving QoS, balancing the load among different base stations, satisfying certain capability needs, etc. Traffic steering can be performed between similar cells, between different layers of the same technology [4] or even between different technologies [5][6][7]. In this paper, we will consider steering users between cells of the same technology, namely LTE. The algorithms that are introduced in this paper could however be extended to other cellular technologies as well.

Also in [8][9][10], the problem of high mobility is considered. Fei and Fan [8] describe two position-assisted handover schemes: one that aims to reduce the handover delay and one that aims to reduce the handover frequency and improve the handover success rate. Both schemes do however rely on the geographical locations of the users, which are not always known. The goal of the schemes is also not targeted at users with high mobility, but instead to high velocity users. Papathanasiou et al. [9], and Cheng and Fang [10] are concerned with applying scanning narrow beams to LTE networks that contain fast moving mobiles.

### III. GENERAL ARCHITECTURE

The general architecture of the proposed High Mobility SON function is shown in Figure 1. It consists of a number of components, which are described below. The *Trajectory Classifier* is the core of the SON function. It is responsible for classifying users according to the trajectory that they follow through a cell. Users that follow similar trajectories than other past users will be identified by it. This information can then be used to predict the future behaviour of currently active users. This component plays an important role in the SON function and will be the focus of this paper.

The set of trajectories that are available to the Trajectory Classifier to map an active user on is determined by the *Trajectory Identifier*. This component decides which trajectories are distinct and useful enough to be considered for matching users to by the Trajectory Classifier.

The *Mobility Classifier* is responsible for determining the mobility type of the users. The mobility type of a user can for instance be vehicular user, pedestrian, stationary user, etc. Knowing what the mobility type of the users is, is important for knowing how they will behave in the future. Pedestrians might for instance regularly stop or go slower or faster in an unpredictable way while vehicular users will more likely move at a predictable pace, only stopping at intersections and traffic lights.

Finally, the *Traffic Steerer* is responsible for the actual decision of when and to which target cell the user should be handed over. Based on the trajectory on which an active user has been mapped and its mobility type the Traffic Steerer will estimate when certain events like call drops, the throughput or some other QoS Key Performance Indicator (KPI) becoming too low, etc. will occur; and which action should be taken and at what time in order to steer the users as appropriate as possible.

The decisions that are taken by the Traffic Steerer are communicated to the *Handover algorithm*. This algorithm is

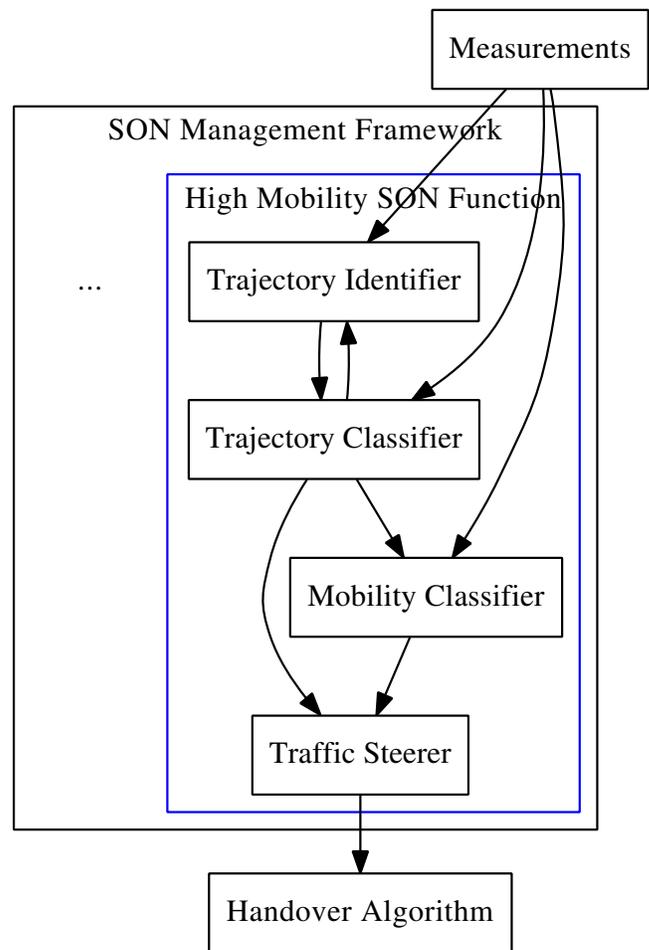


Figure 1. The general architecture of the proposed High Mobility SON function

responsible for performing the actual handovers. Apart from executing the commands coming from the Traffic Steerer, the Handover (HO) algorithm is also responsible for handing over users for which the SON function did not provide instructions as the SON function will only deal with these users for which it sees an opportunity to optimise the handover behaviour.

