



# **ICWMC 2011**

The Seventh International Conference on Wireless and Mobile Communications

June 19-24, 2011

Luxembourg City, Luxembourg

## **ICWMC 2011 Editors**

Dragana Krstic, University of Nis, Serbia

Andreas Löffler, Friedrich-Alexander University of Erlangen-Nuremberg, Germany

Constantin Paleologu, University Politehnica of Bucharest, Romania

# ICWMC 2011

## Foreword

The Seventh International Conference on Wireless and Mobile Communications [ICWMC 2011], held between June 19 and 24, 2011, in Luxembourg, followed on the previous events on advanced wireless technologies, wireless networking, and wireless applications.

ICWMC 2011 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 2011 Technical Program Committee, as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICWMC 2011. 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 2011 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICWMC 2011 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 historic charm Luxembourg.

### ICWMC 2011 Chairs:

Tommi Aihkialo, VTT Technical Research Centre of Finland - Oulu, Finland

Erick Amador, Infineon Technologies, France

David Boyle, Tyndall National Institute, University College Cork, Ireland

Silviu Ciochina, Universitatea Politehnica din Bucuresti, Romania

Guangjie Han, Hohai University, Japan

Ying Jian, Google, Inc., USA

György Kálmán, ABB AS - Akershus, Norway

Dragana Krstic, University of Nis, Serbia

Jong-Hyook Lee, INRIA, France

Jingli Li, TopWorx - Emerson, USA

Antonio Liotta, Eindhoven University of Technology, The Netherlands

Andreas Löffler, Friedrich-Alexander University of Erlangen-Nuremberg, Germany  
Jonathan Loo, Middlesex University - Hendon, UK  
Catherine Meadows, Naval Research Laboratory - Washington DC, USA  
Georg Neugebauer, RWTH Aachen University, Germany  
Christopher Nguyen, Intel Corp., USA  
Homayoun Nikookar, Delft University of Technology, The Netherlands  
Mariusz Skrocki, Orange Labs - Warszawa, Poland  
Naoki Wakamiya, University of Osaka, Japan  
Ouri Wolfson, University of Illinois at Chicago, USA  
Yan Zhang, IMEC- NL, The Netherlands  
Gang Zhao, Digital Technology Laboratory - DMG/Mori Seiki, USA

# ICWMC 2011

## Committee

### ICWMC Advisory Chairs

Jonathan Loo, Middlesex University - Hendon, UK  
Catherine Meadows, Naval Research Laboratory - Washington DC, USA  
Antonio Liotta, Eindhoven University of Technology, The Netherlands  
Ouri Wolfson, University of Illinois at Chicago, USA  
Naoki Wakamiya, University of Osaka, Japan

### ICWMC Industry/Research Chairs

Yan Zhang, IMEC- NL, The Netherlands  
Gang Zhao, Digital Technology Laboratory - DMG/Mori Seiki, USA  
Christopher Nguyen, Intel Corp., USA  
Mariusz Skrocki, Orange Labs - Warszawa, Poland  
Erick Amador, Infineon Technologies, France  
Ying Jian, Google, Inc., USA  
Jingli Li, TopWorx - Emerson, USA  
György Kálmán, ABB AS - Akershus, Norway

### ICWMC Special Area Chairs

#### Signal processing

Silviu Ciochina, Universitatea Politehnica din Bucuresti, Romania

#### Security

Georg Neugebauer, RWTH Aachen University, Germany

#### Mobility

Dragana Krstic, University of Nis, Serbia

#### Vehicular

Jong-Hyouk Lee, INRIA, France

#### Sensors

David Boyle, Tyndall National Institute, University College Cork, Ireland  
Guangjie Han, Hohai University, Japan

#### RFID

Andreas Löffler, Friedrich-Alexander University of Erlangen-Nuremberg, Germany

#### Cognitive radio

Homayoun Nikookar, Delft University of Technology, The Netherlands

#### Applications/Services

Tommi Aihkisalo, VTT Technical Research Centre of Finland - Oulu, Finland

## ICWMC 2011 Technical Program Committee

Jemal H. Abawajy, Deakin University, Australia  
Mohammed Abdel-Hafez, UAE University-Al-Ain, United Arab Emirates  
Seyed Reza Abdollahi, Brunel University, UK  
Tankut Acarman, Galatasaray University - Istanbul, Turkey  
Javier M. Aguiar Pérez, Universidad de Valladolid, Spain  
Tommi Aihkisalo, VTT Technical Research Centre of Finland - Oulu, Finland  
Ahmed Akl, LAAS/CNRS - Toulouse, France  
Ramón Pablo Alcarria Garrido, Universidad Politécnica de Madrid, Spain  
Erick Amador, Infineon Technologies, France  
Samer Ammoun, Modulowatt Ingénierie - Paris, France  
Radu Arsinte, Technical University of Cluj-Napoca, Romania  
Faouzi Bader, CTTC (Centre Tecnologic de Telecommunications de Catalunya) - Barcelona, Spain  
Joonsang Baek, Institute for Infocomm Research, Singapore  
Mohammad M. Banat, Jordan University of Science and Technology, Jordan  
Alireza Borhani, University of Agder - Grimstad, Norway  
David Boyle, Tyndall National Institute, University College Cork, Ireland  
Maria Calderon, University Carlos III of Madrid, Spain  
Juan-Carlos Cano, Universidad Politécnica de Valencia, Spain  
Hsing-Lung Chen, National Taiwan University of Science and Technology - Taipei, Taiwan  
Mohammad Mushfiqur Rahman Chowdhury, University Graduate Center (UniK) / University of Oslo, Norway  
Silviu Ciochina, Universitatea Politehnica din Bucuresti, Romania  
Hugo Coll Ferri, Polytechnic University of Valencia, Spain  
Mauro Conti, Vrije Universiteit - Amsterdam, The Netherlands  
Hector Agustin Cozzetti, Istituto Superiore Mario Boella - Torino, Italy  
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, University Paris EST-UPEC/LISSI Lab - Creteil, France / F'SATIE/TUT, Pretoria, South Africa  
Péter Ekler, Budapest University of Technology and Economics, Hungary  
Karim El Defrawy, Hughes Research Labs (HRL) - Malibu, USA  
Wael M. El-Medany, University of Bahrain, Bahrain  
Lamia Fattouh, King Abdul Aziz University - Jeddah, Saudi Arabia  
Armando Ferro Vázquez, Universidad del País Vasco / Euskal Herriko Unibertsitatea - Bilbao, Spain  
Ana-Belén García-Hernando, Universidad Politécnica de Madrid, Spain  
Sorin Georgescu, Ericsson Research, Canada  
Carlo Giannelli, University of Bologna, Italy  
Michele Girolami, ISTI-CNR, Italy  
Xiang Gui, Massey University, New Zealand  
Guangjie Han, Hohai University, Japan  
Gerhard Hancke, Royal Holloway, University of London, UK  
Mohamad Sayed Hassan, Institut Telecom / Telecom Bretagne, France  
Hani Ragab Hassen, University of Kent at Canterbury, UK  
Vicente Hernández, Universidad Politecnica de Madrid, Spain  
Clemens Holzmann, Upper Austria University of Applied Sciences, Austria  
Muhammad Ismail, University of Waterloo, Canada  
Antonio Jara, University of Murcia, Spain  
Ali Jemmali, École Polytechnique de Montréal, Canada  
Ying Jian, Google, Inc., USA  
Magnus Jonsson, Halmstad University, Sweden

Ghassan Ali Kbar, King Saud University - Riyadh, Saudi Arabia  
Adrian Kacso, University of Siegen, Germany  
Mohamed Abd Rabou Kalil, Ilmenau University of Technology, Germany  
György Kálmán, ABB AS - Akershus, Norway  
Yacine Khaled, Geenov, France  
Zeashan Khan, GIPSA Lab - Grenoble, France  
Hoon Ko, GECAD / ISEP/IPP-Institute of Engineering-Polytechnic of Porto, Portugal  
Dragana Krstic, University of Nis, Serbia  
Zhihua Lai, Ranplan Wireless Network Design Ltd., UK  
Juhani Latvakoski, VTT Technical Research Centre of Finland, Finland  
Jong-Hyouk Lee, INRIA, France  
Jingli Li, TopWorx - Emerson, USA  
Qilian Liang, University of Texas at Arlington, USA  
Fidel Liberal, University of the Vascon Country, Spain  
Antonio Liotta, Eindhoven University of Technology, The Netherlands  
Justin Lipman, Intel R&D China, China  
Donggang Liu, The University of Texas - Arlington, USA  
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain  
Andreas Löffler, Friedrich-Alexander University of Erlangen-Nürnberg, Germany  
Jonathan Loo, Middlesex University - Hendon, UK  
Valeria Loscrí, University of Calabria, Italy  
Friederike Maier, Leibniz Universität Hannover, Germany  
Jamshaid Sarwar Malik, KTH Royal Institute of Technology, Sweden  
D. Manivannan (Mani), University of Kentucky - Lexington, USA  
Catherine Meadows, Naval Research Laboratory - Washington DC, USA  
Hamid Menouar, QUWIC - Qatar University Wireless Innovation Center, Qatar  
Mohamed Moustafa, Akhbar El Yom Academy, Egypt  
Yi Mu, University of Wollongong, Australia  
Subhas Chandra Mukhopadhyay, Massey University (Turitea) - Palmerston North, New Zealand  
Wolfgang Narzt, Johannes Kepler University - Linz, Austria  
Marek Neruda, Czech Technical University in Prague, Czech Republic  
Georg Neugebauer, RWTH Aachen University, Germany  
Christopher Nguyen, Intel Corp., USA  
Hodayoun Nikookar, Delft University of Technology, The Netherlands  
Gabriele Oligeri, ISTI-CNR, Italy  
Tudor Palade, Technical University of Cluj-Napoca, Romania  
Carlos Enrique Palau Salvador, Polytechnic University of Valencia, Spain  
Krassie Petrova, Auckland University of Technology, New Zealand  
Salvatore Flavio Pileggi, Universidad Politécnica de Valencia, Spain  
Thomas Plos, TU-Graz, Austria  
Anastasios Politis, University of Macedonia - Thessaloniki, Greece  
Neeli R. Prasad, Aalborg University, Denmark  
Yusnita Rahayu, Universiti Malaysia Pahang (UMP), Malaysia  
Tomi Rätty, VTT Technical Research Centre of Finland - Oulu, Finland  
Joel Rodrigues, University of Beira Interior, Portugal  
Imed Romdhani, Edinburgh Napier University, UK  
Teng Rui, NICT, Japan  
Sattar B. Sadkhan, University of Babylon, Iraq  
José A. Sánchez Fernández, Universidad Politécnica de Madrid, Spain  
David Sánchez Rodríguez, University of Las Palmas de Gran Canaria (ULPGC), Spain  
José Santa Lozano, Universidad de Murcia, Spain  
Riccardo Scopigno, Istituto Superiore Mario Boella - Torino, Italy  
Jean-Pierre Seifert, Technische Universität Berlin & Deutsche Telekom Laboratories - Berlin, Germany

Sandra Sendra Compte, Polytechnic University of Valencia, Spain  
Adão Silva, University of Aveiro / Institute of Telecommunications, Portugal  
Mariusz Skrocki, Orange Labs - Warszawa, Poland  
Vahid Solouk, Urmia University of Technology, Iran  
Himanshu B. Soni, G H Patel College of Engineering & Technology, India  
Kuo-Feng Ssu, National Cheng Kung University, Taiwan  
Petr Svenda, Masaryk University - Brno, Czech Republic  
Fatma Tansu Hocanin, Eastern Mediterranean University, Turkey  
José Miguel Villalón Millan, University of Castilla La Mancha, Spain  
Natalija Vlajic, York University - Toronto, Canada  
Naoki Wakamiya, University of Osaka, Japan  
Qian Wei, Beijing University of Technology - Beijing, P. R. China  
Yean-Fu Wen, National Chiayi University, Taiwan  
Ouri Wolfson, University of Illinois at Chicago, USA  
Qishi Wu, University of Memphis, USA  
Pei Xiao, Queen's University Belfast, UK  
Xiaochun Xu, University of Florida, USA  
Andrea Zanda, Universidad Politécnica - Madrid, Spain  
Sherali Zeadally, University of the District of Columbia, USA  
Yuanyuan Zeng, Wuhan University, China  
Yan Zhang, IMEC- NL, The Netherlands  
Yu Zhang, Orange Labs Beijing, China  
Gang Zhao, Digital Technology Laboratory - DMG/Mori Seiki, USA

## 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

Efficient Rate Adaptive Resource Allocation Scheme in Uplink OFDMA Wireless Systems <i>Vasileios Papoutsis and Stavros Kotsopoulos</i>	1
Resource Allocation Algorithm for MISO-OFDMA Systems with QoS Provisioning <i>Vasileios Papoutsis and Stavros Kotsopoulos</i>	7
Scrambling-Exact Channel Estimation in W-CDMA Systems <i>Wolfgang U. Aichmann</i>	12
The Mathematical Relationship Between Maximum Access Delay and the R.M.S Delay Spread <i>Khalid Hassan, Tharek Abd Rahman, M. Ramlee Kamarudin, and Faizah Nor</i>	18
An Effective transmit Weight Design for DPC with Maximum Beam in Multi-user MIMO Downlink <i>Cong Li and Yasunori Iwanami</i>	24
MLD-based MFSK Demodulation on MIMO Frequency Selective Fading Channel <i>Yasunori Iwanami and Kenji Nakayama</i>	30
The Impact of Antenna Directivity and Channel Bandwidth on the Power Spectral Density of Wideband and UWB MISO Channels <i>Ana-Maria Pistea, Tudor Palade, Ancuta Moldovan, and Hamidreza Saligheh Rad</i>	36
Angular CMA: A modified Constant Modulus Algorithm providing steering angle updates <i>Koen C. H. Blom, Marcel D. van de Burgwal, Kenneth C. Rovers, Andre B. J. Kokkeler, and Gerard J. M. Smit</i>	42
A Smart Predictive Link Layer Trigger Algorithm to Optimize Homogenous/Heterogeneous Networks WiFi Handover Decisions <i>Chunmei Liu and Christian Maciocco</i>	48
Exploiting Movement synchronization to Increase End-to-end file Sharing efficiency for Delay Sensitive Streams in Vehicular P2P Devices <i>Constandinos Mavromoustakis</i>	53
Vertical Fast Handoff in Integrated WLAN and UMTS Networks <i>Vahid Solouk, Borhanuddin Mohd Ali, and Daniel Wong</i>	59
A Hop-Oriented Position Estimation Algorithm for Wireless Sensor Networks <i>Hsing-Lung Chen, Li-Chung Lin, and Shu-Hua Hu</i>	65
Investigation of the Correlation Effect on the Performance of V-BLAST and OSTBC MIMO Systems <i>Ali Jemmali and Jean Conan</i>	71

Simple Detection Schemes for the Alamouti Code Assisted V-BLAST (ACAV) System <i>David Wee-Gin Lim</i>	75
Radiation Pattern Behaviour of Reconfigurable Asymmetry Slotted Ultra Wideband Antenna <i>Yusnita Rahayu, Azlyna Senawi, and Razali Ngah</i>	81
Radio Resource Unit Allocation and Rate Adaptation in Filter Banks Multicarrier System <i>Faouzi Bader and Mohammad Banat</i>	87
An Experimental Study of ICI Cancellation in OFDM Utilizing GNU Radio System <i>Miftahur Rahman, Sajjadul Latif, and Tasnia Ahad</i>	94
Dynamic Group PCI Assignment Scheme <i>Xu Zhang, DiBin Zhou, Zhu Xiao, Enjie Liu, Jie Zhang, and Andres Alayon Glasunov</i>	101
Performance Analysis of Pilot Patterns in Coded OFDMA Systems over Multi-user Channels <i>Jihyung Kim, Jung-Hyun Kim, Kwang Jae Lim, and Dong Seung Kwon</i>	107
A System Design for The Integration of RFID Systems with Wireless Network Technologies <i>Mehmet Erkan Yuksel and Asim Sinan Yuksel</i>	112
The Selected Propagation Models Analysis of Usefulness in Container Terminal Environment <i>Slawomir Ambroziak and Ryszard Katulski</i>	118
Bandwidth Allocation Algorithms for VOIP Networks: An Experimentation System and Evaluation of Created Algorithm <i>Rafal Orłowski, Michal Chamow, Iwona Pozniak-Koszalka, Leszek Koszalka, and Andrzej Kasprzak</i>	122
On the Receiver Performance in MU-MIMO Transmission in LTE <i>Zijian Bai, Biljana Badic, Stanislaus Iwelski, Tobias Scholand, Rajarajan Balraj, Guido Bruck, and Peter Jung</i>	128
A Hybrid Solution For Coverage Enhancement in Directional Sensor Networks <i>M. Amac Guvensan and A. Gokhan Yavuz</i>	134
Amplifying Matrix Design for Cooperative Relay Networks under Channel Uncertainty and Power Constraint <i>Kanghee Lee, Hyuck Kwon, Yanwu Ding, Zuojun Wang, Yu Bi, and Yazan Ibdah</i>	139
Finite-State Markov Chain Approximation for Geometric Mean of MIMO Eigenmodes <i>Ping-Heng Kuo and Pang-An Ting</i>	145
An Algorithmic Approach for Analyzing Wireless Networks with Retrials and Heterogeneous Servers <i>Nawel Gharbi and Leila Charabi</i>	151

---

Performance Analysis of an Ant-based Routing Algorithm with Enhanced Path Maintenance for MANETs <i>Miae Woo</i>	157
Adaptive Cross Layer Approach for Video Transmission Over Cognitive UWB Network <i>Norazidah Mohd Aripin, Norsheila Fisal, Rozeha A. Rashid, and Anthony C.C. Lo</i>	163
A Survey on Multi-Channel Based, Digital Media- Driven 802.11 Wireless Mesh Networks <i>Christian Kobel, Walter Baluja Garcia, and Joachim Habermann</i>	169
A Bidirectional Semi-Passive DS-CDMA-RFID Transponder for the UHF Band <i>Andreas Loeffler and Ingo Altmann</i>	176
Efficient Symbol Detector for MIMO Communication Systems <i>Soohyun Jang and Yunho Jung</i>	182
Software Defined Radio as a DS-CDMA Receiver <i>Ryszard Katulski, Wojciech Siwicki, and Jacek Stefanski</i>	188
Performance Improvements of Mobile SCTP during handover period <i>Sinh Chung Nguyen, Thi Mai Trang Nguyen, and Guy Pujolle</i>	193
Sink-Connected Barrier Coverage Optimization for Wireless Sensor Networks <i>Yung-Liang Lai and Jehn-Ruey Jiang</i>	198
The Joint Probability Density Function of the SSC Combiner Output Signal at Two Time Instants in the Presence of Hoyt Fading <i>Dragana Krstic, Petar Nikolic, Goran Stamenovic, and Mihajlo Stefanovic</i>	204
Hybrid Log-MAP Algorithm for Turbo Decoding Over AWGN Channel <i>Li Li Lim and David Wee Gin Lim</i>	211
Evaluating Transceiver Power Savings Produced by Connectivity Strategies for Infrastructure Wireless Mesh Networks <i>Pragasen Mudali, Murimo Mutanga, Matthew Adigun, and Ntsibane Ntlatlapa</i>	215
Spectrum Access during Cognitive Radio Mobiles' Handoff <i>Emna Trigui, Moez Esseghir, and Leila Merghem Boulahia</i>	221
Efficiently Using Bitmap Floorplans for Indoor Navigation on Mobile Phones <i>Martin Werner</i>	225
Near Field Sensing in Wireless LANs <i>Wolfgang Narzt and Heinrich Schmitzberger</i>	231

Spectrum Sensing Measurement using GNU Radio and USRP Software Radio Platform <i>Rozeha A. Rashid, Mohd Adib Sarijari, Norsheila Fisal, Anthony Lo, Sh. Kamilah S. Yusof, and Nurhija Mahalin</i>	237
Secure, Privacy-Preserving, and Context-Restricted Information Sharing for Location-based Social Networks <i>Michael Durr, Philipp Marcus, and Kevin Wiesner</i>	243
Seamless Multimedia Handoff for Hierarchical Mobile IPv6 <i>Khaled Zeraoulia and Nadjib Badache</i>	249
Heterogeneous Networks Handover Decision Triggering Algorithm Based on Measurements Messages Transfer using IP Option Header <i>Malak Z. Habeib, Hussein A. Elsayed, Salwa H. Elramly, and Magdy M. Ibrahim</i>	254
The Impacts of Dynamic Broadcast Schemes on the Performance of <i>Dimitrios Liarokapis and Ali Shahrabi</i>	260
Cyclostationary Detection in Spectrum Pooling System of Undefined Secondary Users <i>Nazar Radhi, Kahtan Aziz, Rafed Sabbar Abbas, and Hamed AL-Raweshidy</i>	266
Restore Partitioning in MANETs with Dynamically Deployed Helping Hosts — Message Cabs (MCabs) <i>Ting Wang and Chor Ping Low</i>	271
Increasing the Speed of a Wireless Network by Processing Indoor <i>Alexey Lagunov</i>	277
Thresholds Determination for new Backoff Algorithm in MANETs <i>Muneer Bani Yassein, Wail Mardini, and Zainab AbuTaye</i>	285
An Interference Avoiding Wireless Network Architecture for Coexistence of CDMA 2000 1x EVDO and LTE Systems <i>Xinsheng Zhao, Jia You, and Tao Jin</i>	289
Challenge Token-based Authentication – CTA <i>Ghassan Kbar</i>	294
Efficient Privacy Preservation Protocol Using Self-certified Signature For VANETS <i>Bidi Ying, Dimitrios Makrakis, and Hussein T. Mouftah</i>	301
Security Enhanced Authentication Protocol for UHF Passive RFID System <i>Suna Choi, Sangyeon Lee, and Heyungsub Lee</i>	307

# Efficient Rate Adaptive Resource Allocation Scheme in Uplink OFDMA Wireless Systems

Vasileios D. Papoutsis and Stavros A. Kotsopoulos

Wireless Telecommunications Laboratory, Department of Electrical and Computer Engineering,  
University of Patras, Rio, Greece 265 00,

Email: vpapoutsis@ece.upatras.gr, kotsop@ece.upatras.gr

**Abstract**—The problem of resource allocation for the uplink of wireless SISO-OFDMA systems is investigated. To relieve heavy computational burden, a suboptimal, but efficient scheme is devised which maximizes the sum of the users' data rates subject to constraints on the per user transmitted power and minimum data rate requirements among users. Simulation results indicate that the proposed scheme can satisfy minimum data rate constraints, distributing sum data rate fairly and flexibly among users. In addition, the proposed scheme is complexity effective, and can be applied to latest-generation wireless systems that provide Quality-of-Service (QoS) guarantees.

**Index Terms**—OFDMA, resource allocation, multiuser diversity.

## I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) [1] has developed into a popular scheme for wideband wireless digital communication. In OFDMA, multiple access is achieved by first dividing the spectrum of interest into a number of subcarriers and then assigning subsets of the subcarriers to individual users. OFDMA helps exploit multiuser diversity in frequency-selective channels, since it is very likely that some subcarriers that are “bad” for a user are “good” for at least one of the other users. Because of its superior performance in frequency-selective fading wireless channels, OFDMA is the modulation and multiple access scheme used in latest wireless systems such as IEEE 802.16e (Mobile WiMAX).

There are fixed and adaptive allocations to allocate subcarriers. Fixed allocations use Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA) as multiple access schemes to allocate each user a predetermined time slot or frequency band for transmission. While applying fixed allocation the system neglects the channel diversity and does not use the deep faded subcarriers for other users which do not seem as deep faded to them. In [2], these two fixed allocation schemes are discussed and compared in much detail. On the other hand, adaptive allocations allocate resources to users based on their channel gains. Due to the time-varying nature of the wireless channel, dynamic resource allocation makes full use of the multiuser diversity to achieve higher performance.

Most work on resource allocation has been done for the downlink OFDMA systems. In [3] [4], total transmit power is minimized. In [5] [6], it is proved that the downlink system capacity is maximized when each subcarrier is exclusively

assigned to the user with the best subcarrier gain, eliminating the intra-cell interference (ICI), and power is then distributed by the water-filling algorithm [7]. In [8], the minimum data rate among users is maximized and in [9]-[12], a proportional fairness criterion is employed. In [13] [14], the fulfillment of every user's data rate constraints is guaranteed and in [15], per time-slot resource allocation is introduced with “self-noise” and phase noise. In [16], long term access proportional fairness is introduced. Finally, in [17], the sum of the users' data rates is maximized but the resource allocation unit is not the subcarrier, as in previous algorithms [3]-[16], but a time/frequency unit (slot), in accordance with WiMAX systems.

Recently, uplink resource allocation has received some attention in literature. A practical low-complexity algorithm for a two-user case is proposed in [18]. In [19], the optimality of OFDMA in uplink transmissions has been studied while in [20], a non-iterative and near-optimal joint subcarrier and power allocation scheme is proposed. In [21], the results of [20] are generalized by considering the utility maximization in one time-slot, where the utility is a function of the instantaneous data rate in this time-slot. Another work that focused on per time-slot fairness is [22]. In [23], the uplink resource allocation problem is approached using a gradient scheduler but considers long-term total utility maximization which depends on the average data rate or queue sizes. Finally, in [24], resource allocation algorithms are proposed to find a Nash Bargaining solution according to Game theory.

In this paper, the resource allocation problem in uplink OFDMA systems is investigated. We focus on single antenna systems where at most one user can be assigned per subcarrier. The objective is to maximize the sum of the users' data rates subject to constraints on per user power and minimum data rates among users. The proposed scheme, which is also complexity effective, consists of three algorithms; an algorithm that determines the number of subcarriers for each user, a subcarrier allocation algorithm by dividing the users in two groups and the water-filling algorithm [7]. The first two algorithms assign the available subcarriers to the users of the system and the third one allocates the available power of each user.

The remainder of the paper is organized as follows. The problem of sum data rate maximization using minimum data rate constraints and power constraint for each user is formu-

lated in Section II. The proposed scheme is introduced in Section III and Section IV contains the complexity analysis of the proposed scheme and a complexity comparison with other schemes. Simulation results and a comparison between the proposed scheme and other existing schemes are provided in Section V. Finally, Section VI contains concluding remarks.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

The following assumptions are used in this paper: (i) the time-varying channels between different users and the Base Station (BS) are assumed to be frequency selective wireless channels with independent Rayleigh fading and the channel can be regarded as constant during the resource allocation period; (ii) the ISI is completely removed by exploiting OFDM techniques, *i.e.* the width of each subcarrier is much smaller than the coherence bandwidth of the channel. Thus, each user experiences flat fading in each subcarrier; (iii) the Channel State Information (CSI) is perfectly known by the receiver, and the BS feedbacks a certain form of channel information correctly to each user; (iv) each subcarrier can be used by only one user at each time.

The BS decides uplink resource transmission parameters for all available users based on the feedback CSI. The resource allocation parameters are then sent to each user through a dedicated control channel. Then, each user loads its data into the allocated subcarriers and the BS decodes the data sent from all users. The resource allocation scheme is updated as soon as the channel information is collected and also the resource allocation information is sent to BS for detecting.

Consider an OFDMA uplink transmission in a single cell with  $K$  active users and  $N$  subcarriers.  $P_k$  is the transmit power of each user  $k = 1, 2, \dots, K$  and the channel gain for user  $k$  in subcarrier  $n$  is denoted by  $g_{k,n}$ . Each subcarrier  $n$  of user  $k$  is assigned a power  $p_{k,n}$ . With the noise power spectral density being  $N_0$  and the total bandwidth of all subcarriers being  $B$ , the additive white noise power is  $\sigma^2 = \frac{N_0 B}{N}$ . Therefore, the subcarrier SNR can be expressed as  $h_{k,n} = \frac{g_{k,n}^2}{\sigma^2}$  and the transmitted SNR of user  $k$  in subcarrier  $n$  is  $\gamma_{k,n} = p_{k,n} h_{k,n}$ .

Each of the user's bits are modulated into  $N$  M-level QAM symbols, which are subsequently combined using the IFFT into an OFDMA symbol. For a square M-level QAM using Gray bit mapping as a function of transmitted SNR  $\gamma_{k,n}$  and number of bits of user  $k$  in subcarrier  $n$   $r_{k,n}$ , the BER can be approximated to within 1 dB for  $r_{k,n} \geq 4$  and  $\text{BER} \leq 10^{-3}$  as [25]

$$\text{BER}_{\text{MQAM}}(\gamma_{k,n}) \approx \frac{1}{5} \exp \left[ \frac{-1.6\gamma_{k,n}}{2^{r_{k,n}} - 1} \right] \quad (1)$$

By solving (1),  $r_{k,n}$  is

$$r_{k,n} = \log_2 \left( 1 + \frac{\gamma_{k,n}}{\Gamma} \right) = \log_2 (1 + p_{k,n} H_{k,n}) \quad (2)$$

where  $\Gamma = -\ln(5\text{BER})/1.6$  and  $H_{k,n} = \frac{h_{k,n}}{\Gamma}$  is the effective subcarrier SNR of user  $k$  in subcarrier  $n$ .

Taking into account the conclusions above, the optimization problem is formulated as:

$$\max_{c_{k,n}, p_{k,n}} \frac{B}{N} \sum_{k=1}^K \sum_{n=1}^N c_{k,n} r_{k,n} \quad (3)$$

subject to

$$c_{k,n} \in \{0, 1\}, \forall k, n \quad (4)$$

$$p_{k,n} \geq 0, \forall k, n \quad (5)$$

$$\sum_{k=1}^K c_{k,n} = 1, \forall n \quad (6)$$

$$\sum_{n=1}^N p_{k,n} \leq P_k, \forall k \quad (7)$$

$$\sum_{n=1}^N c_{k,n} r_{k,n} \geq m r_k, \forall k \quad (8)$$

where  $c_{k,n}$  is the subcarrier allocation indicator such that  $c_{k,n} = 1$  if subcarrier  $n$  is assigned to user  $k$  and  $c_{k,n} = 0$  if not. Constraints (4) and (5) ensure the correct values for the subcarrier allocation indicator and the power, respectively. Constraint (6) restricts the assignment of each subcarrier to only one user and (7) is the individual power constraint. Finally, (8) is the minimum data rate constraint. The total data rate for user  $k$ , denoted as  $R_k$ , is defined as

$$R_k = \frac{B}{N} \sum_{n=1}^N c_{k,n} r_{k,n} \quad (9)$$

where  $r_{k,n}$  is given by (2) and  $m r_k$  in (8) is the minimum data rate of each user.

Note that problem (3) is an NP-hard combinatorial optimization problem [26] with non-linear constraints. In a system with  $K$  users and  $N$  subcarriers, there are  $K^N$  possible subcarrier assignments, since it is assumed that no subcarrier can be used by more than one user. For a certain subcarrier assignment, a per user power distribution can be used to maximize the sum of the users' data rates, while guaranteeing minimum data rate constraints. The maximum sum data rate over all  $K^N$  subcarrier assignment schemes is the global maximum and the corresponding subcarrier assignment and per user power distribution is the optimal resource allocation scheme. However, it is difficult to obtain an optimal solution within any reasonable time frame. As a result, a novel and cost-effective resource allocation scheme is formulated to solve this problem.

## III. THE PROPOSED RESOURCE ALLOCATION SCHEME

Ideally, power of each user and subcarriers should be allocated jointly to solve optimization problem (3) optimally. This process has a prohibitive computational complexity. In the following, a suboptimal resource allocation scheme is proposed which consists of three algorithms and assures a low complexity performance:

### A. Number of subcarriers of each user.

In this algorithm, the number of subcarriers  $N_k$ , to be initially assigned to each user, is determined. This process is based on the average effective subcarrier SNR of each user which is calculated by

$$\bar{H}_k = \frac{1}{N} \sum_{n=1}^N H_{k,n}, \quad \forall k = 1, 2, \dots, K \quad (10)$$

The approximate data rate of each user is

$$\bar{R}_k = N_k \frac{B}{N} \log_2 (1 + \bar{H}_k \bar{P}_k), \quad \forall k = 1, 2, \dots, K \quad (11)$$

where  $\bar{P}_k$  is the equal power allocation of each user among respective subcarriers. At each iteration, the user with the minimum difference  $\bar{R}_k - mr_k$  has the option to be assigned one more subcarrier. When all the available subcarriers are assigned to  $K$  users of the system, the approximate number of subcarriers  $N_k$  for each user is got. This algorithm is as follows

#### 1) Initialization:

- Set  $mr_k, \forall k = 1, 2, \dots, K$ , the minimum data rate constraints.
- Set the initial number of subcarriers  $N_k = \lfloor N \frac{mr_k}{\sum_{k=1}^K mr_k} \rfloor, \forall k = 1, 2, \dots, K$  and  $N_{al} = \sum_{k=1}^K N_k$ .
- Get the average effective subcarrier SNR for each user using equation (10).

#### 2) Approximate data rate:

- Get the equal power to each allocated subcarrier  $\bar{P}_k = \frac{P_k}{N_k}, \forall k = 1, 2, \dots, K$ .
- Calculate  $\bar{R}_k, \forall k = 1, 2, \dots, K$ , using equation (11).

#### 3) While $N_{al} < N$ :

- Find  $k^* = \operatorname{argmin}_{k=1,2,\dots,K} \{\bar{R}_k - mr_k\}$ . For the found  $k^*$ , let  $N_{k^*} = N_{k^*} + 1$  and  $N_{al} = N_{al} + 1$ .
- Get the equal power to each allocated subcarrier  $\bar{P}_k = \frac{P_k}{N_k}, \forall k = 1, 2, \dots, K$ .
- Calculate  $\bar{R}_k, \forall k = 1, 2, \dots, K$ , using equation (11).

In initialization step,  $N \frac{mr_k}{\sum_{k=1}^K mr_k}$  is approximated to the lower integer because  $N_k$  should be an integer. Hence, it is not sure that  $N_{al} = N$ ; there might be some remaining subcarriers. That is the reason why step 3) of the algorithm is necessary.

### B. Subcarrier assignment to available users.

In this subsection the  $N_k, \forall k = 1, 2, \dots, K$ , subcarriers are allocated to available users in order to maximize the sum of the users' data rates while guaranteeing minimum data rates of  $K$  users. The algorithm is described below.

#### 1) Initialization:

- Set  $\mathcal{S} = \{1, 2, \dots, N\}, R_k = 0, \forall k = 1, 2, \dots, K, c_{k,n} = 0, \forall k = 1, 2, \dots, K$  and  $n \in \mathcal{S}$ .

- Sort the  $K$  users by average effective subcarrier SNR, i.e.,  $\bar{H}_1 \leq \dots \leq \bar{H}_m \leq \dots \leq \bar{H}_K$  without loss of generality.
- Divide the  $K$  users in bad effective subcarrier SNR group ( $\text{user}_b = \{1, 2, \dots, m\}$ ) and good effective subcarrier SNR group ( $\text{user}_g = \{m + 1, m + 2, \dots, K\}$ ).

#### 2) For $k = 1, 2, \dots, m$ :

- Find  $n$  satisfying  $H_{k,n} \geq H_{k,j}, \forall j \in \mathcal{S}$ .
- For the found  $n$ , set  $c_{k,n} = 1, N_k = N_k - 1, \mathcal{S} = \mathcal{S} - \{n\}$  and update  $R_k$  according to equation (9). In equation (9),  $p_{k,n} = \frac{P_k}{\sum_{n=1}^N c_{k,n}}$ .