Typically, a SON function will be part of a larger SON management framework like for example the SEMAFOUR integrated SON management framework [11]. This framework integrates existing and future SON functions across several radio access technologies and is in charge of functions like policy transformation/supervision and conflict detection/resolution among multiple SON functions.

The High Mobility SON function will be placed at the eNodeB and will operate on a per-cell level. This allows the SON function to be gradually deployed in the network and makes it scalable.

### IV. MEASUREMENT REPORTS

As mentioned before, the Trajectory Classifier is the component of the SON function that is responsible for classifying users based on the trajectory they follow through a cell. It performs a key role in the SON function as it provides information that

serves as input for most of the other components. In order to distinguish between users that follow similar trajectories and other users we need some kind of measurements that are made by the User Equipment (UE) and sent to its Serving eNodeB (SeNB), which can then use them. It is important that these measurements are sent sufficiently frequent in order for the Trajectory Classifier to make actual decisions; furthermore they must allow distinguishing between users that travel along different trajectories through a cell. The measurement reports that will be used by our algorithm are the measurement reports a UE sends to its SeNB as part of the LTE handover process. Both these measurement reports as well as the measurement report triggering have been standardised [12]. Active users monitor the Reference Signal Received Power (RSRP) and/or Reference Signal Received Quality (RSRQ) of their SeNB and its surrounding Neighbouring eNodeBs (NeNBs). Whenever a certain condition involving the RSRP/RSRQ holds and sometimes also when it no longer holds, the UE sends a measurement report to its SeNB containing the list of NeNBs for which the condition (still) holds at that time as well as the RSRP/RSRQ measured for the SeNB and the NeNBs in the list. The triggering of measurement reports at the UE is configured by the SeNB. There are a number of possible event types that can be configured, named event A1 through A6. All these events specify an entering and leaving condition. These conditions are equations that involve the RSRP/RSRQ of the SeNB and/or the NeNB; and/or event-specific thresholds. In order to prevent measurement reports from being triggered too often due to small fluctuations in the RSRP/RSRQ measurements, the triggering conditions also specify a hysteresis value that specifies an extra offset that is added to the equation. Furthermore, the conditions also have to hold for a certain amount of time called the Time-to-Trigger (TTT) before the events are fired. The event-specific thresholds, hysteresis and TTT of an event can all be specified for each individual event that is configured at the UE. In this paper, event A4 is exclusively used as it provides information about the RSRPs of the NeNBs in comparison to a fixed threshold. The other events provide information about the RSRPs relative to the SeNB or about the SeNB itself, which is less usable as the information is more volatile or less exhaustive. An A4 event is triggered when the RSRP/RSRQ of a user becomes better than a certain threshold. By configuring a number of these events the SeNB of a UE will at every point in time know the signal levels of its NeNBs that are observed by the UE within an upper and a lower bound, because whenever the observed RSRP/RSRQ of a NeNB changes and crosses one of the thresholds an update is sent to the SeNB. As multiple events are often triggered at the same time, because of sudden changes in the RSRP/RSRQ, events that occur within a short amount of time ( $< 0.03$  s) are aggregated into a single measurement. This results in a time series where every element represents an update of the RSRP/RSRQ level of the NeNBs. Using this information users will be classified based on their trajectory. The rationale behind this approach is that users that pass through the same parts of a cell will measure similar signal strengths from the same surrounding eNodeBs.

In the following section, the algorithm that matches users that follow similar trajectories based on the collected measurements is presented.

## V. DYNAMIC TIME WARPING

The key to handing over users to the most suited target cell is being able to identify users that follow similar trajectories and to distinguish between users that follow different trajectories.

Measurements that are sent by users can slightly differ due to slight deviations in the trajectory that is followed, slightly different user velocities, time variations in fading, etc. Because of this, it is not just possible to compare the measurements of both the active and reference users by looking for, for instance, the longest common sub-string. In order to make this matching more resilient against slight variations in the measurement data, a modified version of the DTW algorithm is used.

### A. Classical Dynamic Time Warping

The DTW algorithm [2] is used in signal processing to find an optimal alignment between two time series (like the measurements coming from the users). Dynamic Time Warping determines the distance between two time series  $X = (x_1, \dots, x_M)$  and  $Y = (y_1, \dots, y_N)$  by determining the so called optimal warping path through the cost matrix  $C \in R^{M \times N}$  whose elements  $C_{m,n}$  express the distance  $c(x_m, y_n)$  between the elements  $x_m$  and  $y_n$  of the respective series. A warping path through this cost matrix is a series  $p = (p_1, \dots, p_L)$  with  $p_l = (m_l, n_l) \in \{1, \dots, M\} \times \{1, \dots, N\}$ , satisfying the restrictions that  $p_1 = (1, 1)$ ,  $p_L = (M, N)$  and  $p_l = (m_l, n_l)$  for  $l \in \{2, \dots, L\}$  can only be reached from  $p_{l-1} \in \{(m_l - 1, n_l), (m_l, n_l - 1), (m_l - 1, n_l - 1)\}$ , i.e., a warping path is a path through the cost matrix that starts at  $(1, 1)$  and goes to  $(M, N)$  by either increasing the row index, increasing the column index or increasing both at the same time. The cost or distance of a warping path is given by the total cost of the elements along the path:

$$\sum_{l=1}^L c(x_{m_l}, y_{n_l}) \quad (1)$$

The optimal warping path is the warping path that has the lowest total cost of all possible warping paths. The optimal warping path can be efficiently determined by constructing an  $M \times N$  matrix  $D$  in which each element  $(m, n)$  contains the minimal total cost to match the prefix of length  $m$  of  $X$  with the prefix of length  $n$  of  $Y$ . Each element  $(m, n)$  of this matrix (except from the ones on the first row and column) is calculated by adding the cost  $c(x_m, y_n)$  to the minimum of  $D[m-1, n]$ ,  $D[m, n-1]$  and  $D[m-1, n-1]$ . The values of the elements of the first row and column are calculated by adding the cost to the value of the elements to the left or above respectively. Pseudocode for the dynamic time warping algorithm is given in Figure 2. In this code, an additional row and column are added to the matrix  $D$ . The elements in the first row and column of this matrix are initialised to infinity, except for the element in the upper left corner which is set to 0. By doing this, the elements in the remainder of the matrix can all be treated the same.

### B. Modified Dynamic Time Warping

The major shortcoming of the classical DTW algorithm for our purpose is that it matches two time series entirely. This is however not desirable. First, the more recent past of the active user is more interesting as its behaviour in the future

```

1:  $D := \text{array}[0..M, 0..N]$ 
2:  $D[0, 0] := 0$ 
3: for  $m := 1 \rightarrow M$  do
4:    $D[m, 0] := \infty$ 
5: end for
6: for  $n := 1 \rightarrow N$  do
7:    $D[0, n] := \infty$ 
8: end for
9: for  $m := 1 \rightarrow M$  do
10:  for  $n := 1 \rightarrow N$  do
11:     $D[m, n] := c(X[m], Y[n]) + \min($ 
12:       $D[m - 1, n],$ 
13:       $D[m, n - 1],$ 
14:       $D[m - 1, n - 1])$ 
15:  end for
16: end for
    
```

Figure 2. The classical Dynamic Time Warping algorithm.

will be influenced more by this than by its earlier behaviour. Furthermore, it should be possible to match an active user with a reference user without having to match the latest measurement of the active user with the last reference measurement as this usually is when the reference user left the cell and we want to be able to proactively make decisions before the active user also leaves the cell.

In order to do so, the DTW algorithm is adapted as follows: the matrix that is used to calculate the intermediate distances is filled backwards, i.e., the DTW algorithm is applied to the reverse series. By doing this each element of the matrix contains the minimal total cost, when only matching the suffixes of both measurement series up to that point. During construction of the matrix, a current best match is kept and updated each time the value of an element is calculated. By doing this, by the time the upper left corner of the matrix (the element that corresponds to the first elements in both series) is reached the best matching suffixes of the measurement series of the active and reference user has been found.