#### 3) While $|\text{user}_b| \neq \emptyset$ :

- Find  $k^* = \operatorname{argmin}_{k \in \text{user}_b} \{R_k - mr_k\}$ .
- For the found  $k^*$ , if  $N_{k^*} > 0$ 
  - Find  $n$  satisfying  $H_{k^*,n} \geq H_{k^*,j}, \forall j \in \mathcal{S}$ .
  - Set  $c_{k^*,n} = 1, N_{k^*} = N_{k^*} - 1, \mathcal{S} = \mathcal{S} - \{n\}$  and update  $R_{k^*}$  according to equation (9). In equation (9),  $p_{k^*,n} = \frac{P_{k^*}}{\sum_{n=1}^N c_{k^*,n}}$ .
- else
  - $\text{user}_b = \text{user}_b - \{k^*\}$ .

#### 4) Redo:

- Steps 2,3 for the good effective subcarrier SNR group, i.e., for users  $\in \text{user}_g$ .

In step 1) of the subcarrier assignment to available users algorithm, all the variables are initialized.  $\mathcal{S}$  is the set of available subcarriers,  $c_{k,n}$  is the subcarrier allocation indicator and  $R_k$  is a vector which keeps track of the data rate of each user  $k = 1, 2, \dots, K$ . Then, users are divided in two groups according to parameter  $m$ ; the  $\text{user}_b$  and the  $\text{user}_g$ , the group of users with bad average effective subcarrier SNR and the group with good average effective subcarrier SNR respectively. Parameter  $m$  is chosen in such a way that the two user groups contain the same number of users, if  $K$  is an even number. Otherwise, if  $K$  is an odd number, one of the two user groups would contain one more user than the other group.

In step 2), each user of  $\text{user}_b$  group is assigned the available subcarrier on which he has the largest effective subcarrier SNR. Note that an inherent advantage is gained by the fact that users of  $\text{user}_b$  group choose their best subcarrier earlier than the users of the other group.

In step 3), subcarriers are assigned to available users until each user gets his allotment of  $N_k$  subcarriers. The user who has the minimum difference between its data rate and respective minimum data rate constraint has the priority to choose his best subcarrier. The best subcarrier is that on which he has the largest effective subcarrier SNR. The user, who gets his allotment of  $N_k$  subcarriers, can no longer be assigned any more subcarriers.  $|\text{user}_b|$  here denotes the cardinality of set  $\text{user}_b$ .

In step 4), the same procedure takes place but for the  $\text{user}_g$  group; the group of users with good average effective subcarrier SNR. The condition  $k = 1, 2, \dots, m$ , changes to  $k = m + 1, m + 2, \dots, K$ .

### C. Power allocation.

In subcarrier assignment to available users algorithm available power is distributed uniformly among subcarriers. In order to further enhance the sum of the users' data rates, in power allocation algorithm, the subcarrier allocation is kept, but the available power  $P_k$  of each user is assigned to subcarriers of each user using the water-filling [7] algorithm:

$$p_{k,n} = \left( \lambda_k - \frac{1}{H_{k,n}} \right)^+$$

where  $p_{k,n}$  is the allocated power in each subcarrier,  $(\cdot)^+ = \max(0, \cdot)$ , and  $\lambda_k$  satisfies

$$\sum_{n=1}^N p_{k,n} = P_k, \forall k = 1, 2, \dots, K.$$

### IV. COMPLEXITY ANALYSIS

In this section, the computational complexity of the proposed resource allocation scheme is analyzed and compared with that of [5], [20]-[22]. Recall that  $K$  refers to the total number of users in the system and  $N$  refers to the number of subcarriers, which is much larger than  $K$ . As mentioned in Section II, for the exhaustive search algorithm, there are  $K^N$  possible subcarrier assignments which require  $O(K^N)$  time.

Initialization step of the first algorithm of the proposed scheme requires  $K$  multiplications to set the initial number of subcarriers  $N_k$ , and also average effective subcarrier SNR is calculated  $K$  times. Thus, the complexity of this initialization step is  $O(K)$ . In second step of the same algorithm,  $\bar{R}_k$  is calculated  $K$  times which is  $O(K)$ . In third step, the user with the minimum  $\bar{R}_k - mr_k$  among  $K$  users is found and  $\bar{R}_k$  is calculated for  $k = 1, 2, \dots, K$ . This is repeated until  $N_{al} = N$ . Thus, this step requires  $O(K(N - N_{al}))$  complexity which is also the overall complexity of the first algorithm of the proposed resource allocation scheme.

In initialization step of the second algorithm of the proposed resource allocation scheme,  $K$  users are sorted by average effective subcarrier SNR which has  $O(K \log_2 K)$  complexity. Then the complexity of the division of  $K$  users in two groups is  $O(K)$ . Thus, the complexity of this initialization step is  $O(K \log_2 K)$ . In the second step, for each user of one group, the best subcarrier is found which has complexity  $O(KN)$  because in our simulations the two groups contain equal number of users, i.e.,  $m = \frac{K}{2}$ . In step three of this algorithm, subcarriers are allocated to users of one group until each user gets his allotment of  $N_k$  subcarriers. In worst case scenario, the complexity of this step is  $O(KN)$ . In step four, the same procedure takes place but for the other group of users. Thus, because  $K \ll N$  and  $\log_2 K \ll N$  the complexity of the second algorithm of the proposed resource allocation scheme is  $O(KN)$ .

Finally, in the third algorithm of the proposed resource allocation scheme, after subcarrier allocation is found, the water-filling power allocation algorithm is implemented which requires to find  $\lambda_k$ . The update of  $\lambda_k$  can be done by using a

simple bisection method until the power of each user converges [27]. Thus, in order to perform water-filling power allocation for each user the overall time complexity is  $O(KN)$ . Consequently the complexity of the proposed resource allocation scheme is  $O(K(N - N_{al}) + KN + KN) \approx O(KN)$ .

In the resource allocation algorithm proposed in [5], each subcarrier is allocated to the user with the maximum effective subcarrier SNR. Then, after the subcarrier allocation, either water-filling [7] or equal power allocation is applied for each user. Thus, the complexity of [5] is  $O(KN)$ . Algorithm in [5] is optimal in the downlink scenario but not in the uplink because in the latter there are individual power constraints.

In algorithms proposed in [20] [21],  $N$  iterations are required to allocate  $N$  subcarriers to available users. In each iteration, water-filling [7] is performed for each user with time complexity  $O(N)$ . This means that the time complexity of one iteration is  $O(KN)$  and for all iterations is  $O(KN^2)$ . In [21], a fast implementation method is introduced which is based on binary tree data structure and has  $O(KN \log_2 N)$  time complexity but it requires greater storage memory in order to store all the required information in each node of the binary tree.

The algorithm proposed in [22] consists of the step of initial subcarrier allocation and the step of residual subcarrier allocation. For both steps the complexity is  $O(KN)$  as described extensively in [22].

It is easily observed that the proposed resource allocation scheme has a very dramatic reduction in complexity compared to  $O(K^N)$  required by the exhaustive search. In addition it has similar complexity to [5] [22] and smaller than [20] [21], without using the binary search tree introduced in [21].

### V. SIMULATION RESULTS

In this section, the performance of the proposed uplink resource allocation scheme is evaluated using simulation. In all simulations presented in this section, the frequency-selective channel consists of six independent Rayleigh multipath components (taps). As in [9], an exponentially decaying power delay profile is assumed, the ratio of the energy of the  $l$ th tap to the first tap being equal to  $e^{-2l}$ . For each channel realization the proposed scheme is used to perform resource allocation, and the data rates for each user are computed. A maximum delay spread of  $5 \mu\text{s}$  and maximum doppler of 30 Hz is assumed. The channel information is sampled every 0.5 ms to update the resource allocation. As in [9], the total available bandwidth is equal to  $B = 1$  MHz, the number of subcarriers of an OFDM symbol is  $N = 64$ , variance of the additive noise is equal to  $N_0 = -80$  dB-W/Hz (single-sided PSD), and BER =  $10^{-7}$ . Minimum data rate constraints are  $mr_k = 1$  bit/s/Hz for  $k = 1, 2, \dots, K$ , the number of channel realizations is equal to  $10^5$  and parameter  $m = \frac{K}{2}$ .

The proposed resource allocation scheme is compared with the algorithms proposed in [5] (Jang et al.), [20] (Kim et al.), [21] (Ng et al.), [22] (Gao et al.), and a static TDMA scheme. In Figs. 1, 3, 5 the number of users of the system varies from 2 - 8 in increment of 2 and total transmitted power of each

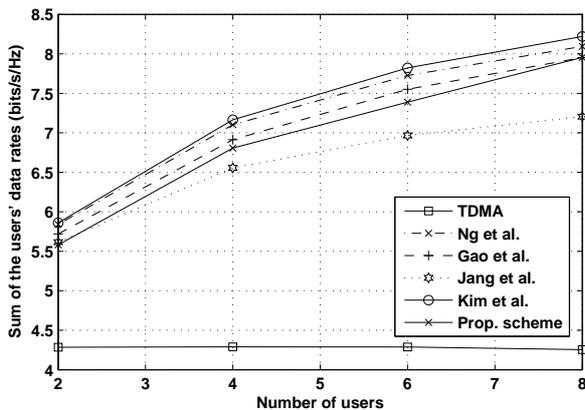


Fig. 1. Sum of the users' data rates vs number of users.

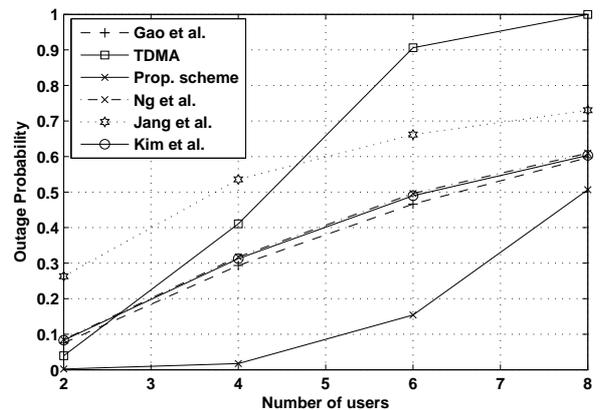


Fig. 3. Outage probability vs number of users.

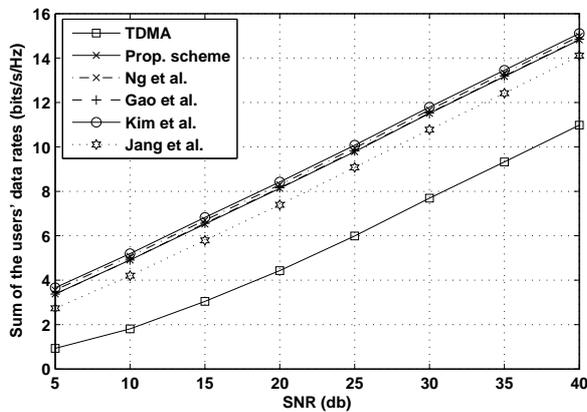


Fig. 2. Sum of the users' data rates vs SNR(db).

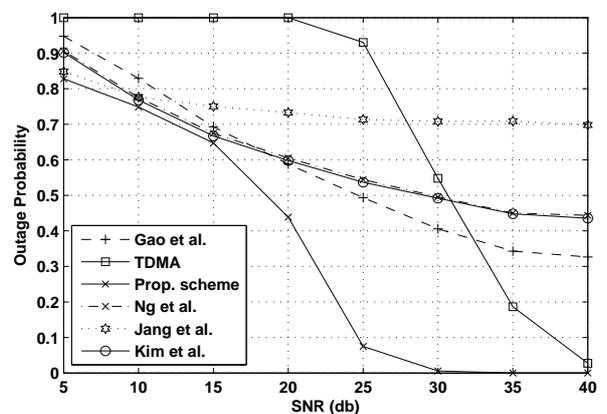


Fig. 4. Outage probability vs SNR(db).

user is equal to  $P_k = 1$  W, while in Figs. 2, 4, 6 the number of users is  $K = 8$  and SNR varies from 5 - 40 in increment of 5.

Figs. 1, 2 depict the comparison of the sum of the users' data rates versus number of users and SNR, respectively. It can be seen, the reasonable price being paid in order to guarantee minimum data rates by using the proposed scheme. In Fig. 1, as the number of users increases, the difference in sum data rates also increases because additional multiuser diversity is available to [20] [21] that only target sum data rate maximization. On the other hand, more users put more constraints to the proposed scheme, because new users need to share the same resources. In addition, sum data rate of the proposed scheme is significantly enhanced over both [5] and static TDMA algorithm as can be seen in Figs. 1, 2.

Figs. 3, 4 depict the comparison of outage probability versus number of users and SNR, respectively. The outage probability of the proposed resource allocation scheme is significantly smaller than any of the other comparing algorithms. This point is very critical in real systems where users should satisfy the minimum data rate criterion.

Figs. 5, 6 depict the comparison of the fairness pointer

versus number of users and SNR, respectively. Fairness pointer  $F$  is the one introduced in [9], and is defined as

$$F = \frac{(\sum_{k=1}^K R_k)^2}{K \sum_{k=1}^K (R_k)^2},$$

where  $F$  is a real number in the interval  $(0, 1]$  with the maximum value of 1 for the case when equal data rates are achieved among users. As can be seen in Figs. 5, 6, users' data rates are almost equal when the proposed resource allocation scheme and static TDMA algorithm are employed with the proposed scheme being more fair. Algorithm in [5] guarantees the least fairness between the users' data rates. Algorithms in [20] [21] guarantee almost the same fairness but it is much lower than that of the proposed scheme and algorithm in [22] guarantees improved fairness than [5] [20] [21] but noticeably less than the proposed scheme. In addition, it can be seen in Fig. 5 that as the number of users increases the fairness pointer increases in all algorithms except the proposed scheme and static TDMA where fairness pointer remains almost constant regardless of the increasing number of users.

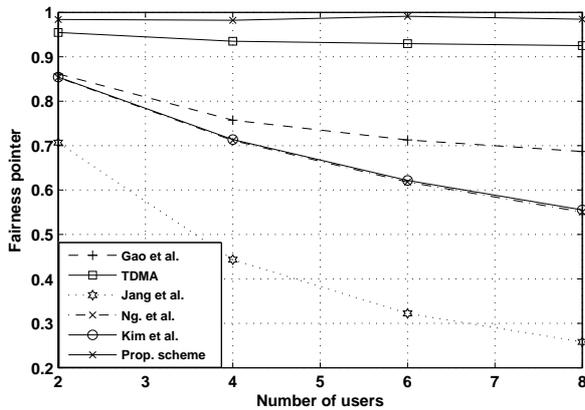


Fig. 5. Fairness pointer vs number of users.

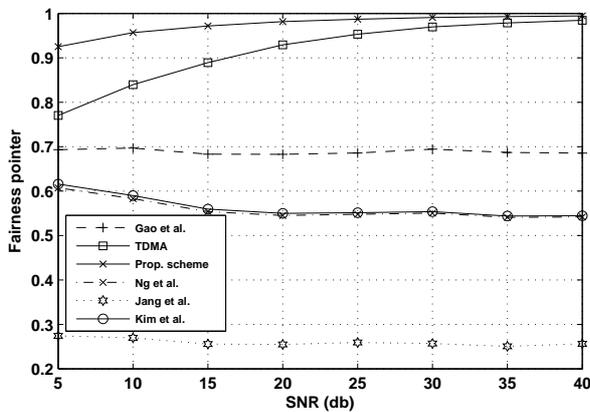


Fig. 6. Fairness pointer vs SNR(db).

## VI. CONCLUSION

A resource allocation scheme for the uplink of SISO-OFDMA wireless systems was introduced which maximizes sum data rate of the system's users while guaranteeing minimum data rates  $mr_k$  among users. It is also complexity efficient and consists of three algorithms; The first algorithm determines the number of subcarriers to be initially assigned to each user, the second algorithm assigns the subcarriers to each user, and the third one allocates the available power  $P_k$  of each user to subcarriers, using the water-filling equation [7]. In addition, its innovative priority scheduling exploits more efficiently the multiuser diversity and makes it perform better than previous schemes. Finally, sum of the users' data rates is distributed more fairly among users.

## REFERENCES

- [1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Prentice Hall, Mar. 2007.
- [2] H. Rohling and R. Grunheid, "Performance comparison of different multiple access schemes for the downlink of an OFDM communication system," *IEEE Vehicular Techn. Conf.*, vol. 3, pp. 1365-1369, Phoenix, USA, May 1997.
- [3] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.
- [4] D. Kivanc, G. Li, and H. Liu, "Computationally efficient bandwidth allocation and power control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150-1158, Nov. 2003.
- [5] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171-178, Feb. 2003.
- [6] G. Li and H. Liu, "On the Optimality of the OFDMA Network," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 438-440, May 2005.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information theory*, John Wiley & Sons, Inc., Hoboken, New Jersey, Jul. 2006.
- [8] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," *IEEE Vehicular Techn. Conf.*, vol. 2, pp. 1085-1089, Tokyo, Japan, May 2000.
- [9] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726-2737, Nov. 2005.
- [10] V. D. Papoutsis, I. G. Fraimidis, and S. A. Kotsopoulos, "Fairness-aware resource allocation for the SISO downlink over frequency-selective channels," *IEEE Wireless Commun. and Networking Conf.*, Sydney, NSW, Australia, Apr. 2010.
- [11] S. Sadr, A. Anpalagan, and K. Raahemifar, "Suboptimal Rate Adaptive Resource Allocation for Downlink OFDMA Systems," *Hindawi Intern. J. Veh. Techn.*, vol. 2009, Article ID 891367, 10 pages, 2009.
- [12] C. Mohanram and S. Bhashyam, "A Sub-optimal Joint Subcarrier and Power Allocation Algorithm for Multiuser OFDM," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 685-687, Aug. 2005.
- [13] Y. J. Zhang and K. B. Letaief, "Multiuser adaptive subcarrier-and-bit allocation with adaptive cell selection for OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1566-1575, Sep. 2004.
- [14] Z. Mao and X. Wang, "Efficient Optimal and Suboptimal Radio Resource Allocation in OFDMA System," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 440-445, Feb. 2008.
- [15] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink Scheduling and resource Allocation for OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288-296, Jan. 2009.
- [16] Y. Ma, "Rate-Maximization for Downlink OFDMA with Proportional Fairness," *IEEE Trans. Veh. Techn.*, vol. 57, no. 5, pp. 3267-3274, Sep. 2008.
- [17] A. Biagioni, R. Fantacci, D. Marabissi, and D. Tarchi, "Adaptive Subcarrier Allocation Schemes for Wireless OFDMA Systems in WiMAX Networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 217-225, Feb. 2009.
- [18] W. Yu and J. M. Cioffi, "FDMA Capacity of Gaussian Multiple-Access Channels With ISI," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 102-111, Jan. 2002.
- [19] H. Li and H. Liu, "An Analysis of Uplink OFDMA Optimality," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2972-2983, Aug. 2007.
- [20] K. Kim, Y. Han, and S. Kim, "Joint Subcarrier and Power Allocation in Uplink OFDMA Systems," *IEEE Commun. Lett.*, vol. 9, no. 6, pp. 526-528, June 2005.
- [21] C. Ng and C. Sung, "Low Complexity Subcarrier and Power Allocation for Utility Maximization in Uplink OFDMA Systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1667-1675, May 2008.
- [22] L. Gao and S. Cui, "Efficient Subcarrier, Power, and Rate Allocation with Fairness Consideration for OFDMA Uplink," *IEEE Trans. Wireless Commun.*, vol. 7, no. 5, pp. 1507-1511, May 2008.
- [23] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry, "Joint Scheduling and Resource Allocation in Uplink OFDM Systems for Broadband Wireless Access Networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 226-234, Feb. 2009.
- [24] Z. Han, Z. Ji, and K. J. R. Liu, "Fair Multiuser Channel Allocation for OFDMA Networks Using Nash Bargaining Solutions and Coalitions," *IEEE Trans. Commun.*, vol. 53, no. 8, pp. 1366-1376, Aug. 2005.
- [25] S. T. Chung and A. Goldsmith, "Degrees of Freedom in Adaptive Modulation: A Unified View," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561-1571, Sep. 2001.
- [26] T. Ibaraki and N. Katoh, *Resource Allocation Problems: Algorithmic Approaches*, M. Garey, Ed. MIT Press, Mar. 1988.
- [27] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Mar. 2004.

# Resource Allocation Algorithm for MISO-OFDMA Systems with QoS Provisioning

Vasileios D. Papoutsis and Stavros A. Kotsopoulos

Wireless Telecommunications Laboratory, Department of Electrical and Computer Engineering,  
University of Patras, Rio, Greece 265 00,

Email: vpapoutsis@ece.upatras.gr, kotsop@ece.upatras.gr

**Abstract**—The problem of user selection and resource allocation for the downlink of wireless systems operating over a frequency-selective channel is investigated. It is assumed that the Base Station (BS) uses many antennas, whereas a single antenna is available to each user (Multiple Input Single Output - MISO case). To relieve heavy computational burden, a suboptimal, but efficient algorithm is devised that is based on Zero Forcing (ZF) beamforming. The algorithm maximizes the sum of the users' data rates subject to constraints on total available power and individual data rate requirements for each user. Simulation results are provided to indicate that the algorithm can satisfy the fairness criterion. Thus, the algorithm can be applied to latest-generation wireless systems that provide Quality-of-Service (QoS) guarantees.

**Index Terms**—MISO, OFDMA, resource allocation, Zero-Forcing, minimum data rate constraints.

## I. INTRODUCTION

Orthogonal Frequency Division Multiple Access (OFDMA) [1] is a multi-user version of the popular Orthogonal Frequency Division Multiplexing (OFDM) digital modulation scheme. In OFDMA, multiple access is achieved by first dividing the spectrum of interest into a number of subcarriers and then assigning subsets of the subcarriers to individual users. OFDMA helps exploit multiuser diversity in frequency-selective channels, since it is very likely that some subcarriers that are “bad” for a user are “good” for at least one of the other users. Because of its superior performance in frequency-selective fading wireless channels, OFDMA is the modulation and multiple access scheme used in latest wireless systems such as IEEE 802.16e (Mobile WiMAX) [1].

In recent years, many dynamic resource allocation algorithms have been developed for the Single Input Single Output (SISO)-OFDMA systems. In [2] [3], the system throughput is maximized with a total power constraint and in [4]-[6], the total power consumption is minimized with constraints on the users' data rates. In [7], minimum data rate is maximized while in [8]-[10], proportional data rate constraints are introduced. In [11] [12], the fulfillment of every user's data rate constraints are guaranteed in order to maximize the sum of the users' data rates and in [13], the sum throughput is maximized with long term access proportional fairness. In addition, in [14] weighted sum data rate is maximized with “self-noise” and phase noise. Finally, in [15], system throughput is maximized but the resource allocation unit is not the subcarrier, as in previous algorithms [2]-[14], but a time/frequency unit (slot),

in accordance with WiMAX systems [1].

An additional major advance in recent wireless systems is the use of Multiple Input Multiple Output (MIMO) transmission to improve communication performance [1]. In fading environments MIMO technology offers significant increase in the data throughput and the link range without additional bandwidth or transmit power requirements by opening up multiple data pipes in the same frequency band of operation [16]. Because of these properties, MIMO systems have received increasing attention in the past decade. MIMO related algorithms can be implemented in each subcarrier and by combining OFDMA with MIMO transmission, wireless systems can offer larger system capacities and improved reliability.

In general, in order to transmit on the boundary of the capacity region, the BS needs to transmit to multiple users simultaneously in each subchannel employing Dirty Paper Coding (DPC) [16]. However, DPC has large implementation complexity. In [17] [18], user selection and beamforming algorithms, that are based on ZF [19], are proposed in order to maximize the system capacity without guaranteeing any kind of fairness among the users' data rates. In [20], proportional data rate constraints are applied and in [21] a kind of fairness is supported.

In this paper, a user selection and resource allocation algorithm for multiuser downlink MISO-OFDMA is developed that is less complex than exhaustive search algorithm and incorporates fairness by imposing minimum data rate constraints among users. As in [20], the beamforming scheme of [17] is applied in each subcarrier, where each user experiences flat fading [1], but the user selection procedure takes minimum data rate constraints into account. A complexity analysis is also presented in order to further support our statements and simulation results indicate that the algorithm can satisfy the fairness criterion.

The remainder of the paper is organized as follows. A description of the MISO-OFDMA system model is introduced in Section II, whereas the problem of sum data rate maximization using minimum data rate constraints is formulated in Section III. The proposed algorithm is introduced in Section IV and Section V contains the complexity analysis of the proposed algorithm and a complexity comparison with exhaustive search algorithm. Simulation results, analysis and a comparison between the proposed algorithm and previous resource allocation schemes are provided in Section VI. Finally, Section VII

contains concluding remarks.

In the following,  $(\cdot)^T$  denotes transpose, whereas  $(\cdot)^*$  denotes conjugate transpose.  $\mathbf{x}$  denotes a column vector,  $\mathbf{A}$  denotes a matrix,  $\|\cdot\|$  represents the Euclidean norm, and  $\mathbb{E}$  is the mean value. Finally,  $[x]_+ = \max\{0, x\}$ .

## II. SYSTEM MODEL

Consider an OFDMA downlink transmission with  $N$  subcarriers,  $T$  transmit antennas at the BS and  $K$  active users, each equipped with a single receive antenna. Also, let  $B$  be the overall available bandwidth, and  $\mathbf{h}_{k,n} = [h_{k,n}^1 \dots h_{k,n}^T]^T$  be the  $T \times 1$  baseband equivalent gain vector of the channel between the BS and user  $k$  in subcarrier  $n$ . Thus, for each subcarrier  $n$ , the baseband equivalent model for the system can be written as

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{x}_n + \mathbf{z}_n, \quad (1)$$

where  $\mathbf{H}_n = [\mathbf{h}_{1,n} \mathbf{h}_{2,n} \dots \mathbf{h}_{K,n}]^T$  is a  $K \times T$  matrix with complex entries,  $\mathbf{x}_n = [x_{1,n} \dots x_{T,n}]^T$  is the  $T \times 1$  transmitted signal vector in subcarrier  $n$ ,  $\mathbf{y}_n = [y_{1,n} \dots y_{K,n}]^T$  is a  $K \times 1$  vector containing the received signal of each user, and  $\mathbf{z}_n = [z_{1,n} \dots z_{K,n}]^T$  is a  $K \times 1$  vector denoting the noise that is assumed to be independent identically distributed (i.i.d.) zero-mean circularly symmetric complex Gaussian with covariance matrix  $\sigma^2 \mathbf{I}_K$ .

It is also assumed that the channel vectors are statistically independent and that their distribution is continuous. Hence,  $\text{rank}(\mathbf{H}_n) = \min(T, K)$ . Moreover, the practically important case where  $K \geq T$  is considered. Hence,  $\text{rank}(\mathbf{H}_n) = T$ . The total transmitted power, in the entire OFDM symbol, is  $P_{tot}$  and equal power is allocated to each subcarrier. Hence,  $\text{trace}[\mathbf{C}_n] \leq \frac{P_{tot}}{N}$ , where  $\mathbf{C}_n = \mathbb{E}[\mathbf{x}_n (\mathbf{x}_n)^*]$  is the covariance matrix of the transmitted signal  $\mathbf{x}_n$ .

Using only transmit beamforming, which is a suboptimal strategy, the following model is obtained. Let  $\mathbf{w}_{k,n} = [w_{k,n}^1 \ w_{k,n}^2 \ \dots \ w_{k,n}^T]^T$  be the  $T \times 1$  beamforming weight vector for user  $k$  in subcarrier  $n$ . Then, the baseband model (1) can be written as

$$\mathbf{y}_n = \mathbf{H}_n \mathbf{W}_n \mathbf{D}_n \mathbf{s}_n + \mathbf{z}_n, \quad (2)$$

where  $\mathbf{W}_n = [\mathbf{w}_{1,n} \ \mathbf{w}_{2,n} \ \dots \ \mathbf{w}_{K,n}]$  is the  $T \times K$  beamforming weight matrix,  $\mathbf{s}_n = [s_{1,n} \ \dots \ s_{K,n}]^T$  is a  $K \times 1$  vector containing the signals destined to each user, and  $\mathbf{D}_n = \text{diag}(\sqrt{p_{1,n}}, \sqrt{p_{2,n}}, \dots, \sqrt{p_{K,n}})$  accounts for the distribution of the power allocated to subcarrier  $n$  among the  $K$  users. According to (2), the resulting received signal vector for user  $k$  in subcarrier  $n$ , is given by

$$\begin{aligned} y_{k,n} &= \sum_{i=1}^K \mathbf{h}_{k,n} \mathbf{w}_{i,n} \sqrt{p_{i,n}} s_{i,n} + z_{k,n} = \\ &= \mathbf{h}_{k,n} \mathbf{w}_{k,n} \sqrt{p_{k,n}} s_{k,n} + \\ &+ \sum_{i=1, i \neq k}^K \mathbf{h}_{k,n} \mathbf{w}_{i,n} \sqrt{p_{i,n}} s_{i,n} + z_{k,n}, \end{aligned} \quad (3)$$

where the term in third line in (3) represents the multi-user interference caused by the simultaneous transmission of data

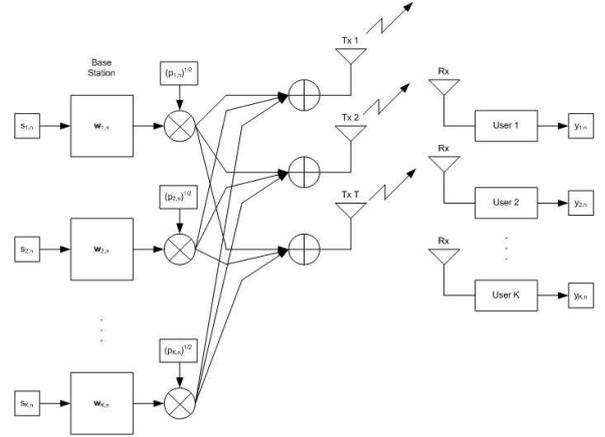


Fig. 1. MISO-OFDMA block diagram.

to other users in subcarrier  $n$ . Concerning (3), a graphic representation of the MISO downlink beamforming block diagram is shown in Fig. 1.

## III. PROBLEM FORMULATION

ZF beamforming is a spatial signal processing by which the multiple antenna transmitter can null multiuser interference signals in wireless communications. It inverts the channel matrix at the transmitter in order to create orthogonal channels between the transmitter and the receiver. The beamforming vectors are selected such that  $\mathbf{h}_{i,n} \cdot \mathbf{w}_{j,n} = 0$ , for  $i \neq j$ , and (3) becomes  $y_{k,n} = \mathbf{h}_{k,n} \mathbf{w}_{k,n} \sqrt{p_{k,n}} s_{k,n} + z_{k,n}$ . It is then possible to encode users individually, and with smaller complexity compared to DPC. ZF at the transmitter incurs an excess transmission power penalty relative to ZF-DPC and the (optimal) MMSE-DPC transmission scheme. If  $K \leq T$  and  $\text{rank}(\mathbf{H}_n) = K$ , the ZF beamforming matrix is the pseudo-inverse of  $\mathbf{H}_n$ , namely  $\mathbf{W}_n = \mathbf{H}_n^* (\mathbf{H}_n \mathbf{H}_n^*)^{-1}$ .

However, if  $K > T$ , it is not possible to use it because  $\mathbf{H}_n \mathbf{H}_n^*$  is singular and low complexity SDMA approaches are required. In that case, it is necessary to select  $t \leq T$  out of  $K$  users in each subcarrier. Hence, there are  $I$  possible combinations of users transmitting in the same subcarrier, denoted as  $A_i$ , where  $A_i \subset \{1, 2, \dots, K\}$ ,  $0 < |A_i| \leq T$ , where  $|A_i|$  denotes the cardinality of set  $A_i$ , and  $I = \sum_{l=1}^T \binom{K}{l}$ .

Let a set of users  $A_i = \{s_1, \dots, s_t\}$ , that produce the row-reduced channel matrix  $\mathbf{H}_n(A_i) = [\mathbf{h}_{s_1,n} \ \mathbf{h}_{s_2,n} \ \dots \ \mathbf{h}_{s_t,n}]^T$  in each subcarrier. When ZF is used, the data rate of user  $k \in A_i$ , in subcarrier  $n$ , is given by [16] [20]

$$r_{k,i,n} = \log_2(\mu_n c_{k,n}(A_i)), \quad (4)$$

where  $c_{k,n}(A_i) = \{[(\mathbf{H}_n(A_i) \mathbf{H}_n(A_i)^*)^{-1}]_{k,k}\}^{-1}$  and  $\mu_n$  is obtained by solving the water-filling equation [20]  $\sum_{k \in A_i} \left[ \mu_n - \frac{1}{c_{k,n}(A_i)} \right]_+ = \frac{P_{tot}}{N}$ . The power loading then yields  $p_{k,i,n} = c_{k,n}(A_i) \left[ \mu_n - \frac{1}{c_{k,n}(A_i)} \right]_+$ ,  $\forall k \in A_i$ .

By applying the conclusions above, the linear beamforming optimization problem, that performs user selection in each

subcarrier and resource allocation in the entire OFDM symbol, can be formulated as

$$\max_{\rho_{k,i,n}, p_{k,i,n}} \frac{B}{N} \sum_{k=1}^K \sum_{n=1}^N \sum_{i=1}^I \rho_{k,i,n} r_{k,i,n} \quad (5)$$

subject to

$$\begin{aligned} \rho_{k,i,n} &\in \{0, 1\}, \forall k, i, n, \\ p_{k,i,n} &\geq 0, \forall k, i, n, \\ \sum_{k=1}^K p_{k,i,n} &\leq \frac{P_{tot}}{N}, \forall n, i, \\ \sum_{k=1}^K \rho_{k,i,n} &\leq T, \forall n, i, \\ R_k &\geq mr_k, \forall k, \end{aligned}$$

where  $mr_k$  is the minimum data rate required by the  $k$ th user.  $\rho_{k,i,n}$  is the subcarrier allocation indicator such that  $\rho_{k,i,n} = 1$  if user  $k \in A_{i,n}$ , and  $A_{i,n}$  is selected in subcarrier  $n$ ; otherwise  $\rho_{k,i,n} = 0, \forall k, i, n$ . The total data rate for user  $k$ , denoted as  $R_k$ , is defined as

$$R_k = \frac{B}{N} \sum_{n=1}^N \sum_{i=1}^I \rho_{k,i,n} r_{k,i,n}. \quad (6)$$

The problem above (5) is an NP-hard combinatorial optimization problem with non-linear constraints. The optimal solution can be obtained by exhaustive search of all possible user assignment sets in all subcarriers but the complexity is given by  $I^N$ , which is extremely complicated even for moderate  $K, N$ .

#### IV. THE PROPOSED RESOURCE ALLOCATION ALGORITHM

In the following, a suboptimal, two-step, low-complexity user selection and resource allocation algorithm is proposed, that selects users independently in each subcarrier, it is based on ZF beamforming and guarantees individual minimum data rates.

1) *Step 1: Algorithm without minimum data rate constraints:* It is a modification of [17] implemented in each subcarrier.