In order to match a suffix of the measurement series of the active user with any interval of the measurements of the reference user, the end of the measurement series of the active user is shifted along the elements of the measurement series of the reference user. The Modified Dynamic Time Warping (MDTW) algorithm is applied to the entire active series and the sub-series of the reference user starting from the first measurement up to the measurement that is aligned with the end of the measurement series of the active user, as is illustrated in Figure 3. As the MDTW algorithm is able to find matches of sub-series of different lengths, it is important that the length of the warping path is taken into account when determining the optimal warping paths. In order to do this, costs will be represented using tuples  $(v, w)$ . The component  $v$  represents the value of the cost and is a value between 0 and 1. When  $v$  has value 0 this means that there is an exact match between two measurements. When  $v$  has value 1 this means that two measurements are completely different. The component  $w$  represents the importance or weight of the value. The exact meaning of the importance depends on how the cost between measurements is defined, but as a general rule the weight should be higher when the cost is based on more information, for instance a longer warping path. Two operations

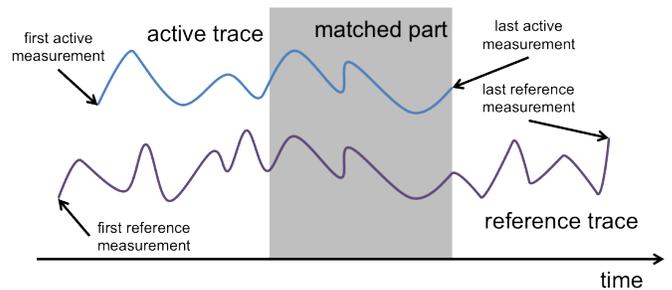


Figure 3. A suffix of the active trace is matched with an interval of the reference trace

are needed by the dynamic time warping algorithm: the addition of costs and the comparison of costs. The addition of two costs is defined as follows:

$$(v_1, w_1) + (v_2, w_2) = \left( \frac{v_1 w_1 + v_2 w_2}{w_1 + w_2}, w_1 + w_2 \right) \quad (2)$$

As can be seen from (2), the value of the sum is equal to the weighed average of both addends while the weight of the sum is equal to the sum of the weights. Two costs are compared by ignoring the weights and just comparing their values:

$$(v_1, w_1) < (v_2, w_2) \iff v_1 < v_2 \quad (3)$$

Due to the way that addition is implemented, tuples will be compared based on the average cost of the elements along the warping path.

Considering every possible suffix of both series has a disadvantage: shorter suffixes will be favoured over longer suffixes as it is more likely to find a short perfect match than it is to find a longer one. In order to mitigate this problem, the initial cost is set to  $(1, x)$  with  $x \geq 0$ . Since after a series of additions, the resulting value of the tuple is the weighted mean of all added values, the importance of this initial cost will become less important as more values are added to it. By using a tuple with non-zero weight as the initial cost, longer matches will be favoured over shorter matches. Note that adding  $(v, w)$  to  $(1, 0)$ , i.e.,  $x = 0$ , will result in  $(v, w)$  removing the influence of the initial cost altogether.

Another problem that arises when allowing the algorithm to match every combination of suffixes is that the cost matrix can be traversed mainly horizontally or vertically, as this produces warping paths of the same length as going diagonally. Going mainly vertically or horizontally is however less favourable as this will match shorter portions of either one of the compared traces for the same length of warping path. This is not a problem with the classical DTW algorithm as it matches the entire series. In order to mitigate this problem an additional cost can be added when traversing the cost matrix in the vertical or horizontal directions, but not when going diagonally. This will 'encourage' the algorithm to match longer portions of both series. Like the initial cost, this additional cost will have the form  $(1, x)$  with  $x \geq 0$ . Pseudocode for this is given in (4).

$$D[i, j] := c + \min ( D[i - 1, j] + \text{extraCost}, \\ D[i, j - 1] + \text{extraCost}, \\ D[i - 1, j - 1] ) \quad (4)$$

```

1: bestMatch := None
2: bestDistance := ∞
3: for start := 1 → m do
4:   distances := array[1..start + 1, 1..n + 1]
5:   for i := 1 → start do
6:     distances[i, n + 1] := ∞
7:   end for
8:   for j := 1 → n do
9:     distances[start + 1, j] := ∞
10:  end for
11:  distances[start + 1, n + 1] := initialCost
12:  for i := start → 1 do
13:    for j := n → 1 do
14:      cost := calculateDistance(
15:        referenceMeasurements[i],
16:        activeMeasurements[j])
17:      distances[i, j] := cost + min(
18:        distances[i + 1, j] + extraCost,
19:        distances[i, j + 1] + extraCost,
20:        distances[i + 1, j + 1])
21:      if distances[i, j] < bestDistance then
22:        bestMatch := (start, i, j)
23:        bestDistance := distances[i, j]
24:      end if
25:    end for
26:  end for
27: end for

```

Figure 4. The modified Dynamic Time Warping algorithm

The pseudocode for the MDTW algorithm is given in Figure 4. Note from this code that the MDTW algorithm will always return a match, together with the distance of that match.

## VI. EVALUATION

In order to assess the ability of the MDTW algorithm to identify users that follow similar trajectories and to distinguish between users that follow different trajectories, simulations were performed using a simulator that is based on the OMNeT++ simulation library [13]. In these simulations, a user moves around producing reference traces. Afterwards, the user starts to produce active traces, which are then matched to the reference traces. Each active trace is compared to all reference traces and the reference trace for which the MDTW algorithm finds the match with the lowest ‘bestDistance’ is considered as the reference trace to which the active trace is matched. The accuracy of this match will be assessed using a performance metric (explained in Section VI-B), which is based on the geographical locations that are visited by the user. These geographical locations are collected during the simulations specifically for the purpose of being able to assess the accuracy of the matches made. In reality these geographical locations will not necessarily be available; therefore it is important that it is possible to specify criteria based on known information that determine whether a match made by the MDTW algorithm can be trusted or not.

### A. Simulation Setup

The simulation area is a rectangle measuring 1732 by 2000 metres featuring wrap-around. It contains 16 three-sectored

cells with directed antennas, placed at regular distances, as is depicted in Figure 5. Pathloss calculations are performed using the Okumura-Hata model for large urban areas. Furthermore, shadow fading that is both auto-correlated in time and cross correlated with the shadow fading of other antennas is considered [14]. Two different scenarios are considered. The goal of

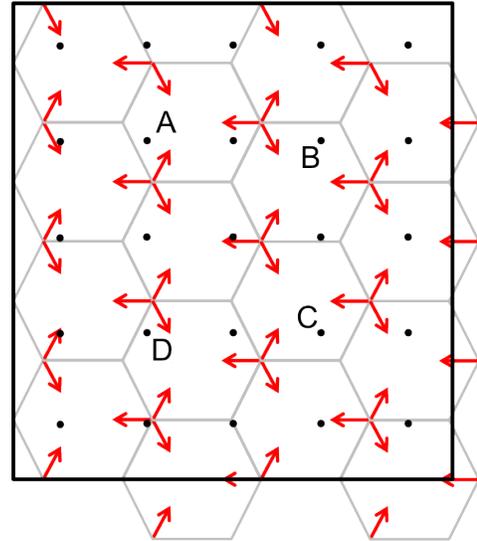


Figure 5. Overview of the simulation area

the first scenario is to assess whether the MDTW algorithm is actually able to find good matches in case it is a certainty that good matches exist. In this scenario a single user follows a rectangular path through the simulation area, passing through the same locations multiple times. The path that is followed by the user is the path between the points A, B, C and D in Figure 5. The simulations consist of two phases: in the first phase reference traces are collected and stored by the different SeNBs the user is connected to. After a certain amount of time (10000 s), the second phase starts. In this phase, the SeNBs start to match the measurements that are generated by the active user with reference traces from the past. These simulations are carried out a number of times. Each time with a different interval from which the user chooses its velocity. The user chooses a different velocity at each corner of its rectangular path. These intervals are  $[5 - \frac{i}{2} \text{ m/s}; 5 + \frac{i}{2} \text{ m/s}]$  with  $i \in \{0, 1, 2, 3, 4, 5\}$ . By doing this, also the ability of the MDTW algorithm to deal with slight variations in the velocity of the users can be assessed.

In the second scenario, the user no longer moves along a fixed path. Instead, it moves from point to point (the black points in Figure 5) in a rectangular grid of points that are equally distributed across the simulation area, i.e., so-called Manhattan mobility [15]. Each time a user reaches a point, it randomly chooses to go either left, right or forward with the probability of going forward being twice as large as the probability of going either left or right. It then travels along the chosen direction with a velocity that is chosen from a certain velocity interval. Similar as in the first scenario this interval is different between simulations. Furthermore, the user not just travels between the the grid points, but instead the points it travels between are chosen uniformly within a circle centred around the grid points. The radius  $r$  of this circle is either 0 m, 1 m, 5 m or 10 m.

When the radius is 0 m the user travels directly between the grid points.

Nineteen different A4 events were configured. Their thresholds range from -100 dBm to -10 dBm in steps of 5 dB. The hysteresis of all these events is set to 2 dB and the TTT to 480 ms. The 'initialCost' and 'extraCost' parameters of the MDTW algorithm are set to 0.1 and 0.01 respectively.