- Set  $R_k = 0, \forall k, \rho_{k,i,n} = 0, \forall k, i$ , and  $\forall n \in \mathcal{S}$ .
- For  $n = 1, 2, \dots, N$ :
  - Set  $\mathcal{U} = \{1, 2, \dots, K\}, |A_{i,n}| = \emptyset$ .
  - Find user  $k = \operatorname{argmax}_{j \in \mathcal{U}} \|\mathbf{h}_{j,n}\|$ .
  - Set  $t = 1, \rho_{k,i,n} = 1, A_{i,n}(t) = \{k\}, \mathcal{U} = \mathcal{U} - \{k\}$ , and compute  $R_k$ , according to (4), (6).  $A_{i,n}(t)$  means the allocation result of the  $t$  step in subcarrier  $n$ .
  - For  $t = 2, 3, \dots, T$ :
    - \* Find a user,  $s_t \in \mathcal{U}$ , such that
 
$$\sum_{k \in A_i(t-1) \cup \{s_t\}} r_{k,i,n} > \sum_{k \in A_i(t-1)} r_{k,i,n}$$
    - \* If user  $s_t$  is found, set  $\rho_{s_t,i,n} = 1, A_{i,n}(t) = A_{i,n}(t-1) \cup s_t$ , and  $\mathcal{U} = \mathcal{U} - \{s_t\}$ .
    - \* Compute  $R_k, \forall k \in A_{i,n}(t)$ , according to (4), (6).

After the initialization, the algorithm finds user  $k$  with the best channel condition in subcarrier  $n$ . Subcarrier  $n$  is then assigned to additional users if the sum data rate in subcarrier  $n$  increases [17]. If there are more than one candidate users  $s_t$  in each step, pick the one with the maximum sum data rate. Procedure continues for all subcarriers.

2) *Step 2: Subcarrier reallocation:* In Step 1, a subcarrier allocation solution is obtained which does not guarantee the fulfillment of every user's data rate constraints. So, some subcarriers need to be allocated to the users whose minimum data rate constraints have not been satisfied yet, a procedure that causes an inevitable decrease in the overall data rate since these subcarriers were first allocated to users with the best channel condition on them. During the reallocation procedure each subcarrier reallocation should cause the least reduction in the sum of the users' data rates and the number of reallocation operations should be kept as low as possible. Thus,

$$e_{k,t,n} = \max \left( \frac{r_{t,i,n} - r_{k,i,n}}{r_{k,i,n}}, \frac{r_{k^*,i,n} - r_{k^*,i',n}}{r_{k^*,i',n}} \right), \quad (7)$$

$$\forall k^* \in A_{i,n} \cap A_{i',n}, \forall n \in \mathcal{S},$$

is defined as the cost function of reallocating subcarrier  $n$  to user  $k$ , forming  $A_{i',n}$ , instead of the originally assigned user  $t$ . If user  $k$  occupies subcarrier  $n$ , instead of user  $t$ , then the data rates of the other users (each  $k^* \in A_{i,n} \cap A_{i',n}$ ) occupying subcarrier  $n$  will be affected too. Hence, the cost function (7) takes into account the change of the data rate of all users in each subcarrier. In addition, max function, in (7), consists of at most  $T$  elements,  $r_{k^*,i,n}$  means the data rate of user  $k^* \in A_{i,n}$  and  $r_{k^*,i',n}$  means the data rate of user  $k^* \in A_{i',n}$ , both in subcarrier  $n$ . The subcarrier reallocation algorithm is as follows.

- For  $k = 1, 2, \dots, K$ :
  - Set  $\mathcal{S} = \{1, 2, \dots, N\}$ .
  - While  $R_k < mr_k$ :
    - \* Calculate the cost function according to (7).
    - \* Find  $[t^*, n^*] = \operatorname{argmin}_{t \in A_{i,n}, n \in \mathcal{S}} e_{k,t,n}$ .
    - \* If  $R_{t^*} - r_{t^*,i,n} \geq mr_{t^*}$  and  $R_m \geq mr_m, \forall m \in A_{i,n} \cap A_{i',n}$ : Set  $\rho_{t^*,i,n} = 0, \rho_{k,i,n} = 1, \mathcal{S} = \mathcal{S} - \{n\}, A_{i,n} = A_{i',n}$  and compute  $R_m, \forall m \in A_{i,n}$ , according to (4), (6).
    - \* Else:  $\mathcal{S} = \mathcal{S} - \{n\}$ .
    - \* If  $\mathcal{S} = \emptyset$ : break.

In Step 2,  $R_k, A_{i,n}$  and  $\rho_{k,i,n}, \forall k, i, n$  are known from Step 1. Subcarrier reallocation is carried out on a user-by-user basis for all users whose minimum data rate requirements have not been satisfied in Step 1. Consider user  $k$  for example. In each stage, user  $t^*$  and subcarrier  $n^*$  with the lowest cost function are selected which cause the least reduction in the sum of the users' data rates. Subcarrier  $n^*$  will be allocated to user  $k$  instead of the originally assigned user  $t^*$ , if this reallocation does not cause  $R_{t^*} - r_{t^*,i,n} < mr_{t^*}$  and  $R_m < mr_m, \forall m \in A_{i,n} \cap A_{i',n}$ . Otherwise, the reallocation will not be done and

new  $t^*$ ,  $n^*$  will be identified from the rest of the subcarriers. This subcarrier reallocation process repeats for user  $k$  until its data rate requirement is satisfied, otherwise an outage is occurred.

## V. COMPLEXITY ANALYSIS

In order to analyze the computational complexity of the proposed algorithm, recall that  $K$  refers to the total number of users in the system and  $T$  refers to the number of transmit antennas at the BS.  $N$  on the other hand refers to the number of subcarriers, which is much larger than both  $K$  and  $T$ .

In Step 1 of the proposed algorithm, the best user  $k$  among  $K$  users is found for subcarrier  $n = 1, 2, \dots, N$ , which requires  $O(KN)$  operations. Then, at most  $T - 1$  other users are found for subcarrier  $n$  which requires the evaluation of at most  $T$  data rates  $r_{k,i,n}$ . In order to evaluate  $r_{k,i,n}$ , inversion of  $\mathbf{H}_n(A_i(t-1) \cup s_t) \mathbf{H}_n(A_i(t-1) \cup s_t)^*$  is required which can be done in time  $O(T^2)$ , for the worst case, when  $T$  users occupy each subcarrier, using the matrix inversion lemma as described in [17]. Repeating this over at most  $K - 1$  users ( $s_t \notin A_i(t-1)$ ) in each one of the  $t = 2$  to  $T$  steps, and over all subcarriers of set  $\mathcal{S}$ , the overall complexity of Step 1 is obtained to be  $O(KNT^3)$ .

In Step 2, while loop runs for at most  $N$  times for each user. In the while loop, cost function is calculated which requires  $T$  comparisons for the max function,  $O(NT^2)$  time for the data rates ( $r_{k,i,n}$ ,  $r_{k^*,i,n}$ ,  $r_{k^*,i',n}$ ) for all subcarriers and  $TN$  multiplications. Therefore, cost function requires  $O(T^4N^2)$ . Finding  $t^*$ ,  $n^*$ , requires  $O(TN)$ . Thus, complexity of Step 2 is  $O(KN^2T^4)$  which is also the complexity of the whole proposed algorithm.

As was mentioned the complexity of exhaustive search for the optimal solution of the original problem is given by  $I^N$ , where  $I = \sum_{l=1}^T \binom{K}{l}$ . Alternatively, the complexity is  $O(K^{NT})$  and is prohibitive even for moderate values of  $K$ ,  $N$ , and  $T$ . Thus, it is easily observed that the proposed algorithm has a very dramatic reduction in complexity compared to  $O(K^{NT})$  required by the exhaustive search.

## VI. SIMULATION RESULTS

The proposed algorithm is compared with the algorithms proposed in [17]-[21], Round Robin (RR) algorithm, and Maximal Ratio Combining (MRC) transmission, only to the user with the strongest channel.  $mr_k = 1.5$  bits/s/Hz,  $\forall k$ , and in [20], proportional data rate constraints are  $\gamma_k = \frac{mr_k}{\sum_{k=1}^K mr_k}$ ,  $\forall k$  and system parameters are  $D = 0.1$ ,  $L = T$ . In RR algorithm, each user is given a fair share of the channel resource regardless of the channel state and  $T$  users are selected in each subcarrier. Both equal power (EQ) allocation and water-filling (WF) power allocation over the parallel subchannels are considered.

In all simulations presented in this section, the frequency-selective channel consists of six independent Rayleigh multipath components. As in [20], an exponentially decaying power delay profile is assumed, the ratio of the energy of the  $l$ th

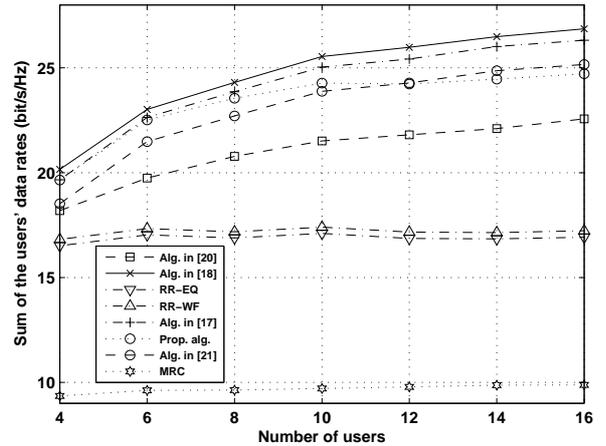


Fig. 2. Sum of the users' data rates vs K.

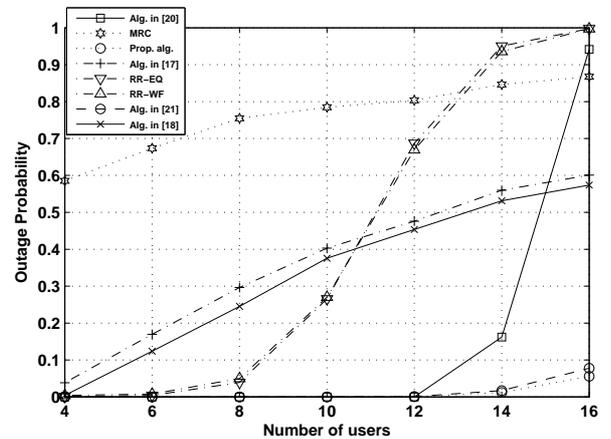


Fig. 3. Outage probability vs K.

tap to the first tap being equal to  $e^{-2l}$ . A maximum delay spread of  $5\mu\text{s}$  and maximum doppler of 30Hz is assumed. The channel information is sampled every 0.5ms to update the proposed algorithm,  $T = 4$ ,  $N = 128$  and the number of channel realizations is equal to  $10^5$ .

In Figs. 2, 3,  $K$  varies from 4 – 16 in increment of 2 and  $SNR = 20$ , while in Figs. 4, 5,  $K = 16$ . Also, Figs. 2, 4 depict the sum of the users' data rates vs number of users and average channel SNR, respectively, and Figs. 3, 5 depict the outage probability for different values of number of users and average channel SNR, respectively. Outage probability is defined as the ratio of the users that have not reached their target data rate over  $K$ .

In Figs. 2, 4, it can be seen the reasonable price being paid in order to guarantee minimum data rate requirements by using the proposed algorithm. In Fig. 2, as the number of users increases, the difference in sum data rates increases because additional multiuser diversity is available to [17] [18] that only target sum data rate maximization. On the other hand, more users put more constraints to the proposed algorithm, because new users need to share the same resources. Algorithm

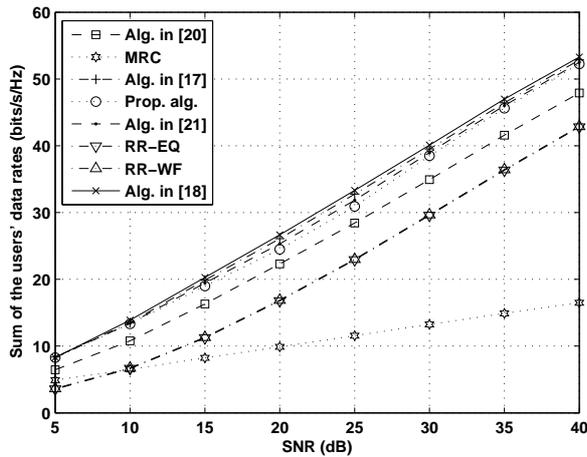


Fig. 4. Sum of the users' data rates vs SNR.

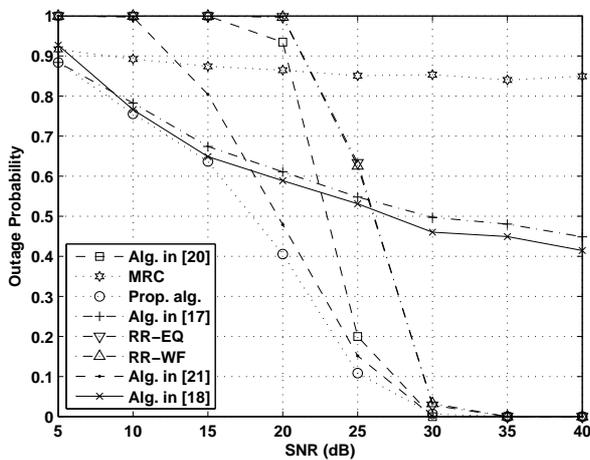


Fig. 5. Outage probability vs SNR.

in [17] is only the first Step of the proposed algorithm where minimum data rate requirements are not yet considered while MRC algorithm is the lower bound of the proposed algorithm as in MRC each subcarrier is allocated to only one user. In [20], although  $\gamma_k = \frac{mr_k}{\sum_{k=1}^K mr_k}$ ,  $\forall k$ , it does not mean that minimum data rate requirements are satisfied, as seen in Figs. 3, 5. Moreover, sum data rate of the proposed algorithm is significantly enhanced over both RR-WF and RR-EQ algorithm, wherein the channel gain information is not exploited. Furthermore, algorithm in [21] imposes a kind of fairness between users' data rates. In addition, in Fig. 3, the outage probability achieved by the proposed algorithm is smaller than that achieved from the reference algorithms. Finally, in Fig. 5, it is also shown that the outage probability achieved by the proposed algorithm decreases quickly over the average channel SNR and is lower than that of the reference algorithms.

## VII. CONCLUSION

A fairness-aware user selection and resource allocation algorithm, which is based on ZF beamforming, for the MISO

downlink over frequency-selective channels was introduced. The main goal was to satisfy the minimum data rate requirements of users despite the loss with respect to the unconstrained case where the only target is the maximization of the sum data rate. Simulation results provide proofs about these statements and complexity analysis shows the dramatic reduction in complexity compared with exhaustive search.

## REFERENCES

- [1] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*, Prentice Hall, 2007.
- [2] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas in Commun.*, vol. 21, no. 2, pp. 171-178, Feb. 2003.
- [3] G. Li and H. Liu, "On the Optimality of the OFDMA Network," *IEEE Commun. Lett.*, vol. 9, no. 5, pp. 438-440, May 2005.
- [4] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with Adaptive Subcarrier, Bit, and Power Allocation," *IEEE J. Sel. Areas in Commun.*, vol. 17, no. 10, pp. 1747-1758, Oct. 1999.
- [5] D. Kivanc, G. Li, and H. Liu, "Computationally Efficient Bandwidth Allocation and Power Control for OFDMA," *IEEE Trans. Wireless Commun.*, vol. 2, no. 6, pp. 1150-1158, Nov. 2003.
- [6] Y. Chen and J. Chen, "A Fast Subcarrier, Bit, and Power Allocation Algorithm for Multiuser OFDM-Based Systems," *IEEE Trans. Veh. Technol.*, vol. 57, no. 2, pp. 873-881, Mar. 2008.
- [7] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," *IEEE Vehicular Technol. Conf.*, vol. 2, pp. 1085-1089, Tokyo, Japan, May 2000.
- [8] C. Mohanram and S. Bhashyam, "A Sub-optimal Joint Subcarrier and Power Allocation Algorithm for Multiuser OFDM," *IEEE Commun. Lett.*, vol. 9, no. 8, pp. 685-687, Aug. 2005.
- [9] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive Resource Allocation in Multiuser OFDM Systems With Proportional Rate Constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726-2737, Nov. 2005.
- [10] V. D. Papoutsis, I. G. Fraimis, and S. A. Kotsopoulos, "Fairness-aware resource allocation for the SISO downlink over frequency-selective channels," *IEEE Wireless Commun. and Networking Conf.*, pp. 1-6, Sydney, NSW, Australia, Apr. 2010.
- [11] Y. J. Zhang and K. B. Letaief, "Multiuser Adaptive Subcarrier-and-Bit Allocation With Adaptive Cell Selection for OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1566-1575, Sep. 2004.
- [12] Z. Mao and X. Wang, "Efficient Optimal and Suboptimal Radio Resource Allocation in OFDMA System," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 440-445, Feb. 2008.
- [13] Y. Ma, "Rate-Maximization for Downlink OFDMA with Proportional Fairness," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3267-3274, Sep. 2008.
- [14] J. Huang, V. G. Subramanian, R. Agrawal, and R. A. Berry, "Downlink Scheduling and Resource Allocation for OFDM Systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 288-296, Jan. 2009.
- [15] A. Biagioni, R. Fantacci, D. Marabissi, and D. Tarchi, "Adaptive Subcarrier Allocation Schemes for Wireless OFDMA Systems in WiMAX Networks," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 217-225, Feb. 2009.
- [16] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian MIMO broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936-3964, Sep. 2006.
- [17] G. Dimic and N. D. Sidiropoulos, "On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3857-3868, Oct. 2005.
- [18] P. W. C. Chan and R. S. Cheng, "Capacity Maximization for Zero-Forcing MIMO-OFDMA Downlink Systems with Multiuser Diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1880-1889, May 2007.
- [19] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-Forcing Methods for Downlink Spatial Multiplexing in Multiuser MIMO Channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.
- [20] V. D. Papoutsis, I. G. Fraimis, and S. A. Kotsopoulos, "User Selection and Resource Allocation Algorithm with Fairness in MISO-OFDMA," *IEEE Commun. Lett.*, vol. 14, no. 5, pp. 411-413, May 2010.
- [21] S. Kai, W. Ying, C. Zi-xiong, and Z. Ping, "Fairness based resource allocation for multiuser MISO-OFDMA systems with beamforming," *J. China Univ. of Posts and Telec.*, vol. 16, no. 1, pp. 38-43, Feb. 2009.

# Scrambling–Exact Channel Estimation in W–CDMA Systems

## An Algorithm Beyond Correlation

Wolfgang U. Aichmann

Nokia Siemens Networks GmbH & Co. KG  
Lise–Meitner–Straße 7, D–89081 Ulm, Germany  
Email: wolfgang.aichmann@nsn.com

**Abstract—** Multiple–input Multiple–output (MIMO) antenna systems require one pilot signal for each transmitting antenna. As scrambling is a quasi–orthogonal operation a small portion of the pilot energy is scattered into other pilot signals even with ideal conditions apart from that small non–orthogonality. In this paper the correlation approach of estimating the channel impulse response is extended analytically such as to take into account the full spreading sequence, improving thereby significantly the performance of high–speed downlink packet access (HSDPA) MIMO systems especially for very high data rates when dual stream transmission is used in combination with cancellation of the interference between the data streams.

**Keywords–** W–CDMA; HSDPA; MIMO; channel estimation; scrambling

### I. INTRODUCTION

Channel estimation plays an important role in wideband code division multiple access (W–CDMA) networks, especially for very high data rates as they are achieved with MIMO [1]. It has therefore been topic of different investigations which go beyond the conventional correlation approach. In [2], a novel type of linear minimum mean square equalizer (LMMSE) was presented which is able to properly take into account all types of interference without being too complex for implementation in real systems, whereas [6,7] investigate smoothening the primary channel estimates by appropriate filtering and cancelling the inter–antenna interference, respectively.

In MIMO systems it is necessary to measure all physical channels from each transmitting to each receiving antenna. Therefore, each transmitting antenna has to be fed with a separate pilot sequence. For HSDPA systems [5], there are defined primary and secondary common pilot channels which are orthogonal either by using orthogonal patterns or by applying orthogonal spreading codes. In any case, the coded pilot sequences are scrambled.

As scrambling is a quasi–orthogonal but not an exact orthogonal operation the channel information derived by the receiver remains distorted even under conditions being ideal beyond these distortions, i.e., absence of receiver noise and additive white Gaussian noise (AWGN) channels. If only

one pilot channel is present, the effect is usually negligible: Although a certain amount of energy is scattered by the spreading operation into other codes the general shape persists as the energy reduction is more or less proportional to the available energy. If there are however two or more pilot sequences the relative error strongly increases as the scattering of energy from other codes also happens at time (slot) positions where the respective pilot signal is weak or even vanishes.

If MIMO is configured for more than one data stream, these data streams will interfere with each other with strength depending on the correlation of the different physical channels. But this interference can effectively be reduced by interference cancellation as all data streams have the same origin and therefore are subject of the same channel conditions. Furthermore, the receiver has to decode all data streams anyway so that the effort of cancellation is limited. As interference cancellations is re–applying the channel transformation to the already detected data, any error in the estimated channel information takes effect three times, namely at first equalization, at re–application of the channel operation in the course of the interference cancellation and at re–equalization of the equivalent single stream data. Any error in channel estimation therefore acts non–linearly on the overall performance of the system.

The scrambling distortion is mainly significant for very high data rates where the respective UE is served with all available resources. This means that intra–cell interference (originating from other users) is small. Furthermore, these mentioned high data rates are achievable under quite good radio conditions only, i.e., also the inter–cell interference must be small. Therefore, a correlation–based estimator might be sufficient in that range of interest but the impact of scrambling induced distortion should be removed.

As the receiver knows the scrambled pilot symbol patterns as they are fed to the transmitting antennas it is possible to calculate the distortion induced by scrambling analytically as a function of the channel impulse response. Inverting this function allows to extend the correlation algorithm such as to take into account the distortion, leading to an estimated channel impulse response being exact with respect to scrambling effects. Although a matrix operation is

required to solve the resulting equation system the corresponding matrix elements depend only on the scrambling vector and can hence be calculated once at the initial phase. As long as the scrambling code is not changed these elements remain stable.

The remainder of this paper is organized as follows: In Section II we introduce the signal model with the basic equations. They are required to describe the data transfer through the channel and will be used in Section III for formulating the analytical solution for scrambling–exact channel estimation. Section IV then demonstrates the gain of the proposed algorithm with some results from link level simulation. Finally, we draw our conclusion in Section V.

## II. BASIC CHANNEL PILOT EQUATIONS

In the following, we assume that two transmitting antennas are present, each fed with a separate pilot sequence, namely  $s_1(n)$  and  $s_2(n)$  with  $n$  counting the oversampled time steps. The signal  $r(l)$  at oversampled chip position  $l$  is for a single antenna then given by (see e.g., [4], Section 1.3):

$$r(l) = \sum_{k=0}^{N_{\text{OSF}}-M} c \left( \frac{l-k}{N_{\text{OSF}}} \right) \left[ h_1(k) s_1 \left( \frac{l-k}{N_{\text{OSF}}} \right) + h_2(k) s_2 \left( \frac{l-k}{N_{\text{OSF}}} \right) \right]. \quad (1)$$

With the conventional correlation approach of channel estimation, the received signal is de–spread ignoring the error made by this quasi–orthogonal operation (see, e.g., [1]):

$$\begin{aligned} \sum_{l=l_0+1}^{N_{\text{SF}}} c * \left( \frac{l}{N_{\text{OSF}}} \right) r(l+d) &= \\ &= \sum_{l=l_0+1}^{N_{\text{SF}}} \sum_{k=0}^{N_{\text{OSF}}-M} c * \left( \frac{l}{N_{\text{OSF}}} \right) c \left( \frac{l-k+d}{N_{\text{OSF}}} \right) \cdot \\ &\quad \cdot \left[ h_1(k) s_1 \left( \frac{l-k+d}{N_{\text{OSF}}} \right) + h_2(k) s_2 \left( \frac{l-k+d}{N_{\text{OSF}}} \right) \right] \approx \\ &\approx h_1(d) s_1(l_0) + h_2(d) s_2(l_0) \end{aligned} \quad (2)$$

In (2), we introduced a start index  $l_0$  for summation over chips in order to cope with handling of more than one symbol. For the first symbol,  $l_0 = 0$ , for the second one  $l_0 = N_{\text{SF}}$ , and so on. If we now define

$$\begin{aligned} \kappa_{l_0,j}(d,k) &\equiv \kappa_{l_0,j}(d-k) = \\ &= \sum_{l=l_0+1}^{N_{\text{OSF}}-N_{\text{SF}}} c * \left( \frac{l}{N_{\text{OSF}}} \right) c \left( \frac{l-k+d}{N_{\text{OSF}}} \right) s_j \left( \frac{l-k+d}{N_{\text{OSF}}} \right) \\ \kappa_{l_0,j}(d-k) &= 0 \quad \forall d-k \notin Z \\ \rho_{l_0}(d) &= \sum_{l=l_0+1}^{N_{\text{SF}}} c * \left( \frac{l}{N_{\text{OSF}}} \right) r(l+d) \end{aligned} \quad (3)$$

eq. (2) can be written as

TABLE 1: NAMING CONVENTIONS FOR IMPORTANT VARIABLES

Received signal (single antenna)	$r(l)$
Sent signal (index $i$ for Tx antenna)	$s_i(l)$
Channel impulse response (index $i$ for Tx antenna)	$h_i(l)$
Scrambling code element at integer chip index $n$	$c(n)$
Channel length	$M + 1$
Spreading sequence length	$N_{\text{SF}}$
Oversampling Factor	$N_{\text{OSF}}$
Number of symbols required for CPICH orthogonality	$N_{\text{ortho}}$
Number of symbols required for CPICH orthogonality, measured in chips	$N = \frac{\langle \text{chips per slot} \rangle}{N_{\text{ortho}}}$

$$\rho_{l_0}(d) = \sum_{k=0}^{N_{\text{OSF}}-M} [h_1(k) \kappa_{l_0,1}(d-k) + h_2(k) \kappa_{l_0,2}(d-k)]. \quad (4)$$

We then are able to split the summation into two parts, namely one counting the chips and the other stepping through the oversampling within each chip. With the definitions

$$\begin{aligned} d &= N_{\text{OSF}} \left\lfloor \frac{d}{N_{\text{OSF}}} \right\rfloor + d \bmod N_{\text{OSF}} = \hat{d} + \check{d} \\ \hat{d} &= N_{\text{OSF}} \left\lfloor \frac{d}{N_{\text{OSF}}} \right\rfloor \\ \check{d} &= d \bmod N_{\text{OSF}} \geq 0 \end{aligned} \quad (5)$$

and assuming that  $l$  is an integer multiple of  $N_{\text{OSF}}$  (generally spoken, oversampling can be accounted for in  $d$ ) we can write finally

$$\begin{aligned} \rho_{l_0}(\hat{d} + \check{d}) &= \\ &= \sum_{k=0}^M \sum_{j=0}^{N_{\text{OSF}}-1} [h_1(N_{\text{OSF}}k + j) \kappa_{l_0,1}(\hat{d} + \check{d} - N_{\text{OSF}}k - j) + \\ &\quad h_2(N_{\text{OSF}}k + j) \kappa_{l_0,2}(\hat{d} + \check{d} - N_{\text{OSF}}k - j)] = \\ &= \sum_{k=0}^M [h_1(N_{\text{OSF}}k + \check{d}) \kappa_{l_0,1}(\hat{d} - N_{\text{OSF}}k) + \\ &\quad h_2(N_{\text{OSF}}k + \check{d}) \kappa_{l_0,2}(\hat{d} - N_{\text{OSF}}k)] \end{aligned} \quad (6)$$

In (6), we have separated integer and oversampled parts on the right hand side of (4) by splitting

$$k = N_{\text{OSF}} \left\lfloor \frac{k}{N_{\text{OSF}}} \right\rfloor + j \rightarrow N_{\text{OSF}}k + j. \quad (7)$$

### III. SCRAMBLING–EXACT CHANNEL ESTIMATION

In [5], there are defined different configurations for the common pilot channel (CPICH) in W-CDMA systems. For the usage in combination with MIMO, two of them are relevant, namely either using only the primary CPICH spread with code  $c_{SF1}$  of length 256 and orthogonal bit sequences on the two transmitting antennas, or using the primary CPICH on the first and a secondary CPICH with code  $c_{SF2}$  on the second transmitting antenna. The secondary CPICH than must use the same bit sequence as the primary CPICH on antenna 1. The same scrambling code for both cases is used in any case.

If two adjacent symbols are merged both options can formally be described by a virtual spreading with length of 512 where the effective spreading codes can be defined in the first case as  $[c_{SF1} \ c_{SF1}]$  and  $[c_{SF1} \ \text{inv}(c_{SF1})]$ , and in the second case as  $[c_{SF1} \ c_{SF1}]$  and  $[c_{SF2} \ c_{SF2}]$ . As long as the channel estimation algorithm finally uses any filtering procedure over an even number of pilot symbols, both options are equivalent with respect to the influence of scrambling. We will restrict therefore the following investigations to the configuration with primary CPICH only and orthogonal bit sequences on the two transmitting antennas.

#### A. Single Symbol Channel Estimation

If we want to calculate  $M + 1$  coefficients for both channels, we need  $2 \cdot (M + 1)$  equations, i.e.,

$$\begin{pmatrix} \rho_{l_0}(N_{OSF} \cdot 0 + \vec{d}) \\ \vdots \\ \rho_{l_0}(N_{OSF} \cdot M + \vec{d}) \\ \vdots \\ \rho_{l_0}(N_{OSF} \cdot (2M + 1) + \vec{d}) \end{pmatrix} = K_{l_0} \begin{pmatrix} h_1(N_{OSF} \cdot 0 + \vec{d}) \\ h_2(N_{OSF} \cdot 0 + \vec{d}) \\ \vdots \\ h_1(N_{OSF} \cdot M + \vec{d}) \\ h_2(N_{OSF} \cdot M + \vec{d}) \end{pmatrix}, \quad (8)$$

where the pilot scrambling matrix  $K_{l_0}$  is defined as

$$K_{l_0} = \begin{pmatrix} \kappa_{l_0,1}(0) & \kappa_{l_0,2}(0) & \cdots & \kappa_{l_0,1}(-M) & \kappa_{l_0,2}(-M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{l_0,1}(M) & \kappa_{l_0,2}(M) & \cdots & \kappa_{l_0,1}(0) & \kappa_{l_0,2}(0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{l_0,1}(2M+1) & \kappa_{l_0,2}(2M+1) & \cdots & \kappa_{l_0,1}(M+1) & \kappa_{l_0,2}(M+1) \end{pmatrix}. \quad (9)$$

If the scrambling sequences would be completely orthogonal, the pilot scrambling matrix (9) would contain non-zero elements ( $\kappa_j(0)$ ) in the first  $M$  rows only, i.e., rows  $M + 1 \dots 2M + 1$ , wouldn't contribute at all to the solution and  $K$  wouldn't be invertible. The accuracy of these equations is therefore small even if the non-orthogonality of spreading is taken into account. To overcome this limitation one has therefore either to average over at least as many symbols as there are pilot sequences (see Section B), or the approach is extended such as to obtain equations of each pilot symbol (see Section C). This becomes obvious when

looking to the pilot scrambling matrix for a single-tap channel including root raised cosine filtering as displayed in Fig. 1.

The channel coefficients (at sampling positions) are finally given by

$$\begin{pmatrix} h_1(0) \\ h_2(0) \\ \vdots \\ h_1(M) \\ h_2(M) \end{pmatrix} = K_{\text{single}, l_0}^{-1} \begin{pmatrix} \rho_{l_0}(0) \\ \vdots \\ \rho_{l_0}(M) \\ \vdots \\ \rho_{l_0}(2M+1) \end{pmatrix}. \quad (10)$$

The upper half of  $K$  shows large real entries on the main diagonal, whereas the other elements are dominated by the scrambling correlation coefficients. As expected, the calculation of the channel impulse response fails at least partly even if afterwards an averaging over all symbols of a frame is performed. This is shown in Fig. 2 for an AWGN channel with raised cosine filter.

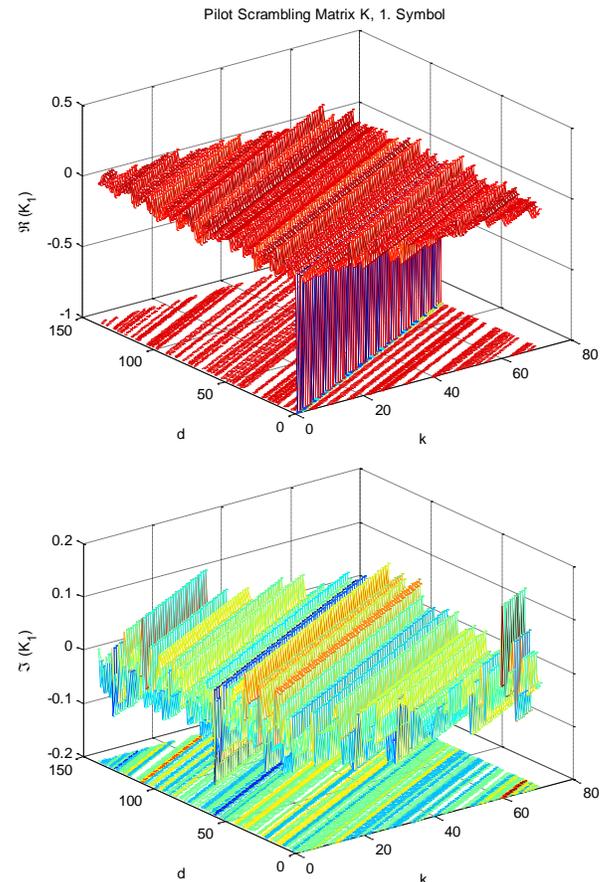


Figure 1. Pilot scrambling matrix for a single tap channel including raised cosine filtering.

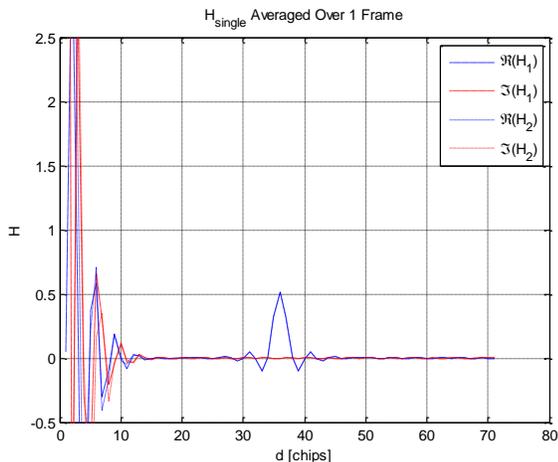


Figure 2. Channel Impulse Response averaged over all symbols of a frame.

### B. Averaging Pilot Scrambling Matrix

The inaccuracy of the channel impulse response calculated for a single symbol is caused by  $\mathbf{K}$  being badly conditioned. This can be improved by averaging both  $\mathbf{K} \rightarrow \langle \mathbf{K} \rangle_{\text{frame}}$  and  $\rho \rightarrow \langle \rho \rangle_{\text{frame}}$ . Although the general shape of the pilot scrambling matrix is preserved by this operation the conditioning is now about 10 orders of magnitude better and  $\langle \mathbf{K} \rangle_{\text{frame}}$  becomes invertible. For an AWGN channel with raised cosine filter it provides a perfectly estimated channel impulse response, see Fig. 3.