An overview of the simulation parameters is given in Table I.

Table I. OVERVIEW OF THE SIMULATION PARAMETERS

Parameter	Value
Simulation area	1732 × 2000 m
Pathloss model	Okumura-Hata (large urban)
User height	1.8 m
Base station height	30 m
Site-to-site distance	500 m
Carrier frequency	2.6 GHz
Antenna model	NGMN [16]
Number of sectors	48
Mean user velocity	5 m/s
Event A4 hysteresis	2 dB
Event A4 time-to-trigger	480 ms
Simulation duration	20000 s
Initial cost (initialCost)	0.1
Extra cost (extraCost)	0.01

### B. Performance Metric

In order to assess whether the matches that are made by the MDTW algorithm are accurate, a metric is needed that assigns an accuracy to the matches. The accuracy metric should reflect how well the algorithm is able to identify the part of the trajectory where the reference and active user followed a similar geographical path at a similar velocity through the cell. It is important that this metric takes into account both the part of the trajectory that was matched by the algorithm, but that does not overlap in reality and the part that does overlap in reality but was not matched by the algorithm, as it is important that the MDTW algorithm is able to identify the exact part of the trajectory that overlaps.

As described in Section VI-A, the users move through the simulation area along certain paths. For each match that is made, the portion of the paths of the active and reference users that overlap geographically is determined. In case there is no geographical overlap, the accuracy metric will be set equal to 0. From this overlapping part, the start and end times at which the active user visited the overlapping part are determined. Suppose  $o_s$  is the time on which the active user enters the overlapping part and  $o_e$  the time on which it leaves it. Similarly,  $m_s$  and  $m_e$  are the respective begin and end times of the match made by the MDTW algorithm. The accuracy metric is then given by (5).

$$\text{Accuracy} = \frac{\max(0, \min(o_e, m_e) - \max(o_s, m_s))}{\max(o_e, m_e) - \min(o_s, m_s)} \quad (5)$$

This formula will yield 0 when the interval of which the MDTW decided that the users followed the same trajectory and the interval corresponding to the geographical overlapping parts are completely disjoint and 1 when they are exactly the same, i.e., there is a perfect match. An illustration of this is given in Figure 6. In this figure, the accuracy metric is the ratio of the common part of both intervals (the part between  $m_s$  and  $o_e$ ) and the width between the leftmost and rightmost ends of the intervals (the part between  $o_s$  and  $m_e$ ).

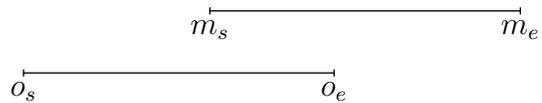


Figure 6. The accuracy metric is calculated as the ratio between the length of the interval  $[m_s; o_e]$  and the length of the interval  $[o_s; m_e]$

### C. Results

Figure 7 shows the results obtained with the first scenario. The x-axis contains the number of active measurements on which the match is based. The y-axis contains the average accuracy of all matches with the corresponding number of measurements on the x-axis. The different curves are for the different velocity intervals. As it can be seen in this figure, in

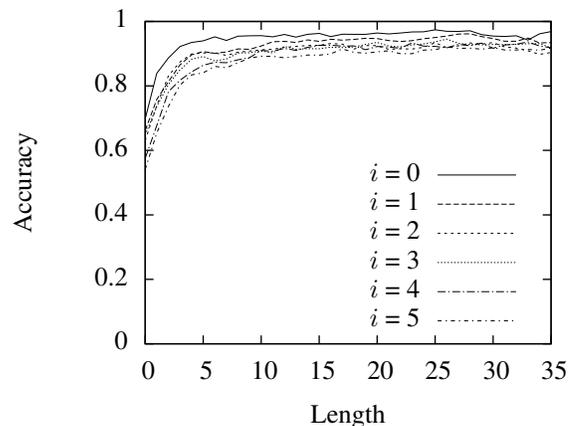


Figure 7. The accuracy metric approaches 1 (perfect match) as the number of measurements on which the match is based increases

case it should be possible to find accurate matches ( $i = 0$ , velocity interval  $[5 \text{ m/s}; 5 \text{ m/s}]$ ) the accuracy of the matches becomes high ( $> 0.9$ ) when the matches are based on 5 or more measurements. These matches are not necessarily perfect, which is due to small deviations on when measurements are generated.