### C. Twin Symbol Channel Estimation

The averaging procedure discussed in Section B assumes that  $K$  and  $\rho$  are statistically independent and the averaging therefore factorizes and can be executed independently. Inversion is then executed with the averaged pilot scrambling matrix. The shown result implies that this assumption is correct at least for a single tap channel including root raised cosine filter.

There is, however, an alternative approach which doesn't need the above restriction: Both pilot scrambling matrix and de-spread received signal vector are constructed using two symbols with different pilot sequences,

$$\begin{pmatrix} \rho_{i_0}(0) \\ \vdots \\ \rho_{i_0}(M) \\ \rho_{i_1}(0) \\ \vdots \\ \rho_{i_1}(M) \end{pmatrix} = \begin{pmatrix} \kappa_{i_0,1}(0) & \kappa_{i_0,2}(0) & \cdots & \kappa_{i_0,1}(-M) & \kappa_{i_0,2}(-M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{i_0,1}(M) & \kappa_{i_0,2}(M) & \cdots & \kappa_{i_0,1}(0) & \kappa_{i_0,2}(0) \\ \kappa_{i_1,1}(0) & \kappa_{i_1,2}(0) & \cdots & \kappa_{i_1,1}(-M) & \kappa_{i_1,2}(-M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{i_1,1}(M) & \kappa_{i_1,2}(M) & \cdots & \kappa_{i_1,1}(0) & \kappa_{i_1,2}(0) \end{pmatrix} \begin{pmatrix} h_1(0) \\ h_2(0) \\ \vdots \\ h_1(M) \\ h_2(M) \end{pmatrix}, \quad (11)$$

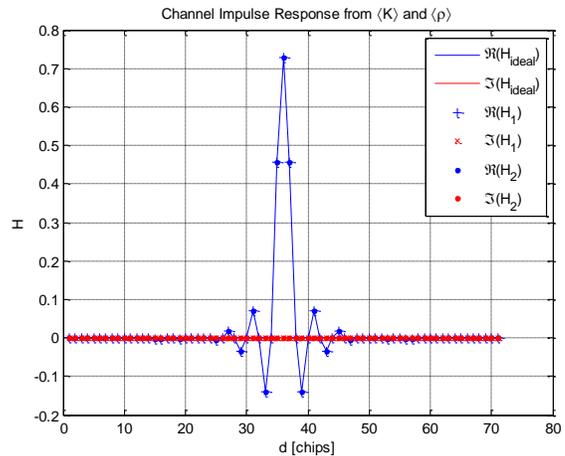


Figure 3. Channel Impulse Response calculated with averaged  $\mathbf{K}$ .

where the pilot scrambling matrix  $\mathbf{K}_{\text{twin}}$  now is defined as

$$\mathbf{K}_{\text{twin},i_0,i_1} = \begin{pmatrix} \kappa_{i_0,1}(0) & \kappa_{i_0,2}(0) & \cdots & \kappa_{i_0,1}(-M) & \kappa_{i_0,2}(-M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{i_0,1}(M) & \kappa_{i_0,2}(M) & \cdots & \kappa_{i_0,1}(0) & \kappa_{i_0,2}(0) \\ \kappa_{i_1,1}(0) & \kappa_{i_1,2}(0) & \cdots & \kappa_{i_1,1}(-M) & \kappa_{i_1,2}(-M) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \kappa_{i_1,1}(M) & \kappa_{i_1,2}(M) & \cdots & \kappa_{i_1,1}(0) & \kappa_{i_1,2}(0) \end{pmatrix}. \quad (12)$$

In the following, these symbols will be called *twin symbols* if the two symbols are adjacent. The channel coefficients are then given by

$$\begin{pmatrix} h_1(0) \\ h_2(0) \\ \vdots \\ h_1(M) \\ h_2(M) \end{pmatrix} = \mathbf{K}_{\text{twin},i_0,i_1}^{-1} \begin{pmatrix} \rho_{i_0}(0) \\ \vdots \\ \rho_{i_0}(M) \\ \rho_{i_1}(0) \\ \vdots \\ \rho_{i_1}(M) \end{pmatrix}. \quad (13)$$

Now, all rows and columns of the pilot scrambling matrix contain (at least) one main element as displayed in Fig. 4.

The real part shows two bars, one completely in the lower half volume, the other in both half volumes. Whereas the former belongs to equal pilot symbols for both channels, the latter results from inverse symbols.

The AWGN channel impulse response including raised cosine filter calculated from the twin pilot correlation matrix again matches perfectly with the ideal channel impulse response, not only when averaged over all symbols of a frame (see Fig. 5), but also for each twin symbol. This means that no averaging is required in the absence of noise.

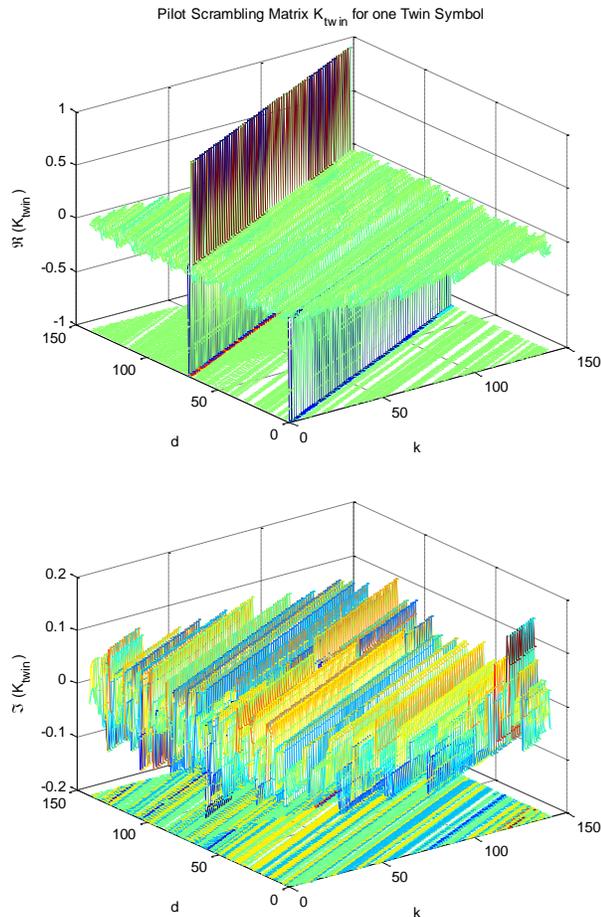


Figure 4. Pilot scrambling matrix for one twin symbol.

#### IV. LINK LEVEL SIMULATIONS

In order to demonstrate the benefit of the scrambling–exact channel estimation we executed link level simulations for a single UE moving with 3km/h and assuming a channel of type Pedestrian A. Two data streams were transmitted in MIMO mode and for both the modulation and coding scheme corresponding to entry 10 in the CQI mapping table K [6] was used. The transmitting antennas were assumed to be mounted crosswise diagonal (X), the receiving antennas crosswise perpendicular and horizontal (+) and crosstalk between the polarization directions was allowed. In Fig. 6, the block error rate (BLER) is displayed for each data stream as a function of SNR. The SNRs for a working point at 10% BLER are given in the legend for the first (solid line) and second (dashed line) data stream.

As expected, there is an appreciable gain of performance notable by a shift of the working point of up to 2dB for the second data stream which mainly benefits from interference cancellation, but even the working point of the first data stream is improved by 0.5dB.

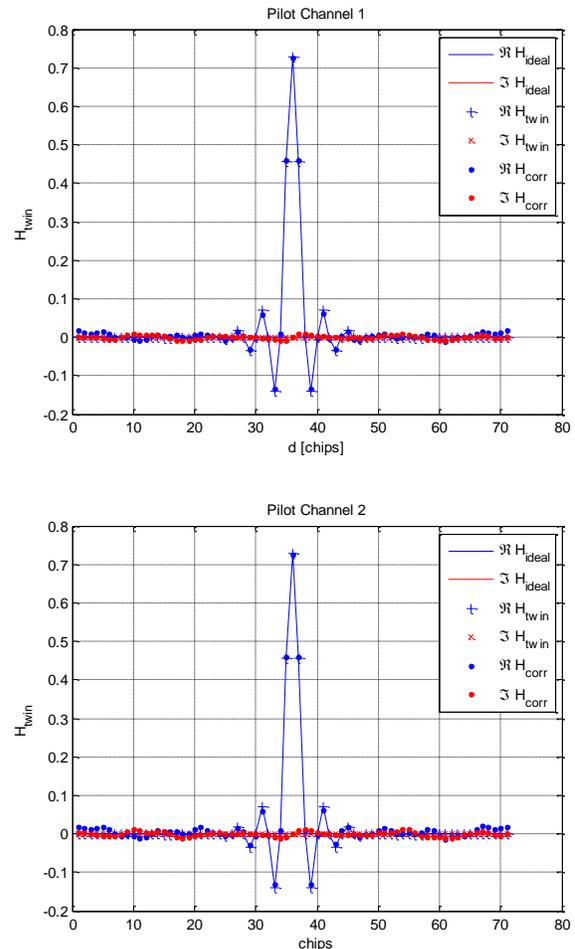


Figure 5. Twin Channel Impulse Response averaged over all symbols, compared with conventional correlation result and ideal case.

#### V. CONCLUSION

The usual procedure for estimating the radio channels discards the non–orthogonality of the spreading sequence which causes an error of several percents in case of more than one pilot sequence is required as is valid for using spatial diversity and MIMO systems. It has to be emphasized that this error is present even in case of pilot channels being fully separated, e.g., by cross polarized wave fronts with each front bearing exactly one pilot and receiving antennas oriented along the polarization directions, as long as this fact is not known a priori on receiving side: The estimator has to assume that signals of both pilots are present at each receiving antenna and hence misinterprets fractions scattered from the ‘real’ pilot by scrambling as contributions of the other pilot. Merely a single pilot being present combined with the knowledge about this fact on receiving side avoids this complication.

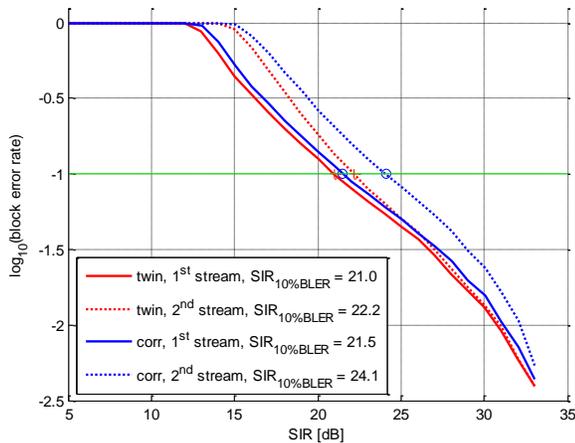


Figure 6. Link level simulation results with simple correlation (upper) and scrambling-exact (lower) channel estimation.

Due to the importance of channel estimation for reaching high data rates, there are published many proposals for improving the basic correlation approach, e.g., by filtering the primary channel estimates [3], cancelling inter-antenna interference [4] or extending the RAKE correlation approach to a LMMSE one [2] and thus taking into account all kinds of interference. All of these suggestions can basically be extended by scrambling-exact dealing of pilot patterns as proposed in this paper even if we have investigated here only the simple RAKE correlation approach, including appropriate filtering as described in [3], in the simulator.

Although the spreading induced error is negligible in many cases, high data rates in combination with MIMO and transmission of two data streams require a higher accuracy. This can be reached by taking into account the spreading sequence exactly in the channel estimation algorithm. As in the correlation case, two symbols (in case of two pilot signals) with different pilot code elements are required to resolve the channels from both transmitting antennas. Therefore, a twin symbol pilot scrambling matrix is introduced with its inverse providing the channel impulse response from the received data.

The numerical effort of this approach is for sure higher than the conventional correlation method. This effort however can at least partly be shifted to an initialization phase because the pilot scrambling matrix only depends on the scrambling sequence, the pilot patterns and the maximum

length of the channel but not on any quantities varying with time. Mainly the second (weaker) data stream benefits from this improved channel estimation if cancellation of inter-stream interference is applied. In link level simulation it was shown that its working point as function of SNR can be improved by 2dB for 10% block error rate.

## VI. ACKNOWLEDGMENT

The Author would like to thank Elke Hallmann and Klaus Knoche for many inspiring discussions during the course of this study.

## REFERENCES

- [1] S. McBeath, M. Ahmed, and K. Rohani, "Impact of imperfect estimators on W-CDMA receiver performance with MIMO antenna systems", in Proc. IEEE 58th Vehicular Technology Conference (VTC 2003-Fall), vol. 2, Oct.2003, pp. 1152-1156. [Online]. Available: <http://ieeexplore.ieee.org/iel5/9004/28569/01285202.pdf?tp=&arnumber=1285202>.
- [2] Ch. Mehlführer, and M. Rupp, "Novel Tap-wise LMMSE Channel Estimation for MIMO W-CDMA" in Proc. 51st Annual IEEE Globecom Conference, 2008, New Orleans, LA, USA, Nov. 2008. [Online]. Available: [http://publik.tuwien.ac.at/files/PubDat\\_169129.pdf](http://publik.tuwien.ac.at/files/PubDat_169129.pdf)
- [3] Bengt Lindoff, Christer Östberg, and Håkan Eriksson, "Channel estimation for the W-CDMA system, performance and robustness analysis from a terminal perspective" in Proc. IEEE 49<sup>th</sup> Vehicular Technology Conference (VTC 1999), vol. 2, July 1999, pp. 1565-1569. [Online]. Available: <http://ieeexplore.ieee.org/iel5/6314/16887/00780649.pdf?tp=&arnumber=780649>
- [4] I.R.S. Casella, E.S. Sousa, and P.J.E. Jeszensky, "Chip space-time block coding using iterative channel estimation with inter-antenna interference cancellation for W-CDMA systems using long scrambling codes", in Proc. IEEE 15<sup>th</sup> International Symposium on Personal, Indoor and Mobile Radio Communications, (PIMRC 2004), vol. 4, pp. 2313-2317. [Online]. Available: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1368732&isnumber=29952>
- [5] 3GPP, "Technical specification group radio access network; physical channels and mapping of transport channels (FDD) (Tech. Spec. 25.211 V8.4.0)", 2009. [Online]. Available: [http://www.3gpp.org/ftp/Specs/archive/25\\_series/25.211/](http://www.3gpp.org/ftp/Specs/archive/25_series/25.211/)
- [6] John G. Proakis and Masoud Saleh, "Digital Communications", 5<sup>th</sup> ed., McGraw-Hill, 2009.
- [7] 3GPP, "Technical specification group radio access network; physical layer procedures (FDD) (Tech. Spec. 25.214 V8.5.0)", 2009. [Online]. Available: [http://www.3gpp.org/ftp/Specs/archive/25\\_series/25.214/](http://www.3gpp.org/ftp/Specs/archive/25_series/25.214/)

# The Mathematical Relationship Between Maximum Access Delay and the R.M.S Delay Spread

Khalid Hassan, Tharek Abd Rahman  
Wireless Communication Center  
Universiti Teknologi Malaysia (UTM)  
81310, Skudai, Malaysia  
Emails: hmmkhalid2@live.utm.my  
tharek@fke.utm.my

M. Ramlee Kamarudin, and Faizah Nor  
Wireless Communication Center  
Universiti Teknologi Malaysia (UTM)  
81310, Skudai, Malaysia  
Emails: ramlee@fke.utm.my  
faizah@fppsm.utm.my

**Abstract**—Currently, the orthogonal frequency division multiplexing (OFDM) systems use a predetermined cyclic prefix (CP) that is conservatively designed for the longest anticipated delay spread to overcome the multipath propagation delays. The most important parameter for determining the CP is power delay profile which is widely accepted to be following a negative exponentially decaying pattern. In this paper, the key parameter root mean square (r.m.s) delay spread of the power delay profile is mathematically derived based on the exponentially decaying power delay profile. The 3rd Generation Partnership Project (3GPP) power delay profiles are fitted into the exponentially decaying power delay profile (EDPDP). The performance measure bit error rate (BER) is used to evaluate the effectiveness of the EDPDP and its r.m.s delay spread. The findings show that EDPDP can be used to characterize the most of the power delay profiles. Subsequently, a mathematical formula to calculate CP estimation is derived. As a result, it is found that the CP is the natural logarithm of the ration of maximum power to minimum power of a particular power delay profile multiplied by its r.m.s delay spread. This finding gives the relationship between the maximum access delay and r.m.s delay spread as far as CP is concerned.

**Index Terms**—Ofdm; cyclic prefix; delay spread.

## I. INTRODUCTION

New generation of wireless mobile radio systems aims to provide higher data rates to the mobile users as well as serving many users. Therefore, adaptation methods are becoming popular for optimizing mobile radio system transmission and reception at the physical layer as well as at the higher layers of protocol stack. These adaptive algorithms offer performance improvement, better radio coverage and high data rates with low power consumption. Several adaptation schemes require a form of measurement or estimation of one or more variables that may change over time [1]. Therefore, one way of increasing the spectral efficiency is to adapt the length of the cyclic prefix (CP) which varies depending on the radio environment [2]. The most important parameter for determining the CP is power delay profile which is widely accepted to follow a negative exponentially decaying pattern [3, 4]. Considering the fact that the multipath effect is highly dependent on the deployed environment in which the wireless system operates, the width of a CP is chosen in such a way that it is larger than the maximum access delay of the propagation channel.

The CP is determined by the maximum access delay or by the root mean square (r.m.s) delay spread of that environment multiplied by a constant in the range between two to four as a rule of thumb [5].

The cyclic prefix adaptation in this study is composed of two parts. The first part is to prove that the exponentially decaying power delay profile (EDPDP) can be used to characterize power delay profiles. Then, the EDPDP model has been used to derive a mathematical formula for estimating the cyclic prefix in OFDM system. The second part of the study focuses on how the channel impulse response (CIR) can be obtained during one OFDM frame. This can be achieved by transmitting predetermined bits in the OFDM pilot subcarriers in the transmitter part. On the receiver part, the minimum mean square error (MMSE) or least square (LS) channel estimators can be used to estimate the CIR. Subsequently, the power delay profile (i.e., the energy taps) can be directly obtained from the CIR [6, 7].

In order to gauge the necessary adaptive CP, the power delay profile (energy taps) firstly should be fitted into the EDPDP model. Consequently, the CP model which is derived in this paper can be used to get the expected CP. This predicted CP can be used by the receiver in the subsequent transmitted frame. Therefore, by using this technique, it is expected to adapt the CP for each OFDM frames dynamically. Assuming that the mobile user is under low mobility condition; consequently, the power delay profile and its r.m.s delay spread are assumed to be stationary during one OFDM frame time.

This paper will focus on the first part of the study. In order to prove the validity of the proposed models, the 3rd Generation Partnership Project (3GPP) channel models are fitted into the EDPDP model. Besides, the decaying constant of the EDPDP is estimated for each power delay profiles. Subsequently, OFDM physical layer simulation is used to validate the accuracy of the proposed models (i.e., the EDPDP and the CP models) in terms of bit error rate (BER). In the transmitter part, the CP is increased in steps in order to reduce the effects of the 3GPP power delay profiles which are used to characterize the effects of the multipath signal. Next, the BER performance is plotted for these different CP values. Lastly, the CP which gives the minimum BER is compared with the

estimated CP. It is found that estimated CP is always in the vicinity of the optimum CP value.

The rest of the paper is structured as follows; in Section II, a formula for r.m.s delay spread is derived based on EDPDP model. The concept of curve fitting technique is explained in Section III. A new CP formula is adopted in Section IV. The OFDM physical layer simulation which is used to validate the effectiveness of the proposed models is clearly explained in Section V. The Performance of the EDPDP and CP models in terms of BER are discussed in Section VI. Meanwhile the main findings are elucidated in Section VII. Finally, the conclusion derived from this study is stated in Section IX.

## II. THE R.M.S DELAY SPREAD OF THE EDPDP

The delay spread is a type of distortion that is caused when identical signals arrive at different times at their destination. The signals usually arrive via multiple paths with different angles of arrival. The time difference between the moment of arrival of the first multipath component (typically the line-of-sight component) and the last multipath component is called the maximum access delay. Additionally, the r.m.s delay spread ( $\tau_{r.m.s}$ ) is defined as the square root of the second central moment of the power delay profile  $P(\tau)$ . The formula for calculating the delay spread for multipath signals is defined as [8]:

$$\tau_{rms} = \sqrt{\frac{\sum_{i=1}^N P_i \tau_i^2}{\sum_{i=1}^N P_i} - \frac{(\sum_{i=1}^N P_i \tau_i)^2}{(\sum_{i=1}^N P_i)^2}} \quad (1)$$

where  $\tau_i$  and  $P_i$  are the arrival time and power of  $i_{th}$  path respectively. In the case of continuous power delay profile, the ( $\tau_{r.m.s}$ ) is defined as [8]:

$$\tau_{rms} = \sqrt{\frac{\int_0^{\infty} (\tau - m)^2 P_h(\tau) d\tau}{\int_0^{\infty} P_h(\tau) d\tau}} \quad (2)$$

where  $m$  is the mean of access delay which is also defined as:

$$m = \frac{\int_{\tau_{min}}^{\tau_{max}} \tau P(\tau) d\tau}{\int_{\tau_{min}}^{\tau_{max}} P(\tau) d\tau} \quad (3)$$

where  $\tau_{min}$  and  $\tau_{max}$  are the arrival times of the first path and the last path respectively. As agreed by some researchers, the power delay profile is negatively exponentially decaying [3,4], it is possible to formulate the power delay profile as:

$$P = P_0 e^{-\alpha\tau} \quad (4)$$

where  $P_0$  is the power when  $\tau = 0$ ,  $\alpha$  is the decaying constant and  $\tau$  is the multipath signal arrival time. The mean of the power delay profile ( $m_1$ ) can be calculated as:

$$m_1 = \int_{\tau_{min}}^{\tau_{max}} \tau P_0 e^{-\alpha\tau} d\tau \quad (5)$$

In turn, Equation 5 is solved to produce:

$$m_1 = \left[ P_0 e^{-\alpha\tau} \left( \frac{\tau}{-\alpha} - \frac{1}{\alpha^2} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (6)$$

$$m_1 = \left[ -P_0 e^{-\alpha\tau} \left( \frac{\tau}{\alpha} + \frac{1}{\alpha^2} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (7)$$

$$m_1 = -\frac{1}{\alpha} \left[ P_0 e^{-\alpha\tau} \left( \tau + \frac{1}{\alpha} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (8)$$

Now,  $\tau_{mean}$  is defined as the mean time which can be obtained by normalizing the mean of power delay profile ( $m_1$ ) by the total sum of all arrived powers ( $I_0$ ) which can be defined as:

$$I_0 = \int_{\tau_{min}}^{\tau_{max}} P_0 e^{-\alpha\tau} d\tau = \left[ \frac{-P_0}{\alpha} e^{-\alpha\tau} \right]_{\tau_{min}}^{\tau_{max}} \quad (9)$$

From the aforementioned definitions, the mean time  $\tau_{mean}$  can be articulated as:

$$\tau_{mean} = \frac{m_1}{I_0} \quad (10)$$

$$\tau_{mean} = \frac{-\frac{1}{\alpha} \left[ P_0 e^{-\alpha\tau} \left( \tau + \frac{1}{\alpha} \right) \right]_{\tau_{min}}^{\tau_{max}}}{\left[ \frac{-P_0}{\alpha} e^{-\alpha\tau} \right]_{\tau_{min}}^{\tau_{max}}} \quad (11)$$

$$\tau_{mean} = \left[ \left( \tau + \frac{1}{\alpha} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (12)$$

Similarly, the second moment ( $m_2$ ) of the power delay profile can be defined as:

$$m_2 = \int_{\tau_{min}}^{\tau_{max}} \tau^2 P_0 e^{-\alpha\tau} d\tau \quad (13)$$

$$m_2 = \left[ -P_0 \alpha e^{-\alpha\tau} \left( \frac{\tau^2}{-\alpha} - \frac{2\tau}{\alpha^2} + \frac{2}{-\alpha^3} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (14)$$

$$m_2 = \left[ -P_0 \alpha e^{-\alpha\tau} \left( -\frac{\tau^2}{\alpha} - \frac{2\tau}{\alpha^2} - \frac{2}{\alpha^3} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (15)$$

$$m_2 = -\frac{P_0}{\alpha} \left[ \alpha e^{-\alpha\tau} \left( \tau^2 + \frac{2\tau}{\alpha} + \frac{2}{\alpha^2} \right) \right]_{\tau_{min}}^{\tau_{max}} \quad (16)$$

Accordingly, the second moment of the arrival times  $\tau_{s-moment}$  can also be expressed as:

$$\tau_{s-moment} = \frac{m_2}{I_0} \quad (17)$$

$$\tau_{s-moment} = \frac{-\frac{P_0}{\alpha} \left[ \alpha e^{-\alpha\tau} \left( \tau^2 + \frac{2\tau}{\alpha} + \frac{2}{\alpha^2} \right) \right]_{\tau_{min}}^{\tau_{max}}}{\left[ \frac{-P_0}{\alpha} e^{-\alpha\tau} \right]_{\tau_{min}}^{\tau_{max}}} \quad (18)$$

$$\tau_{s-moment} = \left[ \tau^2 + \frac{2\tau}{\alpha} + \frac{2}{\alpha^2} \right]_{\tau_{min}}^{\tau_{max}} \quad (19)$$

The r.m.s delay spread ( $\tau_{rms}$ ) is defined as the square root of the second central moment of the arrival times which is given as [8]:

$$\tau_{rms} = \sqrt{\tau_{s-moment} - \tau_{mean}^2} \quad (20)$$

$$\tau_{rms} = \sqrt{\left[ \tau^2 + \frac{2\tau}{\alpha} + \frac{2}{\alpha^2} - \left( \tau + \frac{1}{\alpha} \right)^2 \right]_{\tau_{min}}^{\tau_{max}}} \quad (21)$$

$$\tau_{rms} = \sqrt{\left[ \frac{1}{\alpha^2} \right]_{\tau_{min}}^{\tau_{max}}} \quad (22)$$

$$\tau_{rms} = \frac{1}{\alpha} \quad (23)$$

As can be seen from Equation (23), the r.m.s delay spread is inversely proportional to the decaying constant. Subsequently, the r.m.s delay spread which is just derived from the EDPDP model will be identified as ( $\hat{\tau}_{rms}$ ) in this chapter in the following sections.

### III. FITTING THE POWER DELAY PROFILES INTO THE EDPDP MODEL

Since it is required to get the value of r.m.s delay spread (i.e.,  $\alpha$ ), and the power delay profile is always a discrete profile, there is a need to use the least squares curve fitting technique to fit the power delay profile into the EDPDP model. As a case study, the power delay profiles are obtained from the 3GPP channel models as shown in Table I. These channel models are fitted into the EDPDP model; as a result, the EDPDP coefficients  $P_0$  and  $\alpha$  are estimated. For the sake of simplicity, the  $P$ ,  $P_0$ ,  $\alpha$  and  $\tau$  in Equation (4) are substituted by  $y$ ,  $A$ ,  $-B$  and  $x$  respectively, which finally gives the EDPDP model as:

$$y = Ae^{Bx} \quad (24)$$

taking the natural logarithm for both sides:

$$\ln y = \ln A + Bx \quad (25)$$

let  $a = \ln A$  and  $b = B$ , then the sum of all the errors squared ( $\epsilon^2$ ) can be expressed as:

$$\epsilon^2 = \sum_{i=1}^n y_i (\ln y_i - a - bx_i)^2 \quad (26)$$

applying the least squares fitting technique gives:

$$a \sum_{i=1}^n y_i + b \sum_{i=1}^n x_i y_i = \sum_{i=1}^n y_i \ln y_i \quad (27)$$

$$a \sum_{i=1}^n x_i y_i + b \sum_{i=1}^n x_i^2 y_i = \sum_{i=1}^n x_i y_i \ln y_i \quad (28)$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i & \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i y_i & \sum_{i=1}^n x_i^2 y_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \ln y_i \\ \sum_{i=1}^n x_i y_i \ln y_i \end{bmatrix} \quad (29)$$

$$a = \frac{\sum_{i=1}^n (y_i \ln y_i) \sum_{i=1}^n (x_i^2 y_i) - \sum_{i=1}^n (x_i y_i) \sum_{i=1}^n (x_i y_i \ln y_i)}{\sum_{i=1}^n y_i \sum_{i=1}^n (x_i^2 y_i) - (\sum_{i=1}^n x_i y_i)^2} \quad (30)$$

$$b = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n (x_i y_i \ln y_i) - \sum_{i=1}^n (x_i y_i) \sum_{i=1}^n (y_i \ln y_i)}{\sum_{i=1}^n y_i \sum_{i=1}^n (x_i^2 y_i) - (\sum_{i=1}^n x_i y_i)^2} \quad (31)$$

TABLE I  
POWER DELAY PROFILES OF 3GPP CHANNEL MODELS

RAx		TUx		HTx	
(dB)	time (ns)	(dB)	time (ns)	(dB)	time (ns)
-5.2	0	-5.7	0	-3.6	0
-6.4	42	-7.6	217	-8.9	356
-8.4	101	-10.1	512	-10.2	441
-9.3	129	-10.2	514	-11.5	528
-10.0	149	-10.2	517	-11.8	546
-13.1	245	-11.5	674	-12.7	609
-15.3	312	-13.4	882	-13.0	625
-18.5	410	-16.3	1230	-16.2	842
-20.4	469	-16.9	1287	-17.3	916
-22.4	528	-17.1	1311	-17.7	941
		-17.4	1349	-17.6	15000
		-19.0	1533	-22.7	16172
		-19.0	1535	-24.1	16492
		-19.8	1622	-25.8	16876
		-21.5	1818	-25.8	16882
		-21.6	1836	-26.2	16978
		-22.1	1884	-29.0	17615
		-22.6	1943	-29.9	17827
		-23.5	2048	-30.0	17849
		-24.3	2140	-30.7	18016

TABLE II  
THE R.M.S DELAY SPREAD VALUES USING THE TWO CALCULATION METHODS FOR 3GPPT CHANNEL MODEL

$\tau_{rms}$	RAx	TUx	HTx
Eq (1) ( $\mu s$ )	0.10001	0.50006	3.038
Eq (23) ( $\mu s$ )	0.13362	0.50048	4.220

where  $b = B$ ,  $A = \exp(a)$  and  $\alpha = -B = -b$  as previously assumed.

In Table II, the values of the r.m.s delay spread are calculated by using two methods; the first method is the direct r.m.s delay spread formula as stated in Equation (1), and the second method is as given in Equation (23).

### IV. CYCLIC PREFIX FORMULATION

In this section, a mathematical formula is derived to estimate the CP based on the basic assumption that the channel power profile is exponentially decaying, that is:

$$P = P_0 e^{-\alpha \tau} \quad (32)$$

taking the natural logarithm of both sides and making  $\tau$  as the subject of the formula:

$$\tau = \frac{\ln(P/P_0)}{-\alpha} \quad (33)$$

substituting of Equation (23) into Equation (33) yields:

$$\tau = -\tau_{rms} \ln(P/P_0) \quad (34)$$

The graph in Figure (1) shows that  $\tau_{max}$  is the arrival time of the last path which has a power level of  $\beta$ , from this definition,  $\tau_{max}$  can be expressed as:

$$\tau_{max} = -\tau_{rms} \ln(\beta/P_0) \quad (35)$$

meanwhile the value of  $\tau_{min}$  is the arrival time of the first path which has a power level of  $\gamma$ , it can also be expressed as:

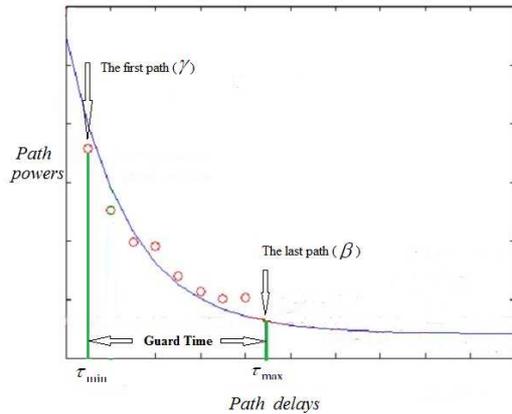


Fig. 1. Illustration Diagram Shows the Maximum Access Delay and CP Formulation Concept

$$\tau_{min} = -\tau_{rms} \ln(\gamma/P_0) \quad (36)$$

Now, the required CP based on maximum access delay concept is defined as:

$$CP = \tau_{max} - \tau_{min} \quad (37)$$

Substitution of Equations (35) and (36) into Equation (37) yields:

$$CP = \tau_{rms} (\ln \gamma/P_0 - \ln \beta/P_0) \quad (38)$$

$$CP = \tau_{rms} \ln \left( \frac{\gamma}{\beta} \right) \quad (39)$$

In Equation (39), the CP is a function of  $\gamma$ , whereby  $\gamma$  is the power of the first path as defined earlier. Interestingly, the value of  $\gamma$  depends on the propagation distance between the mobile station and the base station. Subsequently, the propagation distance is a function in the long term fading model. Accordingly, it is possible to hypothesize that there is a connection between the proposed CP model and the long term fading model.

## V. THE OFDM PHYSICAL LAYER SYSTEM DESIGN

In order to evaluate the performance of the proposed models, a detailed link level simulation has been developed to investigate the physical layer performance of the IEEE 802.16e air interface. The functional blocks used in the transmitter and receiver chain of the link level simulation are shown in Figure (2).

Functional blocks of the OFDM system are implemented according to IEEE 802.16-2004 and IEEE 802.16-2005 with the exception of pulse shaping, which is outside the scope of standards. The first box in the block is the Bernoulli binary generator which is the information bits generator. The function of the second box is for channel coding which is composed of Forward Error Correction (FEC) and interleaving. For FEC,

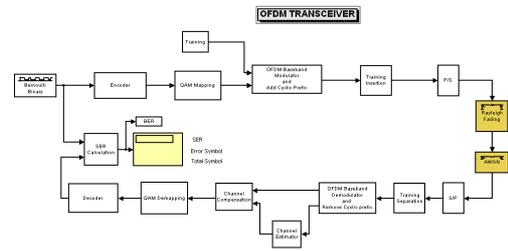


Fig. 2. The OFDM transceiver block diagram

the conventional encoder is used, which is the only mandatory coding scheme according to IEEE 802.16e-2005 specification. The conventional encoder mother code rate (R0) is and its memory size is 6. After bit interleaving is used, the system uses the 16 PSK mapping scheme. At the input of the Inversed Fast Fourier Transform (IFFT), the data selector is used to form the OFDM symbol which consists of 1024 subcarriers (i.e.,  $N_{fft} = 1024$ ). This OFDM symbol is constructed from 720 data subcarriers, 120 pilot subcarriers, 183 guard subcarriers and one direct current (DC) subcarrier which is used as a center frequency. Before transmitting the OFDM symbol, CP is added it in order to mitigate the effect of the multipath. This CP is added at the front of the OFDM symbol which is a duplication of the tail of the useful symbol with ratios of  $G = 1/512, 1/256, 1/128, 1/64, 1/32, 1/16, 1/8, 1/4, 1/2$ ; where G is the CP to useful symbol time ratio. The other system parameters are the channel bandwidth  $BW = 10MHz$ , sampling factor  $n = 28/25$ , sampling frequency  $F_s = (n \times BW/8000) \times 8000 = 11.2MHz$ , subcarrier spacing  $f = F_s/N_{fft} = 10.9375$ , useful symbol time  $T_b = 1/f = 91.42857\mu s$ ,  $CP = G \times T_b$  and OFDM symbol time  $T_s = T_b + CP$ . The value CP and  $T_s$  depend on the values of G as shown in Table (III).

TABLE III  
THE RELATIONSHIP BETWEEN THE PARAMETER G AND AND THE CP

The ratio G	Cyclic Prefix ( $\mu s$ )	OFDM symbol time ( $T_s$ ) ( $\mu s$ )
1/512	0.17	91.6
1/256	0.35	91.7
1/128	0.7	92
1/64	1.4	93
1/32	2.8	94
1/16	5.7	97
1/8	11.4	102
1/4	22.8	114
1/2	45.7	137

## VI. THE PERFORMANCE OF THE CYCLIC PREFIX MODEL

The purpose of this section is to validate the efficiency of the CP model using the 3GPP channel models which consist of three power delay profiles. The first power delay profile is the Rural Area channel model (RAx), the second one is the Typical Urban channel model (TUx) and the last one is the Hilly Terrain channel model (HTx) as shown in Table (I). The 3GPP channel models are fitted into the EDPDP

and the r.m.s delay spread is calculated using the standard formula and the proposed formula as shown in Table (II). The curves in Figure (3) show the BER of RAX power delay profile. It can be clearly seen that the minimum BER occurs when  $G = 1/128$  and its corresponding  $CP = 0.71429\mu s$  as shown in Table (III). On the other hand, the estimated ( $CP$ ) is  $CP = \hat{\tau}_{rms} \ln\left(\frac{\gamma}{\beta}\right) = 0.528\mu s$  as shown in Table (IV). Additionally, it can be observed that when the CP increases (i.e.,  $G$  increases to  $1/64$ ), there is no enhancement in BER performance. This means that the optimal CP occurs when  $G = 1/128$ . Moreover, it is observed that when the CP increases ( $G$  increases), the BER performance degrades. This is due to the fact that the BER is influenced by two factors. The first factor is inter-symbol-interference (ISI) and the second factor is the power consumption. When the CP increases, the effect of ISI is eliminated. Nevertheless, since the CP is considered as an additional transmitted data, transmitting these additional data needs to increase the transmitted power in order to maintain the same BER performance. In other words, increasing the CP size needs to increase transmitted power in order to keep the same BER. Therefore, increasing CP size while keeping the same amount of power for the OFDM frame, this will leads to a degradation in BER performance.

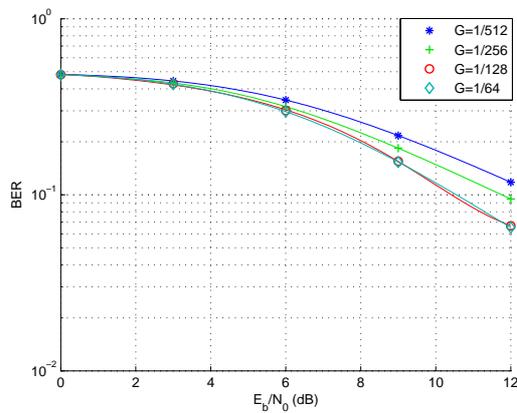


Fig. 3. BER Performance of 3rd Generation Partnership Project (3GPP)-Rural Area Channel Model (RAX) for Different CP Values

Focusing now on Figure (4), it can be clearly seen that the minimum BER of TUX occurs when  $G = 1/32$  and its corresponding  $CP = 2.8\mu s$  as shown in Table (III), meanwhile the estimated CP is  $CP = \hat{\tau}_{rms} \ln\left(\frac{\gamma}{\beta}\right) = 2.14\mu s$  as illustrated in Table (IV).

Moving on, the curves in Figure (5) show that the minimum BER for HTX occurs when  $G = 1/4$  and its corresponding  $CP = 22.8\mu s$  as shown in Table (III), meanwhile the estimated  $CP = \hat{\tau}_{rms} \ln\left(\frac{\gamma}{\beta}\right) = 18.01\mu s$  as depicted in Table (IV). Although there are some differences between the values of the r.m.s delay spread standard formula and the values of the proposed formula, the RAX and HTX estimated CP values are still positioned in the right range. On the other hand, the value of r.m.s delay spread of the standard formula

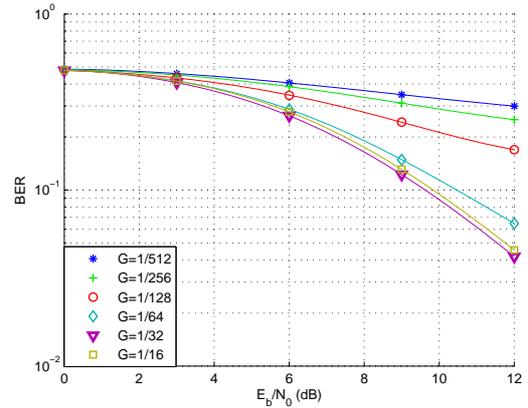


Fig. 4. BER Performance of 3rd Generation Partnership Project (3GPP)-Typical Urban Channel Model (TUX) for Different CP Values

and the value of r.m.s delay spread of the proposed formula for TUX almost have the same values. This is due to the fact that TUX measured data is smoothly decaying, meanwhile there is a slight divergence in RAX and HTX measured data as it can be clearly seen in Table (I).

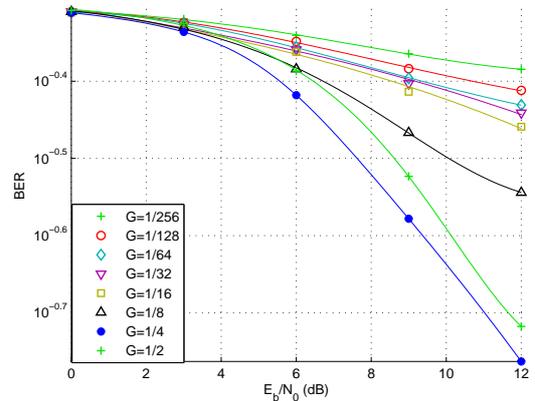


Fig. 5. BER Performance of 3rd Generation Partnership Project (3GPP)-Hilly Terrain Channel Model (HTX) for Different CP Values

TABLE IV  
THE VALUES OF THE PROPOSED CP MODEL PARAMETERS FOR 3GPP CHANNEL MODELS

Parameters	Formula	RAX	TUX	HTX
$P_0$	$P = P_0 e^{-\alpha\tau}$	306.87	268.99	175.46
$\gamma$	$\gamma = P_0 e^{-\alpha\tau_{min}}$	306.87	268.99	175.46
$\beta$	$\beta = P_0 e^{-\alpha\tau_{max}}$	5.90	3.73	2.45
R	$R = \ln\left(\frac{\gamma}{\beta}\right)$	3.95	4.27	4.26
$CP(\mu s)$	$CP = \tau_{rms} \ln\left(\frac{\gamma}{\beta}\right)$	0.52	2.14	18.01

An important finding derived from this study, the ratio  $R = \ln\left(\frac{\gamma}{\beta}\right)$  is approximately in the vicinity of the value 4 for all the 3GPP power delay profiles as shown in Table (IV). As aforementioned, the proposed CP model is derived based on maximum access delay, and it is found that  $CP = 4 \times \tau_{rms}$

as shown in Table (IV). As a result, the proposed CP can be formulated as  $CP = 4 \times \tau_{rms}$ . This result of this study seemed to confirm the finding of a study by Arslan who found that the CP equals  $\tau_{rms}$  multiplied by a constant in the range between two to four [5]. As a result, the calculation of CP based on maximum access delay or based on r.m.s delay spread multiplied by 4 will give the same result.

## VII. FINDINGS

It is possible to consider  $\gamma$  as the main signal meanwhile the rest of the multipath signals can be considered as interference signals (i.e.,  $\beta$ ). Therefore, the main contribution of this paper, it gives the required span between the main signal  $\gamma$  and the others interference signals. This span can be used to eliminate the ISI when the r.m.s delay spread is used to calculate the CP. This finding is parallel with the finding in [10, pp. 77-78], in which the CP length is proposed to be adjusted as:

$$CP = \beta_0 \tau_{rms} \quad (40)$$

Therefore, when we compare Equation (39) with Equation (40), we will find that:

$$\beta_0 = \ln \left( \frac{\gamma}{\beta} \right) \quad (41)$$

As a result, Equation (41) can be used to get the relationship between CP and required span instead of the look up table that has been proposed in [11].

## VIII. DISCUSSIONS OF THE EDPDP AND CP MODELS

What is interesting in Equation (23) is that the r.m.s delay spread depends on the decaying constant  $\alpha$ . The decay of the multipath signals depends on the attenuation caused by reflected and diffracted objects that surround the mobile station. This leads to an important fact that the r.m.s delay spread depends on the environment that surrounds the mobile station[5].

Next, as seen in Table (III) and Table (IV), the EDPDP model gives an acceptable level of precision for the r.m.s delay spread values indicating that all other types of power delay profiles can be represented by the EDPDP. This assumption is compatible with electromagnetic propagation phenomenon where the electromagnetic multipath signals reach the receiver after being reflected or diffracted from the surrounding walls of the buildings. In fact, the multipath signals are propagating through different propagation distances; therefore, they have different power levels. These power levels are inversely proportional to the propagation distances; as the propagation distance increases, the propagation time delay also increases. Therefore, the multipath signals should decay as their propagation distances increase. Moreover, the multipath signals encounter the same amount of power absorption loss. However, it is possible to find some paths that are reflected; for instance, from glass windows and metal doors, these paths that

are reflected from such surfaces face an absorption loss slightly different from the first category. As a result, it is possible to find some paths arriving late but with power levels still higher than the ones which arrived earlier. However, the curve fitting technique will compensate for this kind of effects.

## IX. CONCLUSION

In conclusion, the EDPDP model is proposed to characterize the power delay profiles, and its r.m.s delay spread is mathematically derived. The required CP for mitigating the influence of ISI in OFDM is also formulated. Fitting the power delay profiles into the EDPDP and using Equation (39), the necessary required adaptive CP can be achieved. In addition, there are two methods that are being used to estimate the CP; the first one is based on maximum access delay and the second is based on r.m.s delay spread multiplied by 4. The proposed CP model clearly shows that there is a mathematical relationship between these two methods. In addition, such proposed CP model confirms that these two methods lead to the same result.

## ACKNOWLEDGMENT

This paper was prepared under support of Ministry of Higher Education (MOHE), conducted by Universiti Teknologi Malaysia (UTM) under vot. no: 4F036. Therefore, the authors would like to extend their sincere gratitude to MOHE and UTM.

## REFERENCES

- [1] Jiunn-Tsair Chen, Jenwei Liang, Huan-Shang Tsai, and Young-Kai Chen, *Joint MLSE receiver with dynamic channel description*, IEEE Journal on Selected Areas in Communications, vol. 16, no. 9, pp. 1604-1615, Dec 1998.
- [2] Bernard Henri Fleury, *An uncertainty relation for WSS processes and its application to WSSUS systems*, IEEE Transactions on Communications, vol. 44, no. 12, pp. 1632-1634, Dec 1996.
- [3] Henrik Schober, Riedrich Jondral, Richared A. Stirling-Gallacher, and Zhaocheng Wang, *Delay Spread Estimation for OFDM Based Mobile Communication Systems*, In Proceedings of the European Wireless Conference, 2002.
- [4] Jan Hansen, *An analytical calculation of power delay profile and delay spread with experimental verification*, IEEE Communications Letters, vol. 7, no. 6, pp. 257- 259, June 2003.
- [5] Huseyin Arslan, *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*, pp. 238, 2007, Dordrecht, Netherlands, Springer.
- [6] Jan-Jaap van de Beek, Ove Edfors, Magnus Sandell, Sarah Kate Wilson, and Per Ola Borjesson, *On channel estimation in OFDM systems*, Vehicular Technology Conference, 1995 IEEE 45th, vol. 2, no. 2, pp. 815-819, 25-28 Jul 1995.
- [7] Zhang Zhao-yang and Lai Li-feng, *A novel OFDM transmission scheme with length-adaptive Cyclic Prefix*, Journal of Zhejiang University, pp. 1336-1342, 2004.
- [8] Peyton Peebles, *Probability, Random Variables and Random Signal Principles*, pp. 81-82, 4th-ed McGraw-Hill, Singapore, 2001.
- [9] Fernando Perez Fontn and Perfecto Marino Espineira, *Modeling the Wireless Propagation Channel*, pp. 159-164, John Wiley & Sons, Ltd, 2008.
- [10] Andrea Goldsmith, *Wireless Communication*, Cambridge University Press.
- [11] Andrea M. Tonello, Salvatore D'Alessandro, and Lutz Lampe, *Cyclic Prefix Design and Allocation in Bit-Loaded OFDM over Power Line Communication Channels*, IEEE Transactions on Communications, vol. 58, no. 11, pp. 3265-3276, November 2010.

# An Effective transmit Weight Design for DPC with Maximum Beam in Multi-user MIMO Downlink

Cong Li, Yasunori Iwanami

Department of Computer Science and Engineering, Graduate School of Engineering

Nagoya Institute of Technology, Nagoya, Japan

E-mail: chx17515@stn.nitech.ac.jp, iwanami@nitech.ac.jp

**Abstract**—A novel transmit scheme is proposed for the cancelation of Inter-User Interference (IUI) and the improvement of achievable sum-rate of system in Multi-User Multiple-Input Multiple Output (MU-MIMO) downlinks. Particle Swarm Optimization (PSO) algorithm is employed to solve the constrained nonlinear optimization problem for obtaining the optimal transmit weights. Compared with the conventional Dirty Paper Coding (DPC) having single receive antenna, the proposed scheme applies the principle of DPC to cancel the IUI in MU-MIMO downlinks, where the Mobile Station (MS) users are equipped with multiple receive antennas. With the Channel State Information (CSI) known at the Base Station (BS) and the MS, the eigenvalues for all user channels are calculated first and then the user with the maximum eigenvalue is selected as the 1-st user. For the remaining users, the users are sequentially processed and the transmit weights are generated from the previously selected users by using the Particle Swarm Optimization (PSO) algorithm, which makes the transmit gain for each user as large as possible. The analysis of computational load and simulation results verify the effectiveness of the proposed scheme.

**Keywords**—Multi-user MIMO downlink, Capacity, Dirty Paper Coding (DPC), Particle Swarm Optimization (PSO).

## I. INTRODUCTION

Recently, MU-MIMO schemes have attracted considerable interests toward the next generation wireless communication systems MU-MIMO because of its potential for increasing capacity [1]-[5]. In scenario of MU-MIMO downlink, the base station transmits spatially multiplexed signals to multiple users simultaneously, and attempts to transmit independent signal streams to all users over the same frequency. However, it suffers from the Inter-User Interference (IUI) in the received signal at each user. To mitigate the IUI, many linear precoding techniques have been employed to eliminate the IUI, such as the channel inversion [4] and Block Diagonalization (BD) schemes [5]. The Channel Inversion (CI) method pre-processes the transmit signals with the channel pseudo-inverse to suppress the IUI. In the case of highly independent channels, Zero Forcing (ZF) transmit weight can completely remove the IUI, but it leads to the noise enhancement at the receiver and reduces of achievable rate. Though the transmit weight with Minimum Mean-Squared Error (MMSE) criterion, can achieve the throughput increase, however it results in vestigial level of IUI. BD scheme is a well-known linear precoding technique, which decomposes MU-MIMO channels into multiple parallel SU-MIMO channels to completely cancel the IUI by using orthogonal space. However, these schemes impose a restriction on the system

configuration in respect of the number of receive antennas, that is, only the number of transmit antennas at BS is larger than the sum of receive antennas of all users, BS can provided enough degrees of freedom to force the IUI to zero at each MS. In addition, the correlation among users also degrades the system performance. Thus we should take care to avoid the spatial multiplexed users with highly correlated channel.

DPC technique not only approaches the capacity more closely, but also suppresses the IUI completely [6]-[8]. However, the conventional DPC based on QR decomposition is restricted to the case of single receive antenna at each MS. The Maximum Beam (MB) transmit scheme in [9] utilizes the eigenvector corresponding to the maximum eigenvalue of desired user channel as the transmit weight, but the IUI among users is the serious problem. To assure the independence among users, especially in the case of many users, the author gave an imperfect block diagonalization method to reduce the IUI by employing Gram-Schmidt orthonormalization on transmit weights. But this method needs to keep the balance between IUI and transmit gain. In [10] a jointly design of a transmit scheme is proposed, where the near-orthogonal effective channels are success-sively obtained from 1-th user to the last user by subtracting the interference components from the preceding users. However, the interference among users cannot be cancelled completely because of near orthogonality.

Particle Swarm Optimization (PSO), first proposed by Kennedy and Eberhart in 1995 [11], is developed from the swarm intelligence and based on the research of flock movement behaviour of birds flock finding foods [12]-[15]. PSO is employed in many optimization problems, especially, in the optimization of continue space, it shows the better computational efficiency than the other algorithm, such as Genetic Algorithm (GA). Recently, the optimal algorithms PSO have been considerable interest in Multi-user MIMO. A PSO aided optimal Multi-user MIMO linear precoding scheme is proposed in [14], where PSO is used to search the optimal transmit weight which make SINR maximum at each MS. Though the IUI can be completely moved by multiplying decoder matrix, it leads noise enhancement and large computational load. In this paper, we proposed a novel MU-MIMO downlinks transmit scheme not only to obtain the optimal transmit weights for each user but also to eliminate the IUI completely. We first select the user with the maximum eigenvalue as the 1-st user by calculating eigenvalues of all users. Then we obtain the possible transmit weight for the next user by using the PSO algorithm, which makes sure that the selected user has an

optimal accessible SNR. The remaining transmit weights are generated by the same method as the previous users. In the proposed scheme, we introduced the DPC principle to remove the IUI at the BS.

This paper is organized as follows. In Section II, we will present the system model of MU-MIMO. In Section III, we propose the optimal transmit weight based on the PSO algorithm for each user. The simulation results are given in Section IV, in which the proposed scheme shows the excellent BER performance compared with the conventional schemes. Finally the conclusions will be made in Section V. We illustrate some of the notations as follows; vectors and matrices are expressed by bold letters, we use  $E[\bullet]$ ,  $[\bullet]^T$  and  $[\bullet]^H$  as the expectation, transpose and conjugate transpose of matrix, respectively.

## II. SYSTEM MODEL

We consider the downlink multi-user MIMO system with  $N_T$  transmit antennas and  $n_R^{(k)}$  receive antennas at the  $k$ -th user, as shown in Fig. 1, where  $N_u$  is the number of multiple antenna users. In this paper, we focus on the flat Rayleigh (i.i.d.) fading channel model, because the wideband scheme, such as the OFDM, can be accommodated at each frequency index. We assume that the MIMO channel matrix  $\mathbf{H}_k \subset \mathbb{C}^{n_R^{(k)} \times N_T}$  ( $k=1, 2, \dots, N_u$ ) is available at BS and MS.

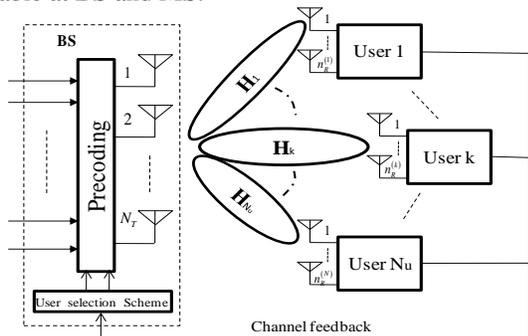


Fig. 1 Downlinks of multi-user MIMO

The  $n_R^k \times 1$  received signal at the  $k$ -th user is written as

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{M}_k \mathbf{x}_k + \sum_{j=1, j \neq k}^{N_u} \mathbf{H}_k \mathbf{M}_j \mathbf{x}_j + \mathbf{n}_k \quad (1)$$

where  $\mathbf{x}_k$  and  $\mathbf{n}_k$  are transmit signal vector additive Gaussian noise and  $\mathbf{M}_k$  is transmit weight at the  $k$ -th user. In this paper, we only focus on one transmit stream for each user and assume that

$$E(\mathbf{x}_k \mathbf{x}_k^H) = \mathbf{I}, E(\mathbf{n}_k \mathbf{n}_k^H) = \sigma_n^2 \mathbf{I}, E(\mathbf{x}_k \mathbf{n}_k^H) = \mathbf{0} \quad (2)$$

## III. PROPOSED OPTIMAL TRANSMIT WEIGHTS BASED ON PARTICLE SWARM OPTIMIZATION

The principle of conventional DPC based on QR decomposition indicates that the sufficient condition of the feasibility of DPC is the existence of lower or upper triangular matrix derived from the channel matrix. Since the IUI cannot be removed at the receiver in the case of MB, we

carefully design the transmit weight for each user to transform the effective channels to the lower or upper triangular matrices. Then we can use DPC to remove the IUI completely. However, how to obtain the optimal transmit weight becomes a problem, referred to as the problem of optimal transmit weight design.

### A. Transmit scheme for eliminating IUI

We design the transmit weights for each user as the following steps. First, we employ Singular Value Decomposition (SVD) for on user channels.

$$\mathbf{H}_k = \mathbf{U}_k \mathbf{A}_k [\mathbf{v}_k^1, \mathbf{v}_k^2, \dots, \mathbf{v}_k^{N_T}]^H, (k=1, 2, \dots, N_u) \quad (3)$$

To achieve the optimal transmit weight for the 1-st user, we perform the following algorithm.

$$\mathbf{H}_k, \mathbf{v}_k^1 = \arg \max_{\mathbf{H}_k} (\|\mathbf{H}_k \mathbf{v}_k^1\|), k \in [1, N_u] \quad (4)$$

Then, the transmit weight for the 1-st user is given by

$$\mathbf{M}_1 = \mathbf{v}_k^1, o(1) = k \quad (5)$$

where  $o(1)$  denotes the number of the selected user in (4). Then the channel matrix of the selected user is arranged on the first layer of the system channel matrix.

$$\mathbf{H} = [\mathbf{H}_{o(1)}^T \quad \mathbf{H}_2^T \quad \dots \quad \mathbf{H}_{N_u}^T]^T \quad (6)$$

In this way, we ensure that the 1-st user uses the MB. For the 2-nd user, in order to use the DPC for eliminating the IUI, we have to transform the effective channel matrix to triangular matrix, and the following operation is performed.

$$(\mathbf{H}_{o(1)} \mathbf{M}_1)^H \mathbf{H}_{o(1)} \stackrel{\text{SVD}}{=} \mathbf{U} \mathbf{A} [\mathbf{v}^1, \mathbf{v}^2, \dots, \mathbf{v}^{N_T}]^H \quad (7)$$

Null space

We can obtain the transmit weight for the 2-nd user from the null space in (7). However, we have to notice that all of these vectors in null space can meet the condition of the availability of DPC. The optimal transmit weight for the 2-nd user can be expressed as follows.

$$\mathbf{M}_2 = \sum_{l=1}^{N_u-1} \alpha_l \mathbf{v}^{l+1} / \left\| \sum_{l=1}^{N_u-1} \alpha_l \mathbf{v}^{l+1} \right\| \quad (8)$$

where  $\alpha_l$  ( $l=1, 2, \dots, N_u$ ) is real coefficient. Then we estimate the channel of the 2-nd user corresponding to  $\mathbf{M}_2$  in (8). It can be obtained by the following algorithm.

$$\mathbf{H}_k = \arg \max_{\mathbf{H}_k} (\|\mathbf{H}_k \mathbf{M}_2\|); k \in [1, N_u], k \neq o(1) \quad (9)$$

by letting  $o(2) = k$ , the 2-nd user is determined. Similarly, the channel matrix corresponding to the selected user in (9) is arranged on the second layer of channel matrix in (6).

We consider the  $k$ -th user and its transmit weight, the same operations as the above are performed to the matrix corresponding to these  $k-1$  selected users.

$$\begin{bmatrix} (\mathbf{H}_{o(1)} \mathbf{M}_1)^H \mathbf{H}_{o(1)} \\ (\mathbf{H}_{o(2)} \mathbf{M}_2)^H \mathbf{H}_{o(2)} \\ \vdots \\ (\mathbf{H}_{o(k-1)} \mathbf{M}_{k-1})^H \mathbf{H}_{o(k-1)} \end{bmatrix} \stackrel{\text{SVD}}{=} \mathbf{U} \mathbf{A} [\mathbf{v}^1, \dots, \mathbf{v}^{k-1}, \underbrace{\mathbf{v}^k, \dots, \mathbf{v}^{N_u}}_{\text{Null space}}]^H \quad (10)$$

from the null space in (10), the transmit weight for  $k$ -th user can be given by

$$\mathbf{M}_k = \sum_{l=1}^{N_u-k+1} \alpha_l \mathbf{v}^{k+l-1} / \left\| \sum_{l=1}^{N_u-k+1} \alpha_l \mathbf{v}^{k+l-1} \right\| \quad (11)$$

$$\mathbf{H}_k = \arg \max_{\mathbf{H}_k} (\|\mathbf{H}_k \mathbf{M}_k\|); k \in [1, N_u], k \neq o(1), o(2), \dots, o(k-1) \quad (12)$$

Similarly, the user channel obtained in (12) is arranged on the  $k$ -th layer of the system channel.

Consequently, we obtain the transmit weights for all users, and the transmit signal vector can be rewritten as

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_u} \end{pmatrix} = \begin{pmatrix} s_1 \\ s_2 - \frac{1}{\|\mathbf{H}_2 \mathbf{M}_2\|} \sum_{j<2} \|\mathbf{H}_2 \mathbf{M}_j\| x_j \\ \vdots \\ s_{N_u} - \frac{1}{\|\mathbf{H}_{N_u} \mathbf{M}_{N_u}\|} \sum_{j<N_u} \|\mathbf{H}_{N_u} \mathbf{M}_j\| x_j \end{pmatrix} \quad (13)$$

where  $s_k$  is the transmitted information for the  $k$ -th user. To avoid the power enhancement problem, the same modulo device as in [8] is used in this paper.

Then we discuss the Signal to Noise Ratio (SNR) at each user terminal. If the Maximum Ratio Combining (MRC) is used to detect the receive signal, the SNR at each MS is given by

$$\langle \text{SNR} \rangle_k = \|\mathbf{H}_k \mathbf{M}_k\|^2 \cdot \|s_k\|^2 / N_k \quad (14)$$

where  $\|\mathbf{H}_k \mathbf{M}_k\|$  can be considered as the transmit gain of the  $k$ -th user. If  $\|\mathbf{v}_k^{\max} \mathbf{M}_k\| = 1$ ,  $\mathbf{M}_k$  is considered as the ideal transmit weight which enables the  $k$ -th user to get the largest transmit gain  $\lambda_k^{\max^2}$ , and  $\mathbf{v}_k^{\max}$  is the vector corresponding to the maximum eigenvalue of  $\mathbf{H}_k$ . However, in fact,  $\|\mathbf{v}_k^{\max} \mathbf{M}_k\|$  lies in between 0 and 1, and it depends on the coefficients  $\alpha_i$  for combining  $\mathbf{M}_k$ , where  $\alpha_i \in R$  lies in a certain lattice. In order to obtain the optimal transmit gain, it is crucial to search the optimal combination coefficients which makes  $\mathbf{M}_k$  parallel to  $\mathbf{v}_k^{\max}$ . We solve the nonlinear optimization problem by employing PSO algorithm to obtain the optimal transmit weights. In this case, if we use the DPC principle to cancel the IUI, the achievable rate can reach optimal. The capacity of proposed scheme can be expressed as

$$\begin{aligned} C &= \max_{\mathbf{M}_k, \mathbf{H}_k \mathbf{M}_j=0, k>j} \log \left| \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{H} \mathbf{M} \mathbf{M}^H \mathbf{H}^H \right| \\ &= \max_{\mathbf{H}_k \mathbf{M}_j=0, k>j} \sum_{k=1}^{N_u} \log \left| \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{H}_k \mathbf{M}_k \mathbf{M}_k^H \mathbf{H}_k^H \right| \end{aligned} \quad (15)$$

The water filling can be used to distribute transmit power. We define the achievable sum-rate of system with users equipped with single receive antenna as

$$C = \sum_{k=1}^{N_u} \left[ \log_2 \mu \|\mathbf{H}_k \mathbf{M}_k\|^2 \right]_+$$

and  $\mu$  solves

$$\sum_{k=1}^{N_u} \left[ \mu - 1 / \|\mathbf{H}_k \mathbf{M}_k\|^2 \right]_+ = P_T \quad (16)$$

where  $P_T$  is the total transmitting power.

### B. Particle Swarm Optimization aided optimal transmit weights

The PSO algorithm, based on the research of flock movement behaviour of birds flock finding foods, where the random flying bird flock are referred to as particles, which represent potential solutions initialized over the whole search space randomly. An objective function is used to evaluate the goodness of positions of particles. Each particle has a fitness value which is evaluated by the objective function to be optimized, and the fitness value is evaluated at each iteration search.

For  $k$ -th user, we assume that the size of swarm and the dimension of search space are  $\Omega_k$  and  $D_k$ , respectively. The position and velocity of particle  $i$  ( $i=1, 2, \dots, \Omega_k$ ) are denoted as  $\chi_{i,k} = [\chi_{i1} \chi_{i2} \dots \chi_{iD_k}]^T$  and  $\zeta_{i,k} = [\zeta_{i1} \zeta_{i2} \dots \zeta_{iD_k}]^T$ . In the course of searching optimal solution, the particle  $i$  knows its best position  $\mathbf{P}_{i,k} = [p_{i1} p_{i2} \dots p_{iD_k}]^T$ , which provides the best information, and the best position so far among the entire particles, is denoted as  $\mathbf{P}_{g,k} = [p_{g1} p_{g2} \dots p_{gD_k}]^T$ , which provides the global best location.  $\mathbf{P}_{i,k}$  and  $\mathbf{P}_{g,k}$  are updated at each iteration by the objective function. Each particle has its own velocity to direct its flying, which relies on its previous speed  $\zeta_{id_k}$  ( $d_k=1, 2, \dots, D_k$ ) as well as its  $p_{id_k}$  and  $p_{gd_k}$  information. At each iteration, the velocity and position of particle  $i$  are updated based on the following equations.

$$\begin{aligned} \zeta_{id_k}^{t+1} &= w \zeta_{id_k}^t + c_1 \varphi_1 (p_{id_k}^t - \chi_{id_k}^t) + c_2 \varphi_2 (p_{gd_k}^t - \chi_{id_k}^t) \\ \chi_{id_k}^{t+1} &= \chi_{id_k}^t + \zeta_{id_k}^{t+1} \end{aligned} \quad (17)$$

where  $t$  is the current iteration number,  $\zeta_{id_k}^t$  and  $\chi_{id_k}^t$  denote the velocity and location of the particle  $i$  in  $d_k$  dimensional space, respectively.  $p_{id_k}^t$  is the individual best location that the particle  $i$  has achieved so far, and  $p_{gd_k}^t$  is the global best location that all particles have achieved so far at the  $i$ -th iteration.  $w$  is the inertia weight which determines to what extent the particle remains along its original course unaffected by the influence of  $p_{id_k}^t$  and  $p_{gd_k}^t$ , it is usually set between 0 and 1.  $c_1$  and  $c_2$  are acceleration constants which are set to 2.  $\varphi_1$  and  $\varphi_2$  are uniformly distributed random number in  $[0,1]$ . This iterative search process will be repeated up to the maximal iteration number or till the termination criterion is satisfied.

Then we have to consider two important issues, one is the particle modelling, and the other is the selection of objective function. Considering the SNR in equation (14), we search the optimal transmit weight  $\mathbf{M}_k$ , which infinitely close to the ideal transmit weight  $\mathbf{v}_k^{\max}$  for the  $k$ -th user,  $\mathbf{v}_k^{\max}$  where  $\mathbf{v}_k^{\max}$  denotes the desired point lying in the best location. So

we define the angle between  $\mathbf{v}_k^{\max}$  and  $\mathbf{M}_k$  as  $\theta_{\mathbf{v}_k^{\max}, \mathbf{M}_k}$  and consider the sine function as the objective function for  $k$ -th user at the  $t$ -th iteration.

$$\left\{ \begin{array}{l} \alpha_t = \chi_{it}^t \\ \mathbf{M}_k = \sum_{l=1}^{N_u-k+1} \alpha_l \mathbf{v}^{k+l-1} / \left\| \sum_{l=1}^{N_u-k+1} \alpha_l \mathbf{v}^{k+l-1} \right\| \\ f_{i,k}^t = \text{function}(\mathbf{v}_k^{\max}, \mathbf{M}_k) \\ \quad = \sin(\theta_{\mathbf{v}_k^{\max}, \mathbf{M}_k}) = \sqrt{1 - \cos^2(\theta_{\mathbf{v}_k^{\max}, \mathbf{M}_k})} \\ \quad = \sqrt{1 - \left\| (\mathbf{v}_k^{\max})^H \mathbf{M}_k \right\|^2} \end{array} \right. \quad (18)$$

Here, we use the sine value of angle between the ideal weight  $\mathbf{v}_k^{\max}$  and the tentative weight  $\mathbf{M}_k$  to measure the degree of approach. During the process of search, the sine value becomes small when the tentative weight gets close to the desired weight, and the search speed and accuracy are conditioned on the size of swarm, iteration number and search dimension. In this paper, the proposed optimal transmit weights based on PSO algorithm is obtained by the following steps.

(a). BS obtains the information of the channels by the feedback from MS and computes the maximum eigenvalues  $\lambda_k^{\max}$  of each user and determines the order of user. By eq. (10), the dimension of search space of PSO is determined for the next user, and then the swarm size  $\Omega_k$  and the maximum iteration number  $I_{\max}^k$  are set for each user.

(b). PSO algorithm is employed to search the optimal transmit weight for each user. For the  $k$ -th user in  $\Omega_k$ , we initialize the velocity and location for each particle as  $\boldsymbol{\zeta}_{i,k}^1 = [\zeta_{i,1}^1, \dots, \zeta_{i,l}^1, \dots, \zeta_{i,D_k}^1]^T$  and  $\boldsymbol{\chi}_{i,k}^1 = [\chi_{i,1}^1, \chi_{i,2}^1, \dots, \chi_{i,D_k}^1]^T$ , where the initial velocity  $\boldsymbol{\zeta}_{i,k}^1$  can be set as a random vector, such as  $\zeta_{i,l}^1 \in [-10, 10]$ , and the initial location is set as  $\boldsymbol{\chi}_{i,k}^1 = [1, 1, \dots, 1] / \sqrt{N_u - k}$ .

(c). Taking  $\boldsymbol{\zeta}_{i,k}^1$  and  $\boldsymbol{\chi}_{i,k}^1$  as starting point, PSO is implemented to search the optimal transmit weights. In each iteration, the temporary best locations will be measured and updated by the objective function in (18) and all the particles achieve the next optimal directions for the search. For example, in the  $t$ -th iteration, all particles use the given best global locations to update the velocity and location by using (17), then compare the location obtained in (17) with the last best location  $p_{id_k}^{t-1}$  and obtain the  $p_{id_k}^t$ , and then the  $p_{id_k}^t$  of each particle will be exchanged to get the global best location  $p_{gd_k}^t$ . Ultimately, all of the particles update their location by the  $p_{gd_k}^t$ . In this paper,  $\boldsymbol{\chi}_{i,k}^t$  is normalized after the update in each iteration.