When the velocity interval becomes larger, the accuracy of the MDTW algorithm becomes slightly worse. This is to be expected as different velocities will cause measurements to be triggered on different locations, which will make it more difficult for the MDTW algorithm to make accurate matches. The obtained results are, however, still very good and accurate enough to make predictions. Note that when the differences in velocity would become even more pronounced it is no longer desired that good matches are made as users with pronounced different velocities should be treated differently anyway.

The results of scenario 1 show that, in case it is certain that good matches exist (because of how the scenario is constructed), these are found by the algorithm, except if they are based on too few measurements.

Figure 8 and Figure 9 show the results obtained with the second scenario. Note that in this scenario good matches will

not necessarily exist, because it is possible that the active user follows a path that is different from all earlier collected reference paths. On the y-axis of Figure 8 and Figure 9, again, the accuracy of the matches is shown. The x-axis now contains the score that is associated with the matches by the MDTW algorithm (i.e., the value of the lowest ‘bestDistance’). In these figures, results for matches based on less than 10 measurements have been left out. Figure 8 shows the results obtained with

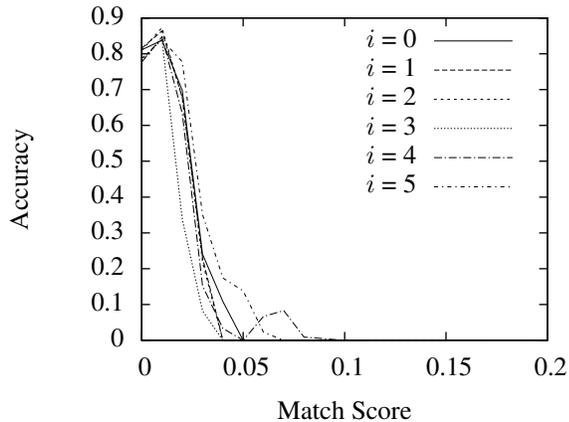


Figure 8. The MDTW algorithm is able to assign low scores to accurate matches and higher scores to inaccurate matches for different velocity intervals

$r = 0$  (i.e., the user travels directly between the grid points of the scenario), for different velocity intervals. This figure shows that there is a strong correspondence between the score that is assigned to a match by the MDTW algorithm and the actual accuracy. Furthermore, there is a clear drop of the match score around 0.03. This drop occurs around the same value for all curves, which means that it is independent from the velocity of the users.

Figure 9 shows the results obtained with  $i = 0$  (i.e., the user always travels at a fixed velocity of 5 m/s) but for different radiuses  $r$  around the grid points of the Manhattan mobility model. The result is similar as for Figure 8; when the score that is assigned by the MDTW algorithm is low, the accuracy of the match is high and it decreases steeply around a match score of about 0.03. From the results obtained with scenario 2 it is clear that the match score of the MDTW algorithm reflects the accuracy of the matches made: if the accuracy of the match is high (i.e., a match we want to trust) then the match score is low (below 0.03). The match score together with the number of measurements on which the match is based can thus be used as criteria that determine whether a match made by the MDTW algorithm can be trusted or not.

## VII. CONCLUSIONS AND WAY FORWARD

In this paper, an algorithm that identifies users that follow similar trajectories through a cell and distinguishes between users that follow different trajectories was discussed. This algorithm is based on the DTW algorithm that is used in signal processing. The DTW algorithm was adapted such that it is able to find the best match of any suffix of one time series

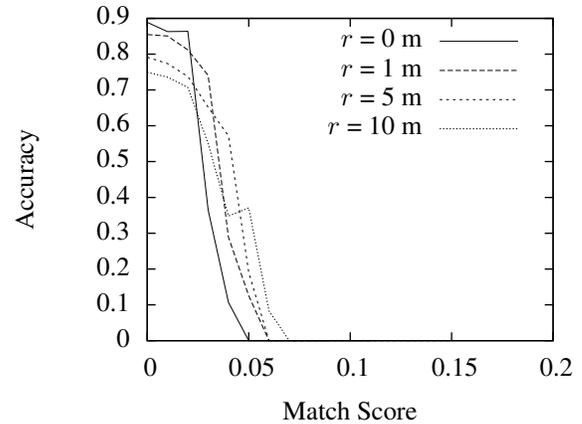


Figure 9. The MDTW algorithm is able to assign low scores to accurate matches and high scores to inaccurate matches for different deviations

(measurements from an active user) with any interval of another series (measurements from a reference user).