(d). PSO algorithm is repeated till the maximum iteration number of  $I_{\max}^k$ .

The above operations are separately implemented at all users and the different user has the different search dimension. In other words, from the 1-st to the  $N_u$ -th user,

the search dimension becomes small, which leads to low SNR at each MS. Thus for those "bad" users, we can enlarge the swarm size or increase the iteration number to obtain the optimal weights. Certainly, the solution of the above is usually achieved by the water filling.

### C. Computational load analysis

Under the same simulation conditions we compare the computational complexity between the proposed scheme and other schemes in [10] and [14]. The comparison is limited in searching the transmit weights for all user. Generally, a SVD has a complexity with order of  $\max(p^2q, pq^2, q^3)$  in the case of  $k$ -th user, where  $p = N_T$  and  $q = k - 1$  [16]. We assume a complex multiplication equal to 4 real multiplications and 2 real additions. In the proposed scheme in this paper and the scheme in [14],  $\Omega_k = 10$  and  $I_{\max}^k = 20$  are set. If the particles denote real number, PSO algorithm has at least a real computation of  $\{\Omega_k [10 + \Phi_f] + \Omega_k - 1\} \times I_{\max}^k$ , where  $\Phi_f$  denotes the computation of objective function. The total numbers of real products and additions required for detecting one transmit symbol are derived as follows.

TABLE I. COMPUTATIONAL COMPLEXITY/SYMBOL

	$N_T = 4, N_u = 4, n_R^{(k)} = 1$	$N_T = 4, N_u = 4, n_R^{(k)} = 2$
$\Phi_L$	8126	11648
$\Phi_{OW-L}$	52400	52400
$\Phi_{OW-DPC}$	15688	16288

$\Phi_L$ : Computational complexity of linear scheme in [10].

$\Phi_{OW-L}$ : Computational complexity of linear scheme with PSO in [14].

$\Phi_{OW-DPC}$ : Computational complexity of the proposed scheme in this paper.

Table 1. shows the PSO algorithm consume the largest computational work in [14] because of continuous complex vector space. However, in this paper, we search the optimal coefficient in real vector space to greatly reduce computational load. Though the scheme in [10] has the least computational load, when the number of receive antennas at MS is increased, the quantity difference becomes small between  $\Phi_{OW-DPC}$  and  $\Phi_{OW-L}$ . Moreover, the demerits of increment of noise in [14] and residual IUI in [10] lead to loss of achievable rate to a certain degree. Taken together, the proposed scheme in this paper is feasible.

## IV. COMPUTER SIMULATION

In this Section, we present simulation results demonstrating the performance of the proposed scheme. To verify the performance of it, first, we compare the achievable sum-rate between the proposed scheme and the other schemes, such as, conventional DPC, Channel Inversion (CI) and BD, with the water filling algorithm. Then we compare the average BER performance of each user and demonstrate the feasibility of the proposed scheme.

We consider the case of  $4 \times (1, 1, 1, 1)$  in MU-MIMO, where the average achievable rate of system is determined for the total 1000 realizations of  $\mathbf{H}$ . The particle number of

$\Omega_k = 10$  and the iteration number of  $I_{\max}^k = 20$  are set for the PSO algorithm. Fig. 2 shows that the channel inversion scheme obtains a close average capacity to BD scheme by employing the water filling, but both of those two methods are inferior to the conventional DPC, which approaches the multiuser capacity more closely. The simulation results show that the proposed scheme obtains the best capacity because of the optimal transmit weights and user order with PSO. In the conventional DPC, the transmit weight for each user in the precoding matrix  $\mathbf{Q}$  is obtained by Gram-Schmidt orthogonalization without sorting.

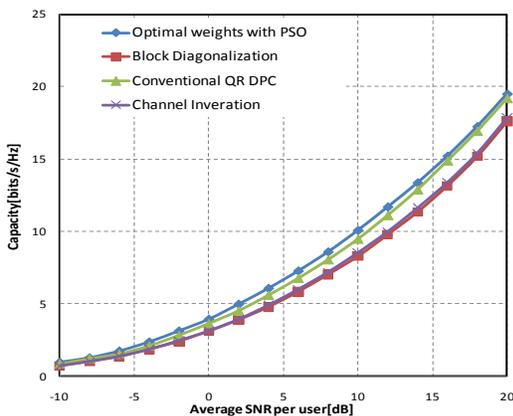


Fig. 2 Capacity comparison among the proposed PSO-DPC with optimal weights DPC and the BD, Channel inversion, Conventional DPC

Next we show the BER performance of proposed transmit scheme and compare it with the BD, channel inversion and conventional DPC schemes. Here we only consider the uniform assignment of transmit power to all users.

Fig. 3 shows the case of  $4 \times (1,1,1,1)$  in MU-MIMO downlink. The particle number of  $\Omega_k = 10$  and the iteration number of  $I_{\max}^k = 20$  are set in the PSO algorithm and QPSK is used to modulate the transmit signal. Without regard to the power distribution on each user, both BD and Channel Inversion with ZF can completely eliminate the IUI on i.i.d Rayleigh fading channels and these two methods show the same performance as SISO in i.i.d Rayleigh fading channel. As shown in Fig. 3, though the channel inversion with MMSE can not exactly cancel the IUI caused by the spatially multiplexed channel and results in some vestigial IUI, it can reduce the effect of noise enhancement between 0 and 12dB. The DPC, referred to as the interference dependent nonlinear precoding, can obtain the better performance compared with the linear BD and channel inversion schemes. In the case of conventional DPC, the QR decomposition only aims to cancel the IUI, but does not involve the optimality of transmit weights in the matrix  $\mathbf{Q}$ . Our proposed transmit scheme with optimal weights by PSO, not only suppresses the IUI completely, but also searches the optimal transmit weights achieving the SNR as large as possible at each MS. Fig. 3 shows the proposed scheme has

obtained the better BER performance by about 4 dB than the conventional DPC at  $\text{BER} = 10^{-3}$ .

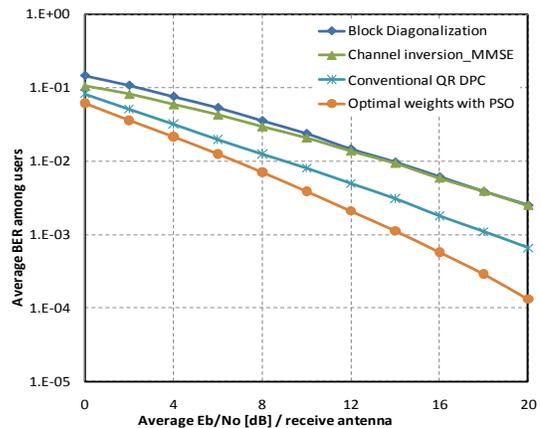


Fig. 3 Comparison of BER characteristics among proposed PSO-DPC with optimal weights, BD, Channel inversion (ZF&MMSE) and Conventional DPC.

Fig. 4 shows the BER performance of proposed scheme with multiple receive antennas and it is compared with the conventional DPC. Similarly, the values of  $\Omega_k = 10$  and  $I_{\max}^k = 20$  are set in the PSO algorithm and QPSK is employed. The conventional DPC based on QR decomposition is restricted to the users with single receive antenna, because in the case of users with multiple receive antennas, if we use the QR decomposition on  $\mathbf{H}$ , some of the users have no throughput. However, the proposed scheme can ensure the throughput for each user and improves the receive SNR by increasing the number of receive antennas. In addition, based on the channel condition and the QoS requirement of users, the PSO search algorithm works with different swarm size of  $\Omega_k$  and the iteration number of  $I_{\max}^k$  to improve the transmit gains. As shown in Fig. 4, with the increase of receive antenna number, the system performance can be greatly improved.

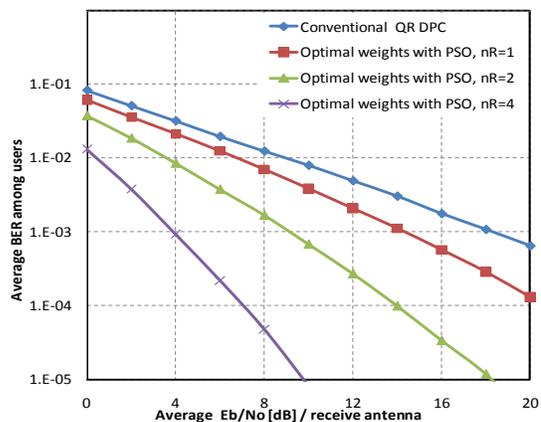


Fig. 4 Comparison of BER characteristics among the proposed PSO-DPC with optimal weights with multiple receives antennas and the conventional DPC

Fig. 5 shows the improvement of BER versus the iteration number. If the iteration number or swarm size is set too small, we cannot obtain the convergence values of transmit weight of users. Thus increasing the number of iteration is required to achieve the better performance, which cannot lead to large calculation if we control the swarm size to likely quantity. From Fig. 5 it is clearly visible that when the value of  $I_{\max}^k$  is set greater than 20, the BER performance cannot be improved any more. So the selection of value of  $I_{\max}^k$  also seems the key design element. The computational load has been discussed in Section IV, from which it is clear that the proposed scheme is feasible when the iteration number is set to  $I_{\max}^k = 20$ .

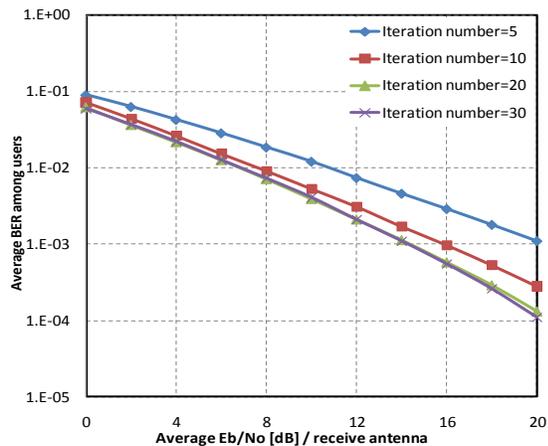


Fig. 5 BER improvement vs. iteration number

## V. CONCLUSIONS

A novel scheme with optimal transmit weights for the downlinks of MU-MIMO has been proposed, in which we employed the PSO algorithm to search the optimal transmit weights and achieved the significantly better performance than the conventional DPC, BD and the channel inversion schemes. With the CSI known at both of BS and MS, BS determine the user order and corresponding transmit weights with as large transmit gain as possible to each user. DPC has been used not only to suppress the IUI, but also maintain the system capacity. In the case of single receive antenna, because of the optimal transmit weight design with the DPC principle and PSO, the proposed scheme approached more closely to the capacity of MU-MIMO than other conventional schemes. In addition, the proposed scheme can employ the multiple receive antennas, whereas the conventional DPC based on QR decomposition is equipped with only one receive antenna at the MS. Accordingly, multiple receive antennas enables the user to obtain better BER performance, especially for the users with weak receive SNR's. For the PSO algorithm, the swarm size and iteration number determine the search accuracy and required computation time, so we can make the optimal search for

each user independently, according to the requirement of each user.

## ACKNOWLEDGMENT

This study has been supported by the Scientific Research Grant-in-aid of Japan No. 21560396, the Telecommunication Advancement Foundation (TAF), the KDDI Foundation and Sharp Corporation. The authors also acknowledge the helps by Dr. Eiji Okamoto.

## REFERENCES

- [1] Q. H. Spencer, C. B. Peel, A. L. Swindlehurst, and M. Haardt, "An introduction to the multi-user MIMO downlink," *IEEE Communications Magazine*, vol. 42, Issue 10, pp. 60-67, Oct. 2004.
- [2] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. Tel.*, vol. 10, no. 6, pp. 585-595, Nov./Dec. 1999.
- [3] D. Aftas, M. Bacha, J. Evans, and S. Hanly, "On the sum capacity of multi-user MIMO channels," in *Proceedings of Intl. Symp. on Inform. Theory and its Applications*, pp. 1013-1018, Oct. 2004.
- [4] T. Haustein, C. von Helmolt, and E. Jorswieck, "Performance of MIMO systems with channel inversion," *IEEE 55th VTC*, vol. 1, pp. 35-39, May 2002.
- [5] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461-471, Feb. 2004.
- [6] N. Jindal, and A. Goldsmith, "Dirty-paper coding versus TDMA for MIMO Broadcast channels," *IEEE Transactions on Information Theory*, vol. 29, no. 5, pp. 1783-1794, 2005.
- [7] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*, New York, pp. 127-151, Wiley, 2002.
- [8] C. Li, Y. Iwanami, and E. Okamoto "Comparative study for Tomlinson-Harashima Precoding based on MMSE criteria in Multi-user MIMO downlink system," pp. 1-5, *IEEE TENCON 2009*, Nov. 2009.
- [9] H. Nishimoto, S. Kato, Y. Ogawa, T. Ohgane, and T. Nishimura, "Imperfect block diagonalization for multiuser MIMO downlink," *IEEE PIMRC 2008*, pp. 1-5, Sept. 2008.
- [10] J. S. Kim, and S. H. Moon, "Linear Beamforming for Multi-user MIMO Downlink Systems with Channel Orthogonalization," *Global Telecommunications Conference, 2009*, pp. 1-6, *GLOBECOM 2009*.
- [11] J. Kennedy, and R. Eberhart, "Particle swarm optimization," *IEEE Int. Conf. Neural Networks*, vol. 4, pp. 1942-1948, Nov. 27 - Dec., 1995.
- [12] Y. Zhao, and J. Zheng, "Multi-user detection employing particle swarm optimization in space-time CDMA systems," *IEEE Int. Symp. Communications and Information Technology*, vol. 2, pp. 940-942, Oct. 12-14, 2005.
- [13] K. K. Soo, Y. M. Siu, W. S. Chan, L. Yang, and R. S. Chen, "Particleswarm-optimization-based multiuser detector for CDMA communications," *IEEE Trans. Vehicular Technology*, vol. 56, pp. 3006-3013, September 2007.
- [14] S. Fang, and L. H. Li, "Optimal Multi-User MIMO Linear Precoding Based on Particle Swarm Optimization," *ICC 2008*, pp. 3355-3359, May 2008.
- [15] S. M. Guru, S. K. Halgamuge, and S. Fernando, "Particle swarm optimisers for cluster formation in wireless sensor networks," *Proc. 2005 Int. Conf. Intelligent Sensors, Sensor Networks and Information Processing*, pp. 319-324, Dec. 2005.
- [16] G. H. Golub, and C. F. Van Loan, *Matrix Computation*, John Hopkins Univ. Press, pp. 253-254, 3rd edition, 1996.

# MLD-based MFSK Demodulation on MIMO Frequency Selective Fading Channel

Yasunori Iwanami

Dept. of Computer Science and Engineering  
Nagoya Institute of Technology  
Nagoya, Japan  
E-mail: iwanami@nitech.ac.jp

Kenji Nakayama

Dept. of Computer Science and Engineering  
Nagoya Institute of Technology  
Nagoya, Japan  
E-mail: nakayama@rose.nitech.ac.jp

**Abstract**— In this paper, we have proposed the novel demodulation scheme of MFSK (M-ary Frequency Shift Keying) signals on MIMO frequency selective channel. For demodulating MIMO MFSK signals, we previously used the FDE (Frequency Domain Equalization) using CP (Cyclic Prefix). In order to further improve the BER characteristics, the novel scheme using MLD (Maximum Likelihood Detection) with ISI canceller and ZP (Zero Padding) has been considered. We further reduced the receiver complexity by replacing MLD with M algorithm. Through computer simulation, we have verified that the proposed scheme using M algorithm with ISI canceller exhibits the excellent BER characteristics compared with the FDE with CP.

**Keywords**-MIMO, MFSK, ISI, IAI, MLD, M-algorithm, FDE, Multipath channel

## I. INTRODUCTION

MFSK signal has the constant envelope property and is appropriate to be amplified by nonlinear amplifier with high power efficiency. However, as the MFSK is the nonlinear modulation scheme, the equalization at the receiver side has been difficult when it is subjected to frequency selective channels. On the other side, due to the increasing demand of high data rate and reliable data transmission, MIMO (Multiple Input Multiple Output) schemes with multiple transmit and receive antennas become quite popular recently. The conventional MIMO scheme processes the received signals using the linear matrix processing. However it has been difficult to apply the linear processing to the nonlinear modulation such as MIMO MFSK, and accordingly there was almost no research on MIMO FSK. So, we aimed to develop the MIMO FSK transmission scheme.

TABLE I LIST of PARAMETERS

$N_s$ : Transmission block length
$N_G$ : Number of zero symbols
$2c$ : Number of samples per symbol
$T_s$ : Symbol duration
$\Delta t$ : Sampling interval; $\Delta t = T_s / (2c)$
$k$ : time index; present time
$(L-1)T_s$ : Maximum symbol delay time
$M$ : Number of modulation level
$n_t$ : Number of transmit antenna
$n_r$ : Number of receive antenna
$J$ : Total number of delay paths
$\Delta \tau$ : Delay path interval
$M_c$ : Parameter of M algorithm

We had already shown that the FDE (Frequency Domain Equalization) scheme using CP (Cyclic Prefix) [1] is applicable to the signal separation and equalization of MIMO MFSK signals [2]-[4], where the FDE is done before the demodulation process of MFSK signal. This method was originally developed for SISO (Single Input Single Output) FSK signals [5]. In this paper, in order to improve the BER characteristics as well as the receiver complexity of previously developed schemes [2]-[4], we consider here the demodulation of MIMO MFSK where the MLD with ISI canceller using ZP is adopted. The receiver minimizes the Euclidian distance between the receive signal and the receive signal replica of candidate transmit signal. Moreover, we have replaced the MLD by M-algorithm to further reduce the receiver complexity. We have verified that the proposed receiver structure using M-algorithm with ISI canceller exhibits the excellent BER characteristics compared with the FDE with CP.

## II. MIMO MFSK DETECTION USING MLD WITH ISI CANCELLER

We consider the detection scheme of MFSK using MLD with ISI canceller in time domain, which improves the BER performance compared with the FDE scheme using CP [2]-[4]. In Fig. 1, we show the block diagram of MIMO MFSK transmission scheme having the receiver using MLD with ISI canceller operating in time domain. At the transmitter side, as shown in Fig. 2,  $N_s$  MFSK modulated symbols with  $N_G$  ZP symbols (zero symbols) are transmitted from each transmit antenna. The ZP symbols are required to circumvent the Inter Block Interference (IBI) between the successive blocks caused by the multipath delay. The ZP is also used to reset the initial phase of continuous phase MFSK signal to be zero at the beginning of the transmit block. The reset of initial phase of continuous phase MFSK signal is necessary for generating the receive signal replica for MLD with phase synchronization at the receiver. Terminating the transmit symbol block by ZP is also useful to prevent the error propagation caused by erroneous ISI cancellation at the receiver. At the receiver side, the detection of Inphase and Quadrature (I and Q) components of MFSK symbols with the block length of  $(N_s + N_G)T_s$  is done to get the complex baseband symbols. Then the discrete time samples at the sampling rate of  $2c$  per one symbol duration  $T_s$  are taken in time domain, i.e., the sampling interval of

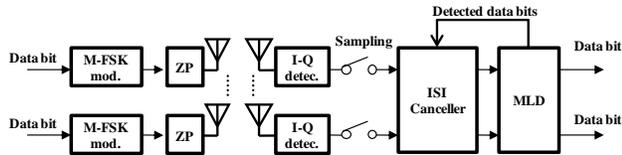


Fig. 1 Block diagram of transmit and receive system of MIMO MFSK using MLD with ISI canceller on frequency selective channel

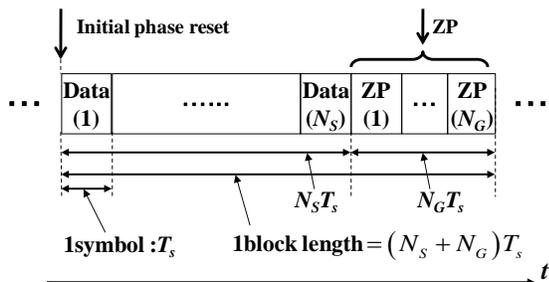


Fig. 2 Transmit signal block and Zero Padding (ZP) at the transmitter side

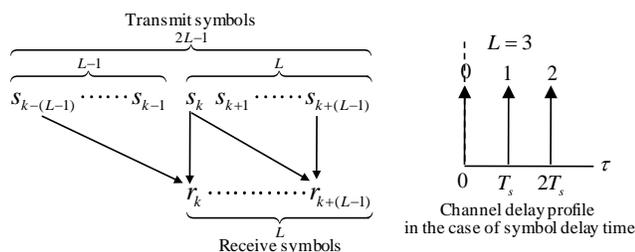


Fig. 3 Relation between transmit and receive signals on multipath delay channel in the case of SISO and symbol delay time

$\Delta t = T_s / (2c)$ . We define the present time index as  $k$  and the maximum symbol delay time due to multipath channel as  $(L-1)T_s$ . For simplicity, we also assume the SISO ( $1 \times 1$ ) transmission in the subsequent illustration. As shown in Fig. 3, to detect the symbol transmitted at time  $k$  using MLD, we need the total  $L$  receive symbols over the symbol time  $k \sim k + (L-1)$ , because the transmit signal at time  $k$  spreads over the receive symbols during  $k \sim k + (L-1)$ . As those  $L$  receive symbols contain the transmit symbol components over  $k - (L-1)$  to  $k + (L-1)$ , it is required for the MLD to search all the transmit symbol pattern over  $(2L-1)$  symbols. Accordingly, the number of searches in MLD becomes  $M^{(2L-1)}$  in the case of SISO [6].

We now consider the symbol time  $k$  in Fig. 3 as the first data symbol in Fig. 2, i.e., we let  $k=1$ . Then all the transmit symbols  $s_{k-(L-1)}, \dots, s_{k-1}$  are zero, because they are included in ZP duration. After receiving total  $L$  symbols of  $r_k, \dots, r_{k+(L-1)}$ , the first transmit symbol  $s_k$  is detected using the MLD. The next  $r_{k+L}$  is then received and the second transmit symbol  $s_{k+1}$  is also detected by using MLD. However, the ISI components caused by the first transmit symbol  $s_k$  is removed through the ISI canceller before the MLD by subtracting the hard receive replicas from the received signals. Similarly, after receiving the next receive symbol  $r_{k+L+1}$ , the third transmit symbol  $s_{k+2}$  is detected by using MLD, however the ISI

components due to  $s_k$  and  $s_{k+1}$  are cancelled by the ISI canceller before the MLD operation. Likewise, the subsequent transmit symbol  $s_{k+i}, i \geq 3 \sim N_s$  is detected using MLD, however the ISI components caused by the transmit symbols already detected are cancelled by the ISI canceller. The ISI cancellation reduces the number of searches of MLD from  $M^{(2L-1)}$  to  $M^L$ . The metric of MLD is taken as the squared Euclidian distance between the receive symbols after ISI cancellation and the receive symbol replicas corresponding to the candidate transmit symbols over  $L$  symbol duration. The transmit symbol is then detected so as to minimize the metric of MLD [6].

The proposed signal separation and equalization algorithm is illustrated in detail as follows. The channel matrix  $H$  is expressed as

$$H = \begin{bmatrix} h_0 & \dots & h_l & \dots & h_{J-1} & 0 & \dots & 0 \\ 0 & & & & & 0 & & \vdots \\ \vdots & 0 & h_0 & \dots & h_l & \dots & h_{J-1} & 0 \\ 0 & \dots & 0 & h_0 & \dots & h_l & \dots & h_{J-1} \end{bmatrix} \quad (1)$$

where  $n_T$  and  $n_R$  are the numbers of transmit and receive antennas respectively. The element  $h_l$  in (1) is also the matrix expressed as

$$h_l = \begin{bmatrix} h_{11}^{(l)} & h_{12}^{(l)} & \dots & h_{1n_r}^{(l)} \\ h_{21}^{(l)} & h_{22}^{(l)} & \dots & h_{2n_r}^{(l)} \\ \vdots & \vdots & & \vdots \\ h_{n_t 1}^{(l)} & h_{n_t 2}^{(l)} & \dots & h_{n_t n_r}^{(l)} \end{bmatrix}, \quad l = 0, \dots, J-1 \quad (2)$$

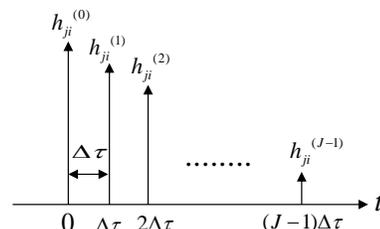
where  $h_{ji}^{(l)}$  is the complex channel gain of  $l$ -th delayed path from transmit antenna  $i$  to receive antenna  $j$  and  $J$  is the total number of delay paths.

The delay profile between transmit antenna  $i$  and receive antenna  $j$  is shown in Fig. 4, where the maximum delay time is  $(J-1)\Delta\tau$ .

The transmit symbol vector, the receive symbol vector and the receive noise vector at symbol time  $k$  are expressed as follows.

$$\left. \begin{aligned} & \text{Transmit symbol vector at symbol time } k : \\ & \mathbf{X}_k = (\mathbf{x}_{k,2c}, \dots, \mathbf{x}_{k,q}, \dots, \mathbf{x}_{k,1})^T, \mathbf{x}_{k,q} = (x_{k,q}^{(1)}, \dots, x_{k,q}^{(i)}, \dots, x_{k,q}^{(n_r)})^T \\ & \text{Receive symbol vector at symbol time } k : \\ & \mathbf{Y}_k = (\mathbf{y}_{k,2c}, \dots, \mathbf{y}_{k,q}, \dots, \mathbf{y}_{k,1})^T, \mathbf{y}_{k,q} = (y_{k,q}^{(1)}, \dots, y_{k,q}^{(j)}, \dots, y_{k,q}^{(n_r)})^T \\ & \text{Receive noise vector at symbol time } k : \\ & \mathbf{N}_k = (\mathbf{n}_{k,2c}, \dots, \mathbf{n}_{k,q}, \dots, \mathbf{n}_{k,1})^T, \mathbf{n}_{k,q} = (n_{k,q}^{(1)}, \dots, n_{k,q}^{(j)}, \dots, n_{k,q}^{(n_r)})^T \end{aligned} \right\} (3)$$

where the element vector  $\mathbf{x}_{k,q}$  in  $\mathbf{X}_k$  is the sampled


 Fig. 4 Delay profile between transmit antenna  $i$  and receive antenna  $j$  on MIMO frequency selective channel

transmit vector at time  $(q-1)\Delta t$  in symbol time  $k$  with  $q=1\sim 2c$  and  $\Delta t=T_s/(2c)$ . Accordingly, one symbol duration  $T_s$  is sampled at every small time interval  $\Delta t$ . For example,  $16 (= 2c)$  sample vectors  $\mathbf{x}_{k,16}, \dots, \mathbf{x}_{k,q}, \dots, \mathbf{x}_{k,1}$  are obtained for  $\mathbf{X}_k$ . Also  $x_{k,q}^{(i)}$  in  $\mathbf{x}_{k,q}$  is the sampled transmit symbol from transmit antenna  $i$  at time  $(q-1)\Delta t$  in symbol time  $k$ .  $y_{k,q}$  in  $\mathbf{Y}_k$ ,  $y_{k,q}^{(i)}$  in  $\mathbf{y}_{k,q}$ ,  $\mathbf{n}_{k,q}$  in  $\mathbf{N}_k$  and  $n_{k,q}^{(i)}$  in  $\mathbf{n}_{k,q}$  have the same meaning as  $\mathbf{x}_{k,q}$  in  $\mathbf{X}_k$  and  $x_{k,q}^{(i)}$  in  $\mathbf{x}_{k,q}$ .

The transmit and receive equation is expressed as

$$\begin{pmatrix} \mathbf{Y}_{k+(L-1)} \\ \vdots \\ \mathbf{Y}_k \end{pmatrix}^T = \mathbf{H} \begin{pmatrix} \mathbf{X}_{k+(L-1)} \\ \vdots \\ \mathbf{X}_k \\ \vdots \\ \mathbf{X}_{k-(L-1)} \end{pmatrix}^T + \begin{pmatrix} \mathbf{N}_{k+(L-1)} \\ \vdots \\ \mathbf{N}_k \end{pmatrix}^T \quad (4)$$

where  $L$  is  $L = \lceil (J-1)\Delta\tau/T_s \rceil + 1$  and  $\lceil x \rceil$  denotes the minimum integer number greater than or equal to  $x$ .  $(L-1)T_s$  also means the maximum symbol delay time.

We further illustrate the transmit and receive equation in detail by setting  $n_T = n_R = 2$ ,  $L = 2$  and  $J = 2c$  for simplicity of explanation.

The element  $y_{p,q}^{(1)}, y_{p,q}^{(2)}$ ,  $p = k, \dots, k+(L-1)$ ,  $q = 1, \dots, 2c$  of receive vector  $\mathbf{y}_p^{(j)}$  in (3) are represented as follows.

$$\left. \begin{aligned} y_{p,1}^{(1)} &= \left\{ h_{11}^{(0)} x_{p,1}^{(1)} + h_{11}^{(1)} x_{p-1,2c}^{(1)} + \dots + h_{11}^{(2c-1)} x_{p-1,2}^{(1)} \right\} \\ &\quad + \left\{ h_{12}^{(0)} x_{p,1}^{(2)} + h_{12}^{(1)} x_{p-1,2c}^{(2)} + \dots + h_{12}^{(2c-1)} x_{p-1,2}^{(2)} \right\} + n_{p,1}^{(1)} \\ &\quad \vdots \\ y_{p,2c}^{(1)} &= \left\{ h_{11}^{(0)} x_{p,2c}^{(1)} + h_{11}^{(1)} x_{p-1,2c-1}^{(1)} + \dots + h_{11}^{(2c-1)} x_{p-1,1}^{(1)} \right\} \\ &\quad + \left\{ h_{12}^{(0)} x_{p,2c}^{(2)} + h_{12}^{(1)} x_{p-1,2c-1}^{(2)} + \dots + h_{12}^{(2c-1)} x_{p-1,1}^{(2)} \right\} + n_{p,2c}^{(1)} \\ y_{p,1}^{(2)} &= \left\{ h_{21}^{(0)} x_{p,1}^{(1)} + h_{21}^{(1)} x_{p-1,2c}^{(1)} + \dots + h_{21}^{(2c-1)} x_{p-1,2}^{(1)} \right\} \\ &\quad + \left\{ h_{22}^{(0)} x_{p,1}^{(2)} + h_{22}^{(1)} x_{p-1,2c}^{(2)} + \dots + h_{22}^{(2c-1)} x_{p-1,2}^{(2)} \right\} + n_{p,1}^{(2)} \\ &\quad \vdots \\ y_{p,2c}^{(2)} &= \left\{ h_{12}^{(0)} x_{p,2c}^{(1)} + h_{12}^{(1)} x_{p-1,2c-1}^{(1)} + \dots + h_{12}^{(2c-1)} x_{p-1,1}^{(1)} \right\} \\ &\quad + \left\{ h_{22}^{(0)} x_{p,2c}^{(2)} + h_{22}^{(1)} x_{p-1,2c-1}^{(2)} + \dots + h_{22}^{(2c-1)} x_{p-1,1}^{(2)} \right\} + n_{p,2c}^{(2)} \end{aligned} \right\} \quad (5)$$

Here we elaborate on the elements  $x_{k,1}^{(1)}, \dots, x_{k,2c}^{(1)}$  of the transmit symbol vector  $\mathbf{x}_k^{(1)}$  from transmit antenna 1 and  $x_{k,1}^{(2)}, \dots, x_{k,2c}^{(2)}$  of the transmit symbol vector  $\mathbf{x}_k^{(2)}$  from transmit antenna 2. The channel responses of transmit symbol vectors  $\mathbf{x}_k^{(1)}$  and  $\mathbf{x}_k^{(2)}$  spread over the receive symbol vectors  $\mathbf{y}_k^{(1)}, \dots, \mathbf{y}_{k+(L-1)}^{(1)}$  and  $\mathbf{y}_k^{(2)}, \dots, \mathbf{y}_{k+(L-1)}^{(2)}$  with the maximum symbol delay time of  $(L-1)T_s$  as illustrated in Fig. 3. In order to detect the transmit symbol vectors  $\mathbf{x}_k^{(1)}$  and  $\mathbf{x}_k^{(2)}$ , the ISI components due to already detected symbols from symbol time  $k-(L-1) \sim k-1$  can be removed from the receive vectors of  $\mathbf{y}_k^{(1)}, \dots, \mathbf{y}_{k+(L-1)}^{(1)}$  and  $\mathbf{y}_k^{(2)}, \dots, \mathbf{y}_{k+(L-1)}^{(2)}$ . The ISI cancelling vector  $\tilde{\mathbf{Y}}_p$  at symbol time  $p$  is expressed as

$$\tilde{\mathbf{Y}}_p = \left( \tilde{y}_p^{(1)}, \dots, \tilde{y}_p^{(n_R)} \right)^T, \quad p = k, \dots, k+(L-1) \quad (6)$$

where  $\tilde{\mathbf{y}}_p^{(j)} = \left( \tilde{y}_{p,2c}^{(j)}, \dots, \tilde{y}_{p,1}^{(j)} \right)^T$  is the element symbol vector for cancelling the ISI at receive antenna  $j$  at symbol time  $p$ . Using the already detected symbols during  $k-(L-1) \sim k-1$  symbol period,  $\tilde{\mathbf{Y}}_p$  in (6) is derived as

$$\left( \tilde{\mathbf{Y}}_{k+(L-1)}, \dots, \tilde{\mathbf{Y}}_k \right)^T = \mathbf{H} \begin{pmatrix} \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{L \text{ symbols}} & \overbrace{\tilde{\mathbf{X}}_{k-1}, \dots, \tilde{\mathbf{X}}_{k-(L-1)}}^{(L-1) \text{ detected symbols}} \end{pmatrix}^T \quad (7)$$

where  $\tilde{\mathbf{X}}_p$ ,  $p = k-1, \dots, k-(L-1)$  are the already detected results. Next, using  $\tilde{\mathbf{Y}}_{k+(L-1)}, \dots, \tilde{\mathbf{Y}}_k$  in (7), the ISI is cancelled as follows.

$$\hat{\mathbf{Y}}_p = \mathbf{Y}_p - \tilde{\mathbf{Y}}_p, \quad p = k, \dots, k+(L-1) \quad (8)$$

where  $\hat{\mathbf{Y}}_p$  is the receive symbol vector after the ISI cancellation. Using the transmit signal  $\mathbf{X}_p$ , receive noise  $\mathbf{N}_p$  and residual ISI component  $\boldsymbol{\varepsilon}_p$ ,  $\hat{\mathbf{Y}}_p$ ,  $p = k, \dots, k+(L-1)$  in (8) is also represented as

$$\left( \hat{\mathbf{Y}}_{k+(L-1)}, \dots, \hat{\mathbf{Y}}_k \right)^T = \mathbf{H} \begin{pmatrix} \overbrace{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k}^{L \text{ symbols}} & \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{(L-1) \text{ symbols}} \end{pmatrix}^T + \begin{pmatrix} \mathbf{N}_{k+(L-1)} \\ \vdots \\ \mathbf{N}_k \end{pmatrix}^T + \begin{pmatrix} \boldsymbol{\varepsilon}_{k+(L-1)} \\ \vdots \\ \boldsymbol{\varepsilon}_k \end{pmatrix}^T \quad (9)$$

The residual ISI component  $\boldsymbol{\varepsilon}_p$  is  $\mathbf{0}$ , if the past decision results are correct. Also when the ISI cancellation is perfect and there is no noise, i.e.,  $\mathbf{N}_p = \mathbf{0}$ , it holds

$$\left( \hat{\mathbf{Y}}_{k+(L-1)}, \dots, \hat{\mathbf{Y}}_k \right)^T = \mathbf{H} \begin{pmatrix} \overbrace{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k}^{L \text{ symbols}} & \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{(L-1) \text{ symbols}} \end{pmatrix}^T \quad (10)$$

where  $\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k$  are transmit signals to be determined.