The ability of the algorithm to match users that follow similar trajectories was assessed by simulations. First a scenario was considered in which it is certain that good matches should exist and then a scenario was examined in which good matches do not necessarily exist. The results show that the algorithm is actually able to identify users that follow similar trajectories through a cell except if there are too few measurements. The algorithm is also able to deal with small variations in the input. This is important as users will never behave exactly the same: there will always be small differences in velocity and the trajectory that the users follows. From the results, it became also clear that the match score the MDTW algorithm assigns to a match reflects the real accuracy of the match: when the accuracy of the match is high the match score is very low. So, the match score together with the number of measurements on which the match is based can be used as criteria for determining whether a match made by the MDTW algorithm can be trusted or not.

In the future, this algorithm will be used as part of a larger SON function that aims at steering users that have a certain mobility behaviour more appropriately in order to reduce the number of call drops and improve the QoS of the users while reducing the signalling overhead in the core network. The applicability of the SON function in other cellular technologies could also be tested as this paper only focused on LTE.

## ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union, Seventh Framework Programme (FP7/2007-2013) under grant agreement n°316384.

## REFERENCES

- [1] S. Hämäläinen, H. Sanneck, and C. Sartori, LTE Self-Organising Networks (SON): Network Management Automation for Operational Efficiency, 1st ed. Wiley Publishing, 2012.
- [2] Muller and Meinard, Information Retrieval for Music and Motion. Springer, 2007.

- [3] C. Ratanamahatana, J. Lin, D. Gunopulos, E. Keogh, M. Vlachos, and G. Das, "Mining time series data," in *Data Mining and Knowledge Discovery Handbook*, O. Maimon and L. Rokach, Eds. Springer US, 2010, pp. 1049–1077.
- [4] P. Fotiadis et al., "Multi-layer mobility load balancing in a heterogeneous LTE network," in *Vehicular Technology Conference (VTC Fall)*, 2012 IEEE, Sept 2012, pp. 1–5.
- [5] S. Frei, W. Fuhrmann, A. Rinkel, and B. Ghita, "Prospects for WLAN in the evolved packet core environment," in *New Technologies, Mobility and Security (NTMS)*, 2012 5th International Conference on, May 2012, pp. 1–5.
- [6] M. D. Nisar, V. Pauli, and E. Seidel, "Multi-RAT traffic steering - why, when, and how could it be beneficial?" Nomor white paper, December 2011.
- [7] N. Jorgensen, D. Laselva, and J. Wigard, "On the potentials of traffic steering techniques between HSDPA and LTE," in *Vehicular Technology Conference (VTC Spring)*, 2011 IEEE 73rd, May 2011, pp. 1–5.
- [8] M. Fei and P. Fan, "Position-assisted fast handover schemes for LTE-advanced network under high mobility scenarios," *Journal of Modern Transportation*, vol. 20, no. 4, 2012, pp. 268–273.
- [9] C. Papathanasiou, N. Dimitriou, and L. Tassiulas, "On the applicability of steerable beams in LTE-advanced networks with high user mobility," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, 2012, p. 234.
- [10] M. Cheng and X. Fang, "Location information-assisted opportunistic beamforming in LTE system for high-speed railway," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, 2012, p. 210.
- [11] A. Eisenblätter et al., "Integrated self-management for future radio access networks: Vision and key challenges," in *Future Network and Mobile Summit (FutureNetworkSummit)*, 2013, July 2013, pp. 1–10.
- [12] "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description; stage 2 (release 11)," 3rd Generation Partnership Project, Tech. Rep. 3GPP TS 36.300 v11.3.0, September 2012.
- [13] OMNeT++ Network Simulation Framework. [Online]. Available: <http://www.omnetpp.org/>
- [14] R. Fraile, J. F. Monserrat, J. Gozálviz, and N. Cardona, "Wireless systems mobile radio bi-dimensional large-scale fading modelling with site-to-site cross-correlation," *European Transactions on Telecommunications*, vol. 19, no. 1, 2008, pp. 101–106.
- [15] F. Bai and A. Helmy, *Wireless Ad Hoc and Sensor Networks*. Springer, October 2006, ch. A Survey of Mobility Modeling and Analysis in Wireless Adhoc Networks.
- [16] "Radio access performance evaluation methodology," Next Generation Mobile Networks, Tech. Rep., January 2008.