The transmit signal vector  $\mathbf{X}_k$  at the present symbol time  $k$  is determined so as to minimize the squared Euclidian distance metric as shown below.

$$\mathbf{X}_k = \arg \min_{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k} \left\| \left( \hat{\mathbf{Y}}_{k+(L-1)}, \dots, \hat{\mathbf{Y}}_k \right)^T - \mathbf{H} \begin{pmatrix} \overbrace{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k}^{L \text{ symbols}} & \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{(L-1) \text{ symbols}} \end{pmatrix}^T \right\|^2 \quad (11)$$

The above equation is the MLD criterion by which all the transmit vector patterns of  $\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k$  are searched to determine the transmit vector  $\mathbf{X}_k$  at symbol time  $k$ . In minimizing the metric in (11), all the  $L$  transmit vectors from symbol time  $k \sim k+(L-1)$  are simultaneously searched, however only the transmit vector  $\mathbf{X}_k$  at symbol time  $k$  is determined as the demodulated output.

The BER characteristics of the proposed MLD with ISI canceller and the previously proposed FDE with CP are simulated under the simulation condition in Table II. The delay profile between each transmit and receive antenna is shown in Fig. 5.

We show the BER results in Fig. 6. From Fig. 6, comparing with the FDE scheme with CP, the proposed MLD with ISI canceller achieves the gain of 13 dB at

TABLE II SIMULATION CONDITION IN FIG. 6

Modulation	$M$ -ary FSK ( $M=4$ )	
Modulation index $h$	$h = 0.7$	
Channel model between Tx and Rx antenna	Quasi-static Rayleigh fading channel with equal power 16 paths	
Interval of delay $\Delta\tau$	$\Delta\tau = T_s / 8$	
Number of antennas	SISO:1×1, MIMO:2×2	
Channel estimation	Perfect at receiver	
Signal separation and equalization	MLD with ISI canceller	FDE-CP (MMSE) with energy detection [2]-[4], [7], [8]
Number of divisions in one symbol: $2c$ ( $T_s = 2c\Delta$ )	$2c = 16$	
Transmit block length $N_s T_s$	$4T_s$ ( $N_s = 4$ )	
Length of ZP or CP	$2T_s$ (ZP)	$2T_s$ (CP for FDE)
FFT size: $2c \times N_s$	---	64

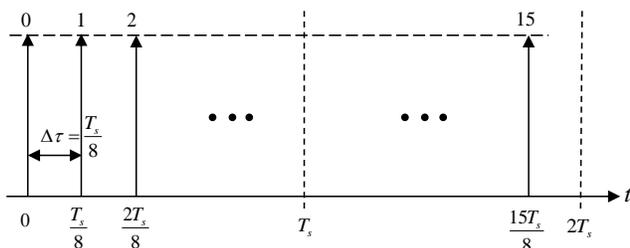
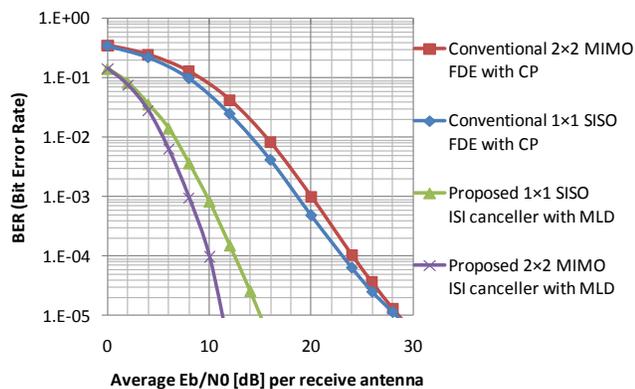


Fig. 5 Delay profile of multipath channel between each transmit and receive antenna in for Table II and Fig. 6

BER =  $10^{-5}$  for  $1 \times 1$ , and 16 dB for  $2 \times 2$ . Thus the proposed MLD with ISI canceller achieves far better equalization and signal separation capability than the previously proposed FDE scheme with CP. By removing the ISI caused by the symbols already detected, the MLD over  $L$  symbol span becomes available regardless of the block length. The long block length of  $N_s T_s$  circumvents the degradation of transmission rate due to the insertion of ZP into the block.


 Fig. 6 BER characteristics of 4FSK ( $h=0.7$ ) on MIMO quasi-static multipath channel

### III. MIMO MFSK RECEIVER USING M ALGORITHM WITH ISI CANCELLER

In the receiver structure proposed in section II, the ISI canceller reduces the amount of calculation in MLD. However, the complexity of MLD is still high when the number of transmit antenna  $n_t$  and the maximum symbol delay time of  $(L-1)T_s$  become large. This is because when  $(L-1)T_s$  is large, the symbol vector  $\mathbf{X}_k$  detected at the symbol time  $k$  influences  $L$  future symbols, thus the MLD window size  $L$  over the transmit symbol vectors  $\mathbf{X}_k, \dots, \mathbf{X}_{k+(L-1)}$  becomes large. In this section, we replace the MLD by the M algorithm to further reduce the amount of calculation in MLD. The M algorithm is the representative of breadth first search algorithms for the tree structure. Based on the MLD criterion in (11), the MLD metric is defined as

$$\left\| \left( \overbrace{\hat{\mathbf{Y}}_{k+(L-1)}, \dots, \hat{\mathbf{Y}}_k}^{L \text{ symbols}} - \mathbf{H} \left( \overbrace{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k}^{L \text{ symbols}}, \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{(L-1) \text{ symbols}} \right)^T \right\|^2 \quad (12)$$

Now, we define the receive replica vector  $\hat{\mathbf{Y}}'_p$ ,  $p = k, \dots, k+(L-1)$ , as

$$\left( \overbrace{\hat{\mathbf{Y}}'_{k+(L-1)}, \dots, \hat{\mathbf{Y}}'_p, \dots, \hat{\mathbf{Y}}'_k}^{L \text{ symbols}} \right)^T = \mathbf{H} \left( \overbrace{\mathbf{X}_{k+(L-1)}, \dots, \mathbf{X}_k}^{L \text{ symbols}}, \overbrace{\mathbf{0}, \dots, \mathbf{0}}^{(L-1) \text{ symbols}} \right)^T \quad (13)$$

where  $\hat{\mathbf{Y}}'_p$  means the receive replica vector when assuming there is no ISI from the past symbol vectors transmitted. Using  $\hat{\mathbf{Y}}'_p$  in (13), we can simplify eq.(12) as follows.

$$\sum_{p=k}^{k+(L-1)} \left\| \hat{\mathbf{Y}}_p - \hat{\mathbf{Y}}'_p \right\|^2 \quad (14)$$

Then we define the cumulative metric from the symbol time  $k$  to  $i$  ( $i=1, \dots, L$ ) as

$$\sum_{p=k}^{k+(i-1)} \left\| \hat{\mathbf{Y}}_p - \hat{\mathbf{Y}}'_p \right\|^2 \quad (15)$$

The cumulative metric in (15) at  $i=L$  coincides with the MLD metric in (12). The above cumulative metric can be used as the likelihood criterion to determine the candidates of transmit symbol vectors up to  $i$ , i.e.,  $\mathbf{X}_{k+(i-1)}, \dots, \mathbf{X}_k$ .

Using this cumulative metric, the transmit symbol vectors at each step  $i$  are selected from the total  $M_c$  (parameter of M algorithm) candidates. The parameter  $i$  is increased at each step. The total number of searches in M algorithm becomes  $\{1+(L-1)M_c\} \times M^{n_t}$ , where  $M$  is the number of modulation levels. On the other hand, the total number of searches in MLD in section II is  $M^{n_t L}$ . By using M algorithm, the total number of searches are reduced greatly, because the parameter of  $L$  is not included in the exponent.

To illustrate the M algorithm in the proposed scheme in detail, we show the case of  $n_t = n_r = 2$ ,  $L = 3$ ,  $M = 2$

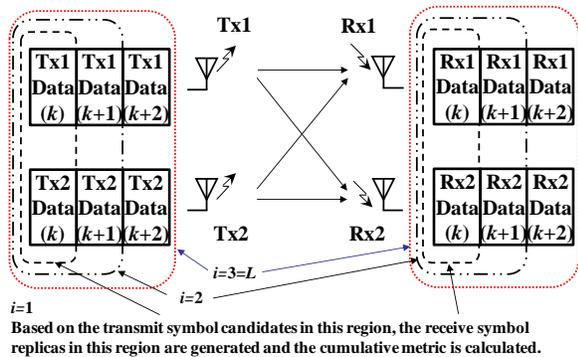


Fig. 7 Example of transmit and receive blocks ( $n_r = 2$ ,  $L = 3$ ,  $M = 2$ ) in M algorithm

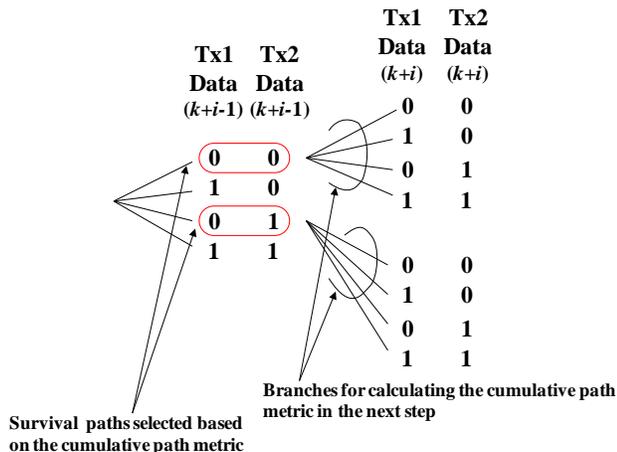


Fig. 8 Example of path selection in M algorithm ( $n_r = 2$ ,  $L = 3$ ,  $M = 2$ ,  $M_c = 2$ )

and  $M_c = 2$  in Fig. 7 and Fig. 8, as an example for simplicity.

- <Step1> Calculation of cumulative path metric at  $i = 1$  (Fig. 7)
- <Step2> Based on the cumulative path metric at  $i = 1$ ,  $M_c (= 2)$  survival paths in the tree are selected. (Fig. 8)
- <Step3> By increasing  $i$  one by one and repeating the step 1 and 2, the transmit data symbols with the survival path having the minimum cumulative path metric are determined at  $i = L$ .

Using the above mentioned M algorithm, the BER characteristics of proposed detection scheme, i.e., M algorithm with ISI canceller, is simulated with the simulation condition listed in Table III and is shown in Fig. 9. The delay profile between each transmit and receive antenna is the same as in Fig. 5.

From Fig. 9, the BER characteristics near MLD are obtained with the value  $M_c = 16$  with less amount of calculation. Accordingly, we can say that the BER

Table III SIMULATION CONDITION IN FIG. 9

Modulation	M-ary FSK ( $M=4$ )
Modulation index: $h$	$h = 0.7$
Channel model between Tx and Rx antenna	Quasi-static Rayleigh fading channel with equal power 16 paths
Interval of delay $\Delta\tau$	$\Delta\tau = T_s / 8$
Number of antennas	MIMO $2 \times 2$
Channel estimation	Perfect at receiver
Signal separation and equalization	M-algorithm/MLD with ISI Canceller
Number of divisions in one symbol: $2c$ ( $T_s = 2c\Delta t$ )	$2c = 16$
Transmit block length $N_s T_s$	$4T_s$ ( $N_s = 4$ )
Length of ZP	$(L-1)T_s = 2T_s$

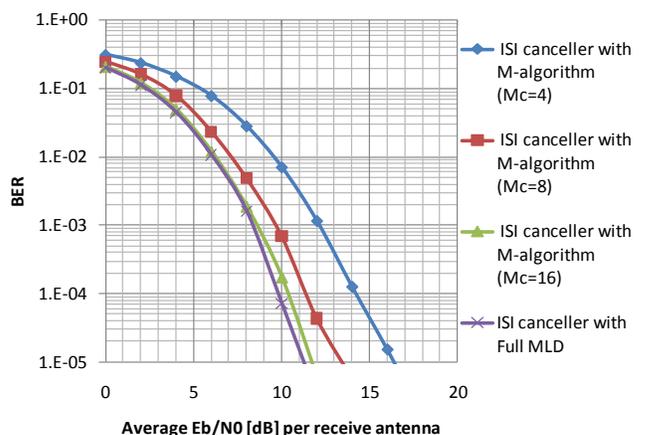


Fig. 9 BER characteristics of 4FSK ( $h=0.7$ ) on MIMO  $2 \times 2$  multipath channel

characteristics approaching the MLD are achievable by selecting the appropriate value of  $M_c$ . The amount of calculation of MLD is proportional to  $M^{n_r L}$ , while the one of M algorithm is  $M^{n_r}$ . Thus, we can cope with the channel with large symbol delay time of  $(L-1)T_s$  and also can handle the larger modulation level of  $M$ .

#### IV. EVALUATION ON AMOUNT OF CALCULATION IN MIMO FSK RECEIVERS

The amount of calculation in MIMO MFSK receiver is estimated through the number of complex additions and multiplications for demodulating one transmit symbol. In Table IV, the numbers of complex additions and multiplications per one symbol for MIMO FSK detection schemes are shown, where we set  $n_r = n_R$  and the approximation of  $N_s \gg L$  is assumed with the notification of

$$A = 2c(2 \cdot J \cdot n_r + 3), \quad B = 2c(2 \cdot J \cdot n_r)L \quad (16)$$

Under the condition in Table IV, the comparison of amount of calculation is evaluated and is shown in Fig. 10. When comparing the MLD with ISI canceller to the M algorithm with ISI canceller in Fig. 10, the amount of calculation of M algorithm with ISI canceller is far less than the MLD with

ISI canceller. Although the FDE with CP is superior to other two schemes in complexity, the BER characteristic is inferior to the other two. Considering the moderate amount of calculation at the receiver and the excellent BER characteristics, the proposed M algorithm with ISI canceller will be considered as the best choice.

TABLE IV NUMBER OF COMPLEX ADDITIONS AND MULTIPLICATIONS PER ONE SYMBOL TO BE DETECTED

FDE with CP	$2c \left[ 10 \log_2 (2c \cdot N_s) \right] + 2n_r^2 + 2n_r - 13$
MLD with ISI canceller	$ALM^{n_r \times L} + B$
ISI canceller with M-algorithm	$A \cdot M^{n_r} \left\{ 1 + M_c \cdot \frac{1}{2} (L^2 + L - 2) \right\} + B$

TABLE V CONDITIONS FOR EVALUATING THE AMOUNT OF CALCULATION IN FIG.10

Modulation	M-ary FSK (M= 4)
Modulation index: $h$	$h = 0.7$
Channel model between Tx and Rx antenna	Quasi-static Rayleigh fading channel with equal power 16 paths
Interval of delay $\Delta\tau$	$\Delta\tau = T_s / 8$
Number of antennas	MIMO 2x2
Channel estimation	Perfect at receiver
Number of divisions in one symbol: $2c$ ( $T_s = 2c\Delta t$ )	$2c = 16$
Length of ZP or CP	$2T_s$
Transmit block length $N_s T_s$	$N_s = 8$

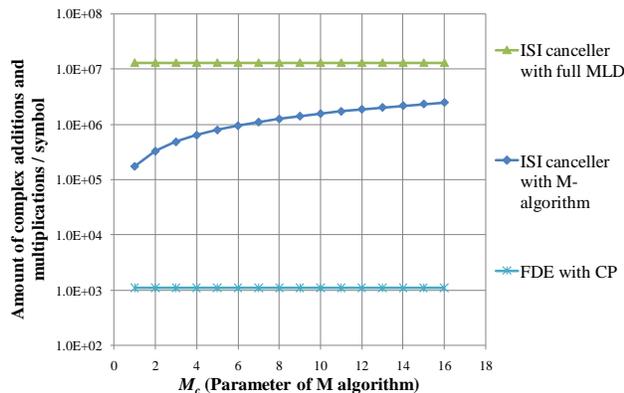


Fig. 10 Amount of calculation at the receiver (MIMO2x2,  $\Delta\tau = T_s / 8$ )

## V. CONCLUSIONS

We have proposed the MFSK receiver structure using M algorithm with ISI canceller on MIMO frequency selective channels. Our previously proposed receiver structure utilized the FDE with CP for signal separation and equalization before the demodulation of MFSK signal followed by energy detectors, however the BER characteristic of FDE scheme with CP was a little bit poor.

To improve the BER characteristics of FDE with CP, we have investigated the receiver structure using MLD with ISI canceller which has the excellent BER characteristics. Although the MLD with ISI canceller exhibits the excellent BER characteristics, the amount of calculation at the receiver, which increases exponentially with the number of transmit antenna  $n_r$  and the MLD window size  $L$ , becomes large and is difficult actually to be implemented. In order to solve the complexity problem, the MLD is replaced by M algorithm. By selecting the appropriate value of M algorithm parameter  $M_c$ , the BER characteristic approaches the lowest value achievable by MLD. In the proposed M algorithm with ISI canceller, the receiver complexity increases exponentially only with the number of transmit antenna, thus we can reduce the receiver complexity. Accordingly the proposed receiver can cope with large delay time of multipath channels. We have developed the MIMO FSK receiver with low BER and modest complexity, which will be effective to increase the transmission rate of conventional SISO FSK channels. As a future study, the measurement of channel state information at the receiver will be needed to assess the actual BER degradation for practical applications.

## ACKNOWLEDGEMENT

This study is partially supported by the Grants-in-Aid for Scientific Research 21560396 of the Japan Society for the Promotion of Science, the Telecommunications Advancement Foundation and the Sharp Corporation. The authors also acknowledge the helps of Dr. Eiji Okamoto and Mrs. Kazumi Ueda.

## REFERENCES

- [1] David Falconer, Sirikiat Lek Ariyavisitakul, Anader Benyamin-Seeyar, and Brian Eidson, "Frequency Domain Equalization for Single-Carrier Broadband Wireless Systems," *IEEE Commun. Mag.*, pp.58-66, April 2002.
- [2] Manabu Hijimoto, Yasunori Iwanami, and Eiji Okamoto, "A study on M-ary FSK with non-coherent detection on MIMO frequency selective channels," *Technical report of IEICE*, RCS2008-230, pp.107-112, Mar. 2009 (in Japanese).
- [3] Kenji Nakayama, Yasunori Iwanami, and Eiji Okamoto, "A comparative study of MIMO-MFSK with frequency and energy detection under frequency-selective fading channels", *IEICE Society Conference*, B-5-34, Sep. 2009 (in Japanese).
- [4] Kenji Nakayama, Yasunori Iwanami, and Eiji Okamoto, "MIMO MFSK receivers using FDE and MLD on quasi-static frequency selective fading channels," *International Symposium on Information Theory and its Applications 2010 (ISITA2010)*, Taichung Taiwan, ES2-Mo-1, 6 pages, Oct. 2010.
- [5] Masahiro Maki and Yoshihiko Akaiwa, "An adaptive equalizer for FSK frequency detection," *IEICE Technical Report*, RCS 93-54, pp.23-28, September, 1993 (in Japanese).
- [6] Satoshi Nakanobu, Yasunori Iwanami, and Eiji Okamoto, "A Comparative Study on Time Domain Sequential Equalizer with MLD and MLSE Equalizer on MIMO Frequency Selective Channels," *IEICE Trans. on Commun.*, vol.E93-B, No.11, pp.3197-3202, Nov. 2010.
- [7] Yasunori Iwanami and Paul H. Wittke, "BER analysis of binary CP-FSK with energy detection and sequence estimation on frequency selective multipath channels", *International Conference on Communications (ICC'99)*, Vancouver, S21.3, vol.2, pp.833-837, June 1999.
- [8] Yasunori Iwanami and Paul H. Wittke, "Error Performance Analysis of an Energy Sequence Estimation Receiver for Binary FSK on Frequency-Selective Fading Channels," *IEEE Trans. on Wireless Communications*, Vol. 2, No. 2, pp. 260-269, March 2003.

# The Impact of Antenna Directivity and Channel Bandwidth on the Power Spectral Density of Wideband and UWB MISO Channels

Ana-Maria Pistea  
Communications  
TUCN  
Cluj Napoca, Romania  
ampistea@gmail.com

Tudor Palade  
Communications  
TUCN  
Cluj Napoca, Romania  
palade@com.utcluj.ro

Ancuta Moldovan  
Communications  
TUCN  
Cluj Napoca, Romania  
Ancuta.Moldovan@com.utcluj.ro

Hamidreza Saligheh Rad  
Department of Radiology  
UPENN  
Philadelphia, USA  
hamid.saligheh@gmail.com

**Abstract**—The combination of MIMO with wideband (WB) and ultra-wideband (UWB) systems imply many new effects, which make the conventional channel models developed for narrowband transmissions to be inadequate for communications with large bandwidth. In this paper we use the Fourier analysis of the cross-correlation functions (CCF)s between the space-time-frequency (STF) transfer functions of two sub-channels of an outdoor WB/UWB MIMO wireless channel to derive the power spectral density (PSD) in a stationary scenario. The assumptions of the stationary scenario transform the MIMO channel into a Multiple-In Single-Out (MISO) channel and the PSD is determined when the channel bandwidth is wide or ultra-wide. The Fourier analysis of the CCF reveals the fact that the PSD deviates from the U-shaped function, i.e., Clarke/Jake's model to a great extent. This deviation is strongly influenced by the antenna directivity and the channel bandwidth. One major phenomenon which become stronger when the bandwidth increases is the frequency selectivity which is obviously larger in case of UWB channels than in case of WB channels. This characteristic makes possible the recognition of the type of the channel for its bandwidth from its PSD only.

*Keywords*-WB; UWB; MIMO; wireless channel; PSD.

## I. INTRODUCTION

The design of high performance wideband (WB) and ultra-wideband (UWB) multiple-input multiple-output (MIMO) wireless systems, requires accurate prediction of the impact of random multipath propagation and reliable MIMO channel models which take into account the impact of the direction-of-departure (DOD), direction-of arrival (DOA), time of arrival (TOA) and the system bandwidth [2], [3].

The cross-correlation function (CCF) of two sub-channels of an outdoor WB/UWB MIMO wireless channel [1] is used to determine the Power Spectral Density (PSD) of WB and UWB channels in a stationary scenario. The mathematical set-up of the stationary scenario transforms the MIMO channel into a Multiple-In Single-Out (MISO) channel. We analyzed a stationary MISO channel as a special case of MIMO channel. The PSD is derived for outdoor channels, when omnidirectional and directional antennas are employed at the mobile station (MS), in a 2D non-isotropic propagation environment. The CCF expression is based on the space-time-frequency (STF)

channel transfer functions (CTF) which is represented by a sum of cluster waves over a number of dominant paths. We derive the PSD using the Fourier analysis of the CCF of MIMO channels. In the literature are other works which present results obtained by using the classical approach that applies the direct Fourier relation between the correlation and the PSD [2], [3]. The majority of these models employ a specific geometry for the scatterers around the MS [4], [5], [6]. Based on this approach each model is just capable to predict the behavior of that particular propagation scenario [8]. Moreover, they are not able to investigate the spatial, the temporal, and the frequency aspects of the wireless channel in one single model. Alternatively in this paper, we do not assume a certain geometry to describe the relative distribution of scatterers in the space. We establish a mathematical relation between the random time-delay and the random channel-gain associated with each scattered waveform within each cluster and use appropriate pdfs for the parameters such as the time-delay, the DOA and the DOD. As a result the derived PSD is based on a non-geometry approach, describing the non-isotropic propagation by using appropriate pdfs for the parameters such as the time-delay, DOA and DOD. The expression of the PSD is the Fourier transform of a linear series expansion of a number of Bessel functions of the first kind. The coefficients of the expansion of the PSD are described by three categories of Fourier series coefficients (FSC): a) linear convolution of the FSC of the antenna pattern, b) the FSC of the azimuth angular spread and c) the Fourier series expansion (FSE) of the pdfs describing the non-isotropic environment.

This paper is organized as follows: the 2D channel model is described in Section 2. In Section 3, the mathematical model of the CCF it is presented. In Section 4, the PSD equation is derived and it is numerically evaluated under different circumstances. Conclusions are presented in Section 5.

## II. TWO DIMENSIONAL WB/UWB MIMO MODEL DESCRIPTION

In this section, we describe the propagation scenario and the notations used throughout this paper in which superscripts B

and  $M$  indicate the variables at BS and MS sides, respectively. We also emphasize the differences which exist between WB and UWB channel models and how these differences are included in our model.

We consider a moving MS with a constant speed vector  $V(\frac{m}{sec})$  and a fixed BS in a 2D non-isotropic propagation environment where the multipath components arrive in clusters. The resulted CTF determined by the  $p$ th transmitting antenna element at BS side, the propagation environment and the  $m$ th receiving antenna element at MS side is the summation of the dominant  $I$  paths and  $L$  clusters. CTF expression includes the following elements:

**E1)** The antenna propagation patterns (APP)s of the  $p$ th and  $m$ th antenna of the BS and MS array,  $G_p^B(\Theta_{il}, \omega)$ ,  $G_m^M(\Theta_{il}, \omega)$  give the response of the antenna in terms of the propagation directions and frequency. These functions implicitly include the effect of mutual coupling caused by the neighboring antenna elements. Are all periodic functions of  $\Theta_{il}^B$ - DOD of the  $i$ th dominant path in the  $l$ th cluster from the BS; and  $\Theta_{il}^M$ - DOA of the  $i$ th dominant path in the  $l$ th cluster to the MS, with the same period of  $2\pi$ . Therefore, we represent them by their FSEs as follows:

$$G(\Theta; \omega) = \sum_{k=-\infty}^{\infty} \mathcal{G}_k e^{jk\Theta}, \quad \mathcal{G}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} G(\Theta; \omega) e^{-jk\Theta} d\Theta \quad (1)$$

For the WB channel, it is assumed that the response of the antenna does not change significantly over the bandwidth since the relative bandwidth is a small fraction of the central frequency. UWB antenna patterns are different at different frequencies and this characteristic should be considered in our model. In fact this is the main difference between channel models for WB signals and channel models for UWB signals. Thus, depending on the signal bandwidth, we will have two approaches for APP calculation:

- for WB signals, APP is calculated depending on the central frequency,  $\omega$ ;
- for UWB signals, we calculate the APP depending on the central frequency and by integrating  $G(\Theta, \omega)$  across all the frequencies of the transmitted signal.

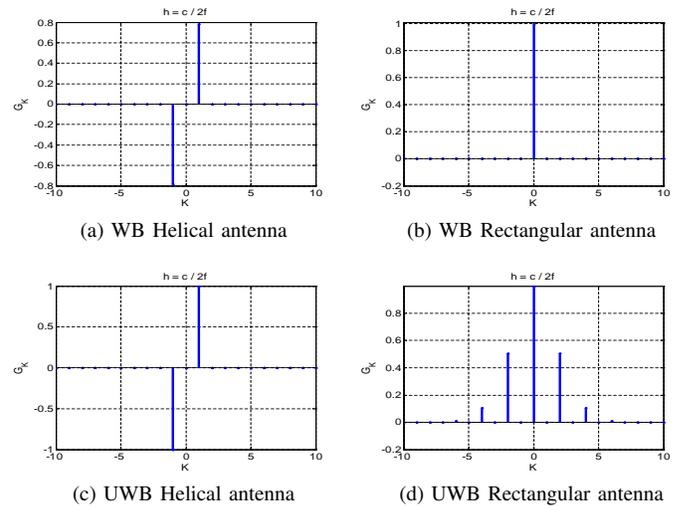
Table I presents the APPs of two WB/UWB antennas. The helical (directional) and rectangular (omnidirectional) antennas are often used for antenna arrays and WB/UWB applications [7]. The Fourier coefficients of omnidirectional antenna can be simply given by  $\mathcal{G}_k = \delta_k$ , where  $\delta_k$  is the unit impulse [8]. In this case, the corresponding coefficients,  $\mathcal{G}_k$ , vanish from the expression of the CCF and the channel bandwidth influence can not be determined. Since in the case of WB and UWB channels, the bandwidth has a great impact on the channel statistics, it's necessary to replace  $\mathcal{G}_k = \delta_k$  with the  $\mathcal{G}_k$  of rectangular antennas which allow us to see the influence of the channel bandwidth.

Table I: 2D Antenna propagation patterns

Antenna Type	APP, $G_W(\Theta, \omega), \forall \Theta \in [-\pi, \pi)$	
Helical antenna	WB	$j G_0 \cdot \frac{\omega}{2c} \cdot h \cdot \sin\Theta$
	UWB	$\frac{1}{(f_H - f_L)} \int_{f_L}^{f_H} j G_0 \cdot \frac{\omega}{2c} \cdot h \cdot \sin\Theta d\omega$
Rectangular antenna	WB	$j G_0 \frac{\sin(\frac{\omega}{2c} \cdot h \cdot \sin\Theta)}{\frac{\omega}{2c} \cdot h \cdot \sin\Theta}$
	UWB	$\frac{1}{(f_H - f_L)} \int_{f_L}^{f_H} j G_0 \frac{\sin(\frac{\omega}{2c} \cdot h \cdot \sin\Theta)}{\frac{\omega}{2c} \cdot h \cdot \sin\Theta} d\omega$

$\omega$  is the channel center frequency,  $f_H, f_L$  are the upper and lower frequencies of the UWB channel bandwidth, the parameter  $h$  is proportional with the size of the antenna and  $G_0$  is the real and positive constant antenna gain that varies for each antenna.

Figure 1 shows the FSCs of the propagation patterns for WB and UWB antennas when  $h = \frac{c}{2f}$ . In the case of WB channels the antenna size is considered at the frequency  $f = \frac{\omega}{2\pi} = 2.4$  GHz and for UWB channels we calculated the antenna size at  $f = \frac{\omega}{2\pi} = 10.6$  GHz. We observe that for these antennas, the value of  $\mathcal{G}_k$  is considerable only for a limited number of coefficients.


 Figure 1: Normalized Fourier Series Coefficients for WB-APPs and UWB-APPs,  $\frac{\mathcal{G}_k}{\max_i |\mathcal{G}_i|}$ , with antenna size:  $h = \frac{c}{2f}$ .

**E2)** The PDF of the propagation directions,  $f^B(\Theta^B)$  and  $f^M(\Theta^M)$ , characterizes the non-isotropic environment. One candidate for the pdf of the non-isotropic AAS called Laplace distribution is presented in [9]. Another distribution which characterizes the non-isotropic environment is the von-Mises pdf[5]. Since these pdfs are periodic functions with period  $2\pi$ , in Table II we represent them by the Fourier series coefficients (FSCs):

$$f_\Theta(\Theta) = \sum_{k=-\infty}^{+\infty} \mathcal{F}_k e^{jk\Theta}, \quad \mathcal{F}_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_\Theta(\Theta) e^{-jk\Theta} d\Theta \quad (2)$$

The von-Mises pdf is strongly influenced by the parameter  $n$  which determine the order of the channel non-isotropy. In other

words  $n$  controls the width of DOA of scatter components. The values of  $n$  can be chosen between  $n \in [0, \infty)$ . When  $n = 0$ ,  $f_{\Theta M}(\Theta) = \frac{1}{2\pi}$  this is equivalent with isotropic scattering.

Table II: Non-isotropic AAS and corresponding Fourier series coefficients

PDF, $f_{\Theta}(\Theta)$ , $\forall \Theta \in [-\pi, \pi)$ , $\mathcal{F}_k$	
Laplace	$f_{\Theta}(\Theta) = \frac{e^{-\frac{ \sqrt{2}\Theta }{\sigma}}}{\sqrt{2}\sigma}$
	$\mathcal{F}_k = \frac{e^{-\frac{\pi(\sqrt{2}+jk)\sigma}{2\pi(-2+j\sqrt{2}k\sigma)}} \left( e^{\frac{2\sqrt{2}\pi}{\sigma}} - e^{j2k\pi} \right)}{2\pi(-2+j\sqrt{2}k\sigma)}$
von-Mises	$f_{\Theta}(\Theta) = \frac{e^{- n*\cos(\Theta-\mu) }}{2\pi J_0(n)}$
	$\mathcal{F}_k = \frac{J_k(n)}{J_0(n)}$

When  $n \rightarrow \infty$ ,  $f_{\Theta M}(\Theta) = \delta(\Theta - \mu)$ , the propagation environment is considered extreme non-isotropic scattering concentrated at  $\Theta = \mu$ , where  $\mu \in [-\pi, \pi)$  is the mean DOA at the MS. For large  $n$ , say  $n \geq 3$  we have a typical non-isotropic environment [10]. When FSCs are determined the parameter  $n$  appears as the argument of the Bessel functions, where  $J_k(\cdot)$  is the modified Bessel function of the first kind and  $J_0(\cdot)$  is the zero-order modified Bessel function.

Figure 2 compares the FSCs of Laplace and von Mises pdf (at the MS). For the von-Mises distribution, FSCs are presented when the propagation environment has two different orders of non-isotropy.

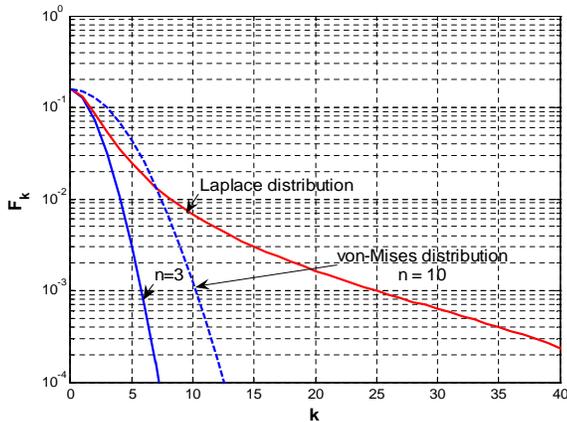


Figure 2: Fourier series coefficients for different AAS pdfs, to approximate Laplace and von-Mises distributions (with different orders of non-isotropy  $n = 3, 10$ ) determined to have a good match to the real pdf for the non-isotropic propagation

Comparing the distributions in Figure 2, we see that the necessary number of FSCs for the Laplace pdf is larger than the necessary number of FSCs for the von-Mises pdf. One can observe that the required number of the FSCs increases when the non-isotropic characteristic of the propagation environment becomes more pronounced.

**E3)** The most significant effect for WB/UWB channels is the different attenuation that the sub-bands undergo. This

phenomenon is known as frequency selectivity.

In our model the frequency selectivity of the radio channel is characterized by the term  $(\frac{\omega_{bw}}{\omega})^\eta$ , where  $\omega_{bw}$  is the signal bandwidth  $\omega_{bw} = \omega_H - \omega_L$ ,  $\omega_L$ , and  $\omega_H$  are the lower and the upper frequencies,  $\omega$  is the central frequency.  $\eta$  depends on the geometric configuration of the objects which produce signal's diffraction. Depending on the geometric configuration of the objects,  $\eta$  can take the values: -1 (diffraction by corner or tip), 0.5 (diffraction by axial cylinder face), 1 (diffraction by broadside of a cylinder).

**E4)** The  $i^{th}$  path (of the  $l$ th cluster) propagation delay,  $\tau_{pm,il}$  is decomposed into three components [8]: one component which represents the delay depending on the distances between BS and MS, and another two components which vary as a function of the local coordinates of BS and MS:

$$\tau_{pm,il} = \tau_{il} - (\tau_{p,il}^B + \tau_{m,il}^M) \quad (3)$$

$$\tau_{p,il}^B \triangleq \frac{a_p^B \Theta_{il}^B}{c}; \quad \tau_{m,il}^M \triangleq \frac{a_m^M \Theta_{il}^M}{c} \quad (4)$$

where  $\tau_{i,l}$  is the delay between local coordinates  $O^B$  or  $O^M$ ,  $\tau_{p,il}^B$ ,  $\tau_{m,il}^M$  represents the propagation delays from antenna  $a_p^B$  to  $a_m^M$  located in their corresponding coordinates  $O^B$  or  $O^M$ .  $T_l$  is the cluster arrival rate and is considered to be constant in time.  $\Theta_{il}^B$  is the unity vector pointing to the DOD of the  $i \times l$ th dominant path from the BS and  $\Theta_{il}^M$  is the unity vector pointing to the DOA of the  $i \times l$ th dominant path to the MS.

**E5)** When modeling narrowband channels, it was adequate to define the path gain depending on the path time-delay [8]. This is not sufficient for WB/UWB MIMO channels where the frequency selectivity phenomenon influences the gain of the channel. In the ray based propagation models which can be applied to signals transmitted at high frequencies range, like WB and UWB signals, can be assumed that one propagation path has DOA and TOA that does not depend on frequency, but has a frequency dependent complex path gain. In our model the multipath gain is expressed as the extension of Friis' Transmission Formula [11]:  $g_{mp,il} = \frac{1}{2\omega\tau_{il}}$ .

**E6)** Assuming plane wave propagation the path phase shift,  $\psi_{mp,il}$ , can be accurately approximated by  $\psi_{il}$ . This is absolutely a function of the distance between the BS and the MS and on the signal frequency, and not a function of the position of different antenna elements in their coordinates [8]. The term  $\psi_{il} = \phi_{il} - \omega(\tau_{il} + T_l)$  characterizes the phase shift depending only on the distance between the BS and the MS and the signal frequency, and not a function of the position of different antenna elements in their coordinates [12].  $\psi_{il} \sim U[-\pi, \pi)$  and illustrates the phase contribution of surrounding scatterers. We express the path phase shift in the form  $e^{j(\phi_{il} - \omega(\tau_{il} + T_l))}$ .

When putting together the elements described in E1÷E6, the CTF has the following expression:

$$h_{pm}(t, \omega) = \left(\frac{\omega_{bw}}{\omega}\right)^\eta \sum_{l=1}^L \sum_{i=1}^I G_p^B(\Theta_{il}^B; \omega) G_m^M(\Theta_{il}^M; \omega) \times \\ \times g_{pm,il} e^{j(\phi_{il} - \omega(\tau_{il} + T_l) - \bar{\omega}_{il}t - \omega\tau_{il} - \omega T_l)} \quad (5)$$

In the resulted CTF each  $l$  cluster, and implicitly each  $i$  wave, is associated with a path attenuation gain,  $g_{pm,il}$ , a path phase shift,  $\psi_{il}$ , a time-varying delay,  $\tau_{pm,il}$ . The Doppler shift of the  $i$ th received wave, within the  $l$ th cluster, is represented by  $\overline{\omega_{il}} = \frac{\omega}{c} V^T \Theta_{il}^M$  where  $V$  and  $c$  are the MS velocity vector and the speed of light, respectively.

### III. TWO DIMENSIONAL CROSS-CORRELATION FUNCTION OF WB/UWB MIMO CHANNELS

The CCF expression of the TFs,  $h_{pm}(t_1, \omega_1)$  and  $h_{qn}(t_2, \omega_2)$ , of two arbitrary sub-channels of a MIMO channel is the result of the following definition:

$$R_{pm,qn}(t_1, t_2, \omega_1, \omega_2) \triangleq E [h_{pm}(t_1, \omega_1) h_{qn}^*(t_2, \omega_2)] \quad (6)$$

In the CCF expression are three dimensions: space, two pairs of antenna elements ( $p, m, q, n$ ), time ( $t_1, t_2$ ), and central frequencies ( $\omega_1, \omega_2$ ). According to these three dimensions, we call it STF-CCF. The expectation operation is performed over all introduced random variables. In the presence of enough number of multi-paths by invoking the central limit theorem the TF can be considered a Gaussian random process. Therefore, the above second-order statistics fully characterize statistical behavior of the channel.

By replacing (5) in (6), regrouping dependent and independent random variables, using the elements described in E1-E6 and the results presented in [1] the CCF results in the following expression:

$$\begin{aligned} R_{pm,qn}(t_1, t_2, \omega_1, \omega_2) &= \frac{(\omega_{bw1}\omega_{bw2})^\eta}{(\omega_1\omega_2)^\eta (4\omega_1\omega_2)} \\ &\times \Phi_\tau(j(\omega_1 - \omega_2)) \Phi_\tau^{-1}(j(\omega_1 - \omega_2)) \Phi_T(j(\omega_2 - \omega_1)) \\ &\times \{ \mathcal{W}(d_{p,q}^B, \mathcal{G}_{p,k}^B(\omega_1) \otimes \mathcal{G}_{q,-k}^{B*}(\omega_2) \otimes \mathcal{F}_k^B) \\ &\times \mathcal{W}(d_{m,n}^M, \mathcal{G}_{m,k}^M(\omega_1) \otimes \mathcal{G}_{n,-k}^{M*}(\omega_2) \otimes \mathcal{F}_k^M) \} \quad (7) \end{aligned}$$

where

$$\mathcal{W}(d, \mathcal{H}_k) \triangleq 2\pi \sum_{k=-\infty}^{\infty} j^k e^{jk\angle d} \mathcal{H}_k(\omega) J_k\left(\frac{|d|}{c}\right) \quad (8)$$

$$\begin{aligned} d_p^B &\triangleq \omega_1 d_p^B, \quad d_q^B \triangleq \omega_2 d_q^B, \quad d_{p,q}^B \triangleq \omega_1 a_p^B - \omega_2 a_q^B \\ d_m^M &\triangleq \omega_1 (a_m^M - t_1 V), \quad d_n^M \triangleq \omega_2 (a_n^M - t_2 V) \\ d_{m,n}^M &\triangleq (\omega_2 t_2 - \omega_1 t_1) V + (\omega_1 a_m^M - \omega_2 a_n^M) \end{aligned}$$

$\mathcal{G}_{(\cdot,k)}^{(\cdot)}$  and  $\mathcal{F}_k^{(\cdot)}$  are the  $k^{\text{th}}$  FSCs of the APP and the AAS in the corresponding coordinates, respectively.  $J_k(u) \triangleq \frac{j^{-k}}{\pi} \int_0^\pi e^{j(k\xi + u \cos \xi)} d\xi$  is the  $k^{\text{th}}$ -order Bessel function,  $\otimes$  and  $|\cdot|$  denotes linear convolution and the Euclidian norm, respectively.

The norm of the separation vectors  $d_{p,q}^B, d_{m,n}^M$  represent shifted distances between  $\omega_1 a_p^B$  and  $\omega_2 a_q^B$  at the BS, and between  $\omega_1 (a_m^M - t_1 V)$  and  $\omega_2 (a_n^M - t_2 V)$  at the MS respectively. Large distances often result in less STF correlation as the Bessel functions asymptotically decrease. Parameters  $d_{(\cdot,k)}^{(\cdot)}$  contain space, time, and frequency separations between  $h_{pm}(t_1, \omega_1)$  and  $h_{qn}(t_2, \omega_2)$ .

### IV. TWO-DIMENSIONAL POWER SPECTRAL DENSITY OF 2D WB AND UWB MISO CHANNELS

We analyze the CCF derived in equation (7) in the frequency domain in order to see the temporal variations of the wireless channel. This analysis corresponds to the stationary scenario when  $\omega_1 = \omega_2 = \omega$  and  $m = n = 1$ . This is the case MISO system. From  $\angle d_{1,1}^M = \angle V + \angle(t_2 - t_1)$  we get:

$$\begin{aligned} R_{p1,q1}(t_1, t_2, \omega, \omega) &= \pi \frac{\omega_{bw}^{2\eta}}{2\omega^{2\eta+2}} \mathcal{W}(d_{p,q}^B, \mathcal{H}_k^B) \\ &\times \sum_{k=-\infty}^{\infty} j^k e^{jk\angle V} (\mathcal{G}_{1,k}^M(\omega) \otimes \mathcal{G}_{1,-k}^{M*}(\omega) \otimes \mathcal{F}_k^M) \\ &\times J_k\left(\frac{\omega(t_2 - t_1)|V|}{c}\right) \quad (9) \end{aligned}$$

Using the Fourier transform of  $J_k(u)$ , the Fourier transform of the CCF derived for stationary case versus the time-difference index,  $\Delta t \triangleq t_2 - t_1$  results in the following equation:

$$\begin{aligned} R_{p1,q1}(\Lambda, \omega) &\triangleq \int_{-\infty}^{\infty} e^{-j\Lambda \Delta t} R_{p1,q1}(t_1, t_2, \omega, \omega) d\Delta t \\ &= \frac{\omega_{bw}^{2\eta}}{2\omega^{2\eta+2}} \mathcal{W}(d_{q,p}^M, \mathcal{H}_k^B) \frac{\pi c}{|V|} \\ &\times \sum_{k=-\infty}^{\infty} (e^{jk\angle V} \mathcal{G}_{1,k}^M(\omega) \otimes \mathcal{G}_{1,-k}^{M*}(\omega) \otimes \mathcal{F}_k^M) \frac{T_k\left(\frac{c\Lambda}{|V|\omega}\right)}{\sqrt{1 - \left(\frac{c\Lambda}{|V|\omega}\right)^2}} \quad (10) \end{aligned}$$

where  $\mathcal{H}_k^B \triangleq \mathcal{G}_{p,k}^B(\omega) \otimes \mathcal{G}_{q,-k}^{B*}(\omega) \otimes \mathcal{F}_k^B$  and  $\Lambda$  is a frequency variable in the interval  $\frac{\omega}{c}|V| < \Lambda < \frac{\omega}{c}|V|$ . Note that  $R_{p1,q1}(\Lambda, \omega) = 0$  for all  $|\Lambda| \geq \frac{\omega}{c}|V|$ . The Chebyshev polynomials form a complete orthogonal set on the interval  $-1 \leq u < 1$ , with respect to the weighting function  $\frac{1}{\sqrt{1-u^2}}$ . Therefore, any bandlimited CCF (on the interval  $-\frac{\omega}{c}|V| \leq \Lambda \leq \frac{\omega}{c}|V|$ ) can be expanded in terms of Chebyshev polynomials as shown in the above expression.

In the following,  $R^M(\Lambda)$  is the last term in 10, and represents the impact of the non-isotropic environment, the APP, and the direction of the MS speed:

$$\begin{aligned} R^M(\Lambda) &\triangleq \sum_{k=-\infty}^{\infty} e^{jk\angle V} (\mathcal{G}_{1,k}^M(\omega) \otimes \mathcal{G}_{1,-k}^{M*}(\omega) \otimes \mathcal{F}_k^M) \\ &\times \frac{T_k\left(\frac{c\Lambda}{|V|\omega}\right)}{\sqrt{1 - \left(\frac{c\Lambda}{|V|\omega}\right)^2}} \quad (11) \end{aligned}$$

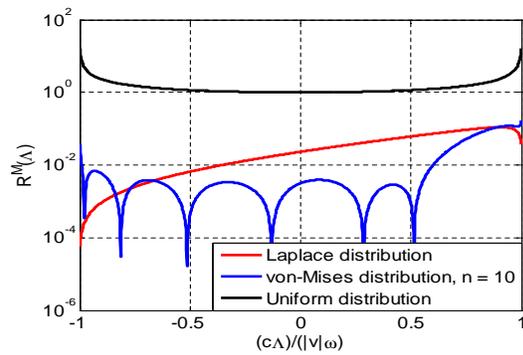
The term  $R^M(\Lambda)$ , is a PSD that represents the channel variations caused around or by the MS.

In Figures 3 and 4, this PSD is depicted for WB signals depending on the central frequency  $f = 2.5 \text{ GHz}$  and for UWB for the bandwidth  $3.1 \div 10.6 \text{ GHz}$ , depending on the following elements:

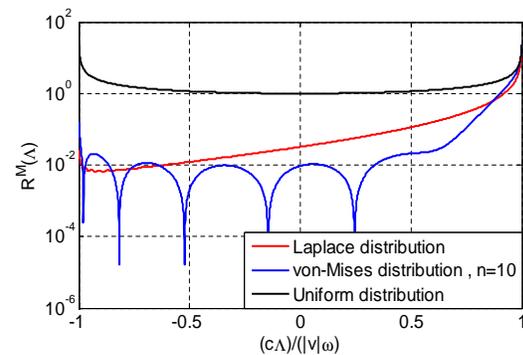
- i) the non-isotropic propagation environment around the MS is represented by Laplace and von-Mises distributions,
- ii) the WB and UWB antennas employed at the MS side are represented by the helical and rectangular APPs,
- iii) the direction of the MS speed is on the positive x-axis or the positive y-axis direction.

Analyzing the presented results we observe that three of the most important parameters which influences the PSD shape are APPs, the pdf of the propagation directions, the mobile speed direction and the channel bandwidth. The following observations can be formulated:

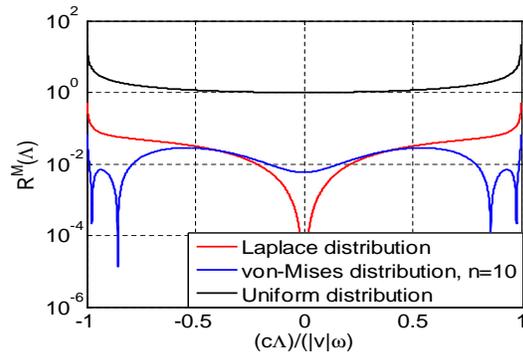
- For both WB and UWB channels, the maximum Doppler shift is  $\frac{\omega}{c}|V|$  (i.e.,  $R^M(\Lambda) = 0$ , if  $|\Lambda| \geq \frac{\omega}{c}|V|$ ).
- These results are consistent with the results proposed in [8], [13], [14]. In [8] it is presented the PSD of a narrowband channel in non-isotropic 2D propagation. The results presented in this work are similar to the PSD shape we obtained for WB channels. Between our results and the results presented in [8] there are similarities regarding the U-shape of the PSD but there are also differences determined by parameters characteristic to wideband channels like frequency selectivity, higher central frequency and APPs typically used for these types of channels. Comparing the PSD obtained for WB channels with those obtain for UWB channels, and even with those obtained for narrowband channels in [8] we can conclude that the channel bandwidth has a great influence on the PSD shape. When the bandwidth increases the channel frequency selectivity also increases and larger variations can be observed over the PSD envelope. The increased frequency selectivity is obviously larger in case of UWB channels than in case of WB channels. This feature offers the possibility to recognize the type of the channel for its bandwidth from its PSD only.
- It is clear that the majority of incoming/outgoing waves do travel in nearly horizontal directions and when the APP is directed along the vehicle motion the PSD has an asymmetrical shape. This is what it can be observed in our results: in Figure 3 when MS moves in the positive direction of the x-axis the PSD is larger at positive  $\Lambda$  than at negative  $\Lambda$ . This phenomenon is the consequence of the interaction between the beam of the antenna pattern, the direction of the MS speed, and the distribution of the propagation directions around the MS. This asymmetry of the PSD, is also determined by the Doppler spectrum which concentrates towards positive frequency axis.
- In Figure 4.4 all the PSD curves are symmetrical around the axis  $\Lambda = 0$ , because the pdf of the path directions and the APPs are symmetrical around  $\Theta^M = 0$ , and are perpendicular toward the speed direction.
- Generally speaking the PSD resulted for WB channels is less fluctuating than the PSD resulted for UWB channels. These fluctuations in the shape of the PSD of WB/UWB channels is the consequence of the frequency selectivity which increases with the signal frequency and signal bandwidth.



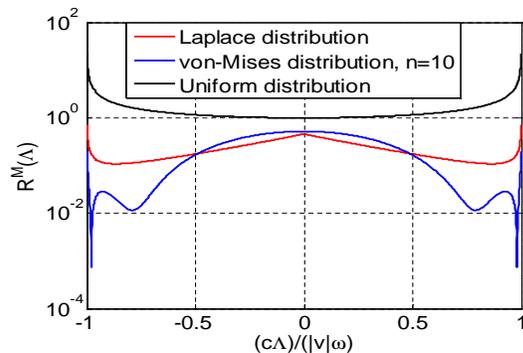
(a) Helical antenna



(b) Rectangular antenna



(c) Helical antenna



(d) Rectangular antenna

Figure 3: PSD of WB channels, MS moves in the positive direction of the x-axis (a,b) and y-axis (c,d), two antenna types employed at the MS, non-isotropic propagation (Laplacian or von-Mises distributed) and isotropic environment (uniformly distributed).

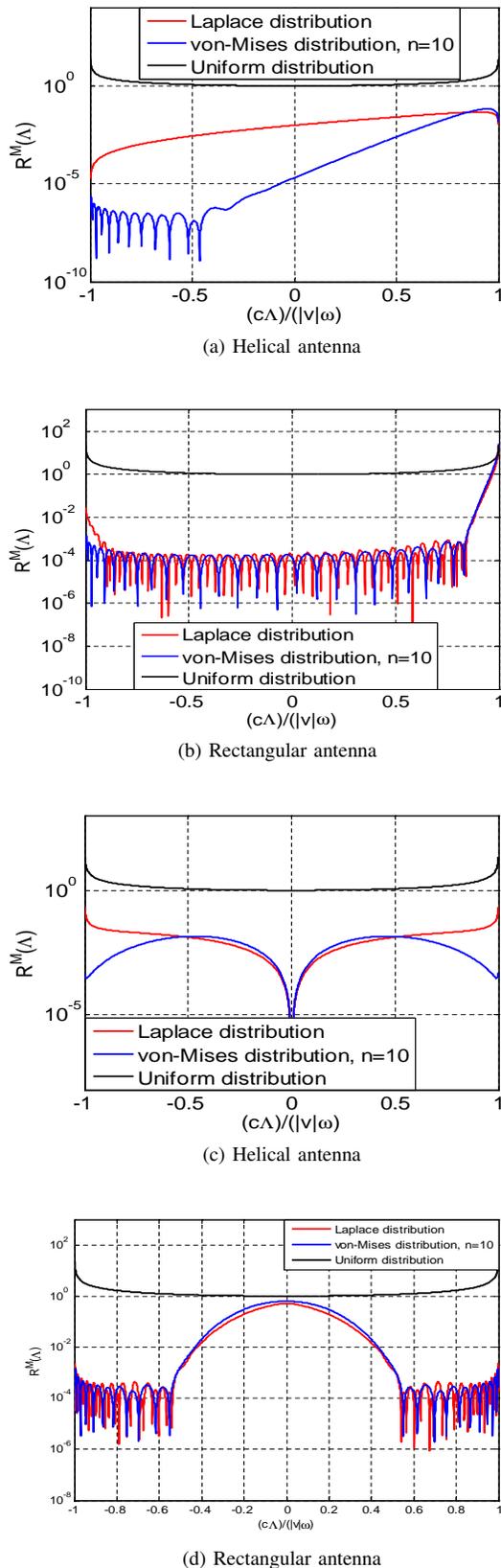


Figure 4: PSD of UWB channels, MS moves in the positive direction of the x-axis (a,b) and y-axis (c,d), two antenna types employed at the MS, non-isotropic propagation (Laplacian or von-Mises distributed) and isotropic environment (uniformly distributed) .

## V. CONCLUSION

In this paper, we investigated the impact of the non-uniform distribution of scatterers along with the non-omnidirectional APPs on the PSD for a 2D-WB/UWB MISO channel. The PSD is the result of the Fourier analysis of the CCF in a stationary case. It was observed that the PSD deviates from the U-shaped function, i.e., Clarke/Jake model. This deviation depends on the AAS, the employed antennas, and the direction of the MS speed. The results also prove that the range of frequencies over which the channel operates as well as the channel bandwidth have a great impact on the behavior of these characteristics.

## REFERENCES

- [1] A.M. Piştea, H. Rad, T. Paladeand A. Moldovan, Cross correlation function for wideband MIMO channels: derivation and analysis. Proceedings of the 6th International Wireless Communications and Mobile Computing Conference, pp. 819-823, July 2010,
- [2] R. B. Ertel, P. Cardieri, K.W. Sowerby, T. S. Rappaport, and J. H. Reed, Overview of spatial channel models for antenna array communication systems. *IEEE Pers. Commun.*, vol. 5, pp. 10–22, February 1998.
- [3] K. Yu. Multiple-Input Multiple-Output Radio Propagation Channels: Characteristics and Models. Doctoral thesis, Signals, Sensors and Systems, Royal Institute of Technology (KTH), 2005.
- [4] Y. Ma and M. Pätzold, Wideband two-ring MIMO channel models for mobile-to-mobile communications. Proc. 10th International Symposium on Wireless Personal Multimedia Communications, WPMC 2007. Jaipur, India, pp. 380–384, Dec. 2007.
- [5] M. Pätzold and B. O. Hogstad, A wideband MIMO channel model derived from the geometric elliptical scattering model. Processing of 3rd International Symposium on Wireless Communication System, pp. 138–143, Sept. 2006.
- [6] C.-C. Chong, C.-M. Tan, D. I. Laurenson, S. McLaughlin, M. A. Beach, and A. R. Nix. A new statistical wideband spatio-temporal channel model for 5-GHz band WLANsystems. *IEEE J. Select Areas Commun.*, vol. 21, pp. 139–150, Feb. 2003.
- [7] A. Balanis. *Antenna theory: analysis and design*. John Wiley & Sons, 2 edition, 1996.
- [8] H.S. Radand S. Gazor, The Impact of Non-Isotropic Scattering and Directional Antennas on MIMO Multicarrier Mobile Communication Channels. *IEEE Transactions on Communications* vol. 56, no. 4, pp. 642-652, Apr. 2008.
- [9] Q. Spencer, M. Rice and M. Jensen. A statistical model for the angle of arrival in indoor multipath Propagation, *IEEE Vehicular Technology Conference*, pp. 1415-1419, 1997.
- [10] A. Abdi, J. Barger, and M. Kaveh, A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station. *IEEE Trans. Veh. Technol.*, vol. 51, no. 1, pp. 425–434, May 2002.
- [11] K. Haneda and J. Takada, An application of SAGE algorithm for UWB propagation channel estimation. Proc. IEEE Conf. Ultra Wideband Systems and Technologies 2003 (UWBST2003), Reston, VA, USA, pp. 483-487, Nov. 2003.
- [12] G. D. Durgin, Theory of stochastic local area channel modeling for wireless communications. Ph.D. dissertation, Virginia Tech, Blacksburg, VA, Dec. 2000.
- [13] J. D. Parson and A. M. D. Turkmani, Characterisation of mobile radio signals: Base station crosscorrelation. *IEE Proceedings I. Communications, Speech and Vision*, 138(6), pp.557–565, December 1991.
- [14] J. D. Parson and A. M. D. Turkmani, Characterisation of mobile radio signals: Model description. *IEE Proceedings I. Communications, Speech and Vision*, vol. 138(6). pp. 549–556, Dec. 1991.

# Angular CMA: A modified Constant Modulus Algorithm providing steering angle updates

Koen C. H. Blom, Marcel D. van de Burgwal, Kenneth C. Rovers,  
André B. J. Kokkeler and Gerard J. M. Smit  
*Dep. of Electrical Engineering, Mathematics and Computer Science*  
*University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands*  
*Email: k.c.h.blom@utwente.nl, m.d.vandeburgwal@utwente.nl, k.c.rovers@utwente.nl,*  
*a.b.j.kokkeler@utwente.nl, g.j.m.smit@utwente.nl*

**Abstract**—Conventional blind beamforming algorithms have no direct notion of the physical Direction of Arrival angle of an impinging signal. These blind adaptive algorithms operate by adjusting the complex steering vector in the case of changing signal conditions and directions. This paper presents Angular CMA, a blind beamforming method that calculates steering angle updates (instead of weight vector updates) to keep track of the desired signal. Angular CMA and its respective steering angle updates are particularly useful in the context of mixed-signal hierarchical arrays as means to find and distribute steering parameters. Simulations of Angular CMA show promising convergence behaviour, while having a lower complexity than alternative methods (e.g., MUSIC).

**Keywords**—blind beamforming; CMA; hierarchical arrays

## I. INTRODUCTION

Adaptive beamforming algorithms operate by adjusting the beamformer steering vector in the case of changing signal (conditions and) directions. A subclass of these adaptive algorithms is the subclass of the blind beamforming algorithms. These algorithms use structural properties of the desired signal as a reference to calculate appropriate weight adjustments and have no direct notion of the physical Direction of Arrival (DOA) angle of this signal [1].

Existing blind beamforming algorithms operate on all antenna inputs and update the entire complex steering vector. However, in mixed-signal hierarchical arrays this situation is different. Here, beamforming is performed on multiple levels, partly analog and partly digital. Spatial interference is already being suppressed by the analog beamformers before the signal is being digitized [2]. These analog beamformers decrease the dynamic range of the input signals to the Analog-to-Digital Converters (ADCs), therefore requirements for the ADCs are lowered. The remaining interference can be further suppressed or nulled in the digital domain. Within this layered architecture, a classical digital blind beamforming algorithm has merely the results of the analog beamformers available and can only update its own digital steering parameters. However, for deterministic steering of

the complete hierarchical system a novel adaptive algorithm is required that can efficiently track a desired signal and distribute steering parameters throughout the complete array.

This paper presents a modified blind beamforming algorithm that senses the presence of mispointing and calculates the required steering angle updates (instead of weight vector updates) to keep track of the desired signal. The desired steering angle can now be calculated at the digital level of the layered architecture and thereafter distributed to both the analog and digital beamformers. This approach is based on the conventional Constant Modulus Algorithm (CMA) algorithm, but since it operates on the steering angle instead of the steering vector, it is called Angular CMA.

Both CMA and Angular CMA use the Constant Modulus (CM) property of the received signal as a reference to calculate appropriate weight adjustments. The search space of CMA constitutes of all complex weights that make up the steering vector [3]. However, in Angular CMA this search space is reduced to a single steering angle.

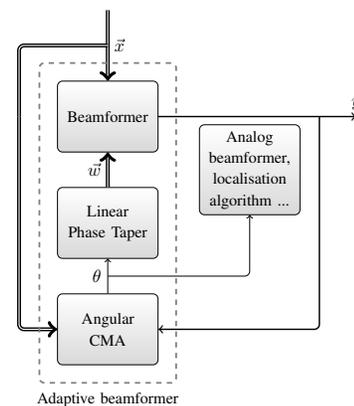


Figure 1. Angular CMA based adaptive beamformer.

This work describes Angular CMA in the context of a simplified (non-hierarchical) array architecture that is shown in Figure 1. Herein, Angular CMA operates on narrowband signals received by an  $N$ -element Uniform Linear Array

Supported through STW projects: CMOS Beamforming (07620) and SeaSTAR (10552).

(ULA). The output of the ULA at each sample instant is a vector of  $N$  quadrature baseband samples and is indicated by  $\vec{x}$ . A changing DOA angle affects the antenna data  $\vec{x}$ . Traditional CMA compensates for these effects by updating the steering vector  $\vec{w}$  directly. Angular CMA first calculates the desired steering angle  $\theta$  and uses a Linear Phase Taper (LPT) to find  $\vec{w}$ . As shown in Figure 1, the steering angle  $\theta$  can also be used for steering analog beamformers and for applications that require a notion of the physical DOA angle.

A basic classification of adaptive array algorithms can be found in Section II, followed by a concise description of CMA in Section III. Thereafter, the derivation of Angular CMA is given in Section IV. Section V compares simulation results of Angular CMA and traditional CMA with as main point of comparison their convergence behaviour. The computational complexity of Angular CMA is discussed in Section VI. Advantages and disadvantages of Angular CMA are discussed in Section VII. Finally, a short overview of the most significant results and future work is given in Section VIII.

## II. RELATED WORK

Adaptive beamforming algorithms can be categorized in three classes: temporal reference, spatial reference and blind beamforming algorithms [1]. Temporal reference algorithms use training sequences in the received signal to calculate adjustments for the array steering vector. These algorithms can only be used in case training sequences are available and therefore their use is limited. We aim for a more generic solution, e.g., to track Digital Video Broadcasting Satellite (DVB-S) signals.

Spatial reference algorithms are based on the spatial correlation characteristics of the received antenna signals. A major application of these algorithms is DOA angle determination of the impinging signals. Some methods in this class are: Multiple Signal Classification (MUSIC), Estimation of Signal Parameters by Rotational Invariance Techniques (ESPRIT) and Maximum Likelihood (ML) based techniques [4][5][6]. Typically spatial reference algorithms have a high computational complexity.

Blind beamforming algorithms, also known as blind deconvolution algorithms, perform adaptive inverse filtering in an unsupervised manner [7]. In this paper the CM property of the received signal is exploited to perform blind deconvolution. Originally, this idea was proposed by Godard [8]. Independently of Godard, a special case of this idea was published by Treichler and Agee [9]. They named their algorithm the Constant Modulus Algorithm (CMA). The algorithms from Godard and from Treichler and Agee, minimize a nonconvex cost function by adapting complex weights. In contrast to adapting all complex weights this work introduces a special case of CMA that directly acts on the array steering angle.

## III. CONSTANT MODULUS ALGORITHM (CMA)

CMA penalizes deviation of the beamformer output from a constant modulus. The cost function  $J$  is defined as the expected deviation of the squared modulus beamformer output  $y$  with respect to the constant  $R_2$  [7]:

$$J = E\{(|y[k]|^2 - R_2)^2\} \quad (1)$$

Herein,  $k$  is an index for different sampling instants and ‘ $E$ ’ represents expectation.  $R_2$  is chosen such that the gradient of  $J$  is zero when minimum costs are reached and is written as:

$$R_2 = \frac{E[|x[k]|^4]}{E[|x[k]|^2]} \quad (2)$$

For an ideal normalized CM beamformer output the constant  $R_2$  is equal to one. Therefore, in our further analysis of CMA and in the derivation of Angular CMA the constant  $R_2$  is assumed to be one. The cost function  $J$  is illustrated in Figure 2.

Minimum costs are reached when the beamformer output  $y$  has unit modulus. The beamformer output can be expressed as  $\vec{w}^H \vec{x}$ . The aim of CMA is to minimize  $J$  by altering  $\vec{w}$ . A stochastic gradient descent technique can be used to attain this goal. The gradient of  $J$  with respect to  $\vec{w}$  is found as follows:

$$J = E\{(|y|^2 - 1)^2\} \quad (3)$$

$$\nabla_{\vec{w}} J = 2 \cdot E\{(|y|^2 - 1) \cdot \nabla_{\vec{w}} (|\vec{w}^H \vec{x}|^2 - 1)\} \quad (4)$$

$$\nabla_{\vec{w}} J = 2 \cdot E\{(|y|^2 - 1) \cdot \nabla_{\vec{w}} (\vec{w}^H \vec{x} \vec{x}^H \vec{w} - 1)\} \quad (5)$$

Continue using the fact that  $\vec{x} \vec{x}^H$  is a Hermitian matrix and  $\nabla_{\vec{w}} \vec{w}^H A \vec{w} = 2A\vec{w}$  holds for every Hermitian matrix  $A$  [10]:

$$\nabla_{\vec{w}} J = 4 \cdot E\{(|y|^2 - 1) \cdot \vec{x} \vec{x}^H \vec{w}\} \quad (6)$$

$$\nabla_{\vec{w}} J = 4 \cdot E\{(|y|^2 - 1) \cdot y^* \vec{x}\} \quad (7)$$

Based on Equation 7 the minimizer for the stochastic gradient descent version of CMA is found:

$$\vec{w}[k+1] = \vec{w}[k] - \mu \cdot (|y[k]|^2 - 1) \cdot y[k]^* \vec{x}[k] \quad (8)$$

Herein,  $\mu$  absorbs the factor 4 and determines the convergence speed.

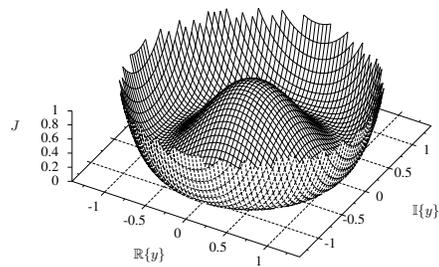


Figure 2. Surface plot of the CMA cost function ( $R_2 = 1$ ).

#### IV. ANGULAR CMA

The previous section mentions that the CMA minimizer is found by expressing  $y$  in Equation 3 as  $\vec{w}^H \vec{x}$ , followed by derivation of the gradient of  $J$  with respect to  $\vec{w}$ . Angular CMA is constructed by expressing  $y$  in Equation 3 as  $\vec{w}(\theta)^H \vec{x}$ , followed by derivation of  $J$  with respect to  $\theta$ . The term  $\vec{w}(\theta)$  is called a Linear Phase Taper (LPT) and is explained in the next section. Successively, the error-performance surface and the derivation of Angular-CMA minimizer are discussed.

##### A. Linear Phase Taper (LPT)

A LPT is a linear phase variation across the array aperture, which produces a beamshift of the main beam without any change in sidelobe structure [11]. Figure 1 shows that given a desired steering angle  $\theta$  the LPT calculates the required output steering vector  $\vec{w}$ . If one of the outer array element acts as a phase reference then the LPT for an  $N$ -element ULA can be written as:

$$\vec{w}(\theta) = e^{\phi(\theta) \cdot \vec{n}}, \quad \vec{n} = [0 \dots (N-1)]^T \quad (9)$$

With  $\phi(\theta)$  defined as:

$$\phi(\theta) = \frac{j2\pi \cdot d \sin(\theta)}{\lambda} \quad (10)$$

Herein,  $d$  represents the spacing of the antenna elements and  $\lambda$  the wavelength of the desired signal.

##### B. Error-performance surface

The dependence of the cost function  $J$  on the weights  $\vec{w}$  is called the error-performance surface [7]. For CMA this dependence is an  $(N+1)$ -dimensional surface with  $N$  degrees of freedom, where  $N$  is the number of antenna elements. In the case of more than two degrees of freedom such a surface is hard to visualize.

The error-performance surface of Angular CMA is two-dimensional with only one degree of freedom (the steering angle  $\theta$ ). Therefore, its error-performance surface can be visualized using a two-dimensional plot. Mathematically, this surface can be written as:

$$J(\theta) = (|\vec{w}(\theta)^H \vec{x}|^2 - 1)^2 = \left( \left| \left( e^{\frac{j2\pi \cdot d \sin(\theta)}{\lambda} \cdot \vec{n}} \right)^H \vec{x} \right|^2 - 1 \right)^2 \quad (11)$$

Herein, the expectation operator 'E' is dropped because noiseless antenna data is taken for plotting.

Figure 3 shows the error-performance surface of Angular CMA for a  $\frac{\lambda}{2}$  element spacing and signals arriving from broadside ( $\theta = 0$ ). The surface is drawn for a two-, four- and eight-element array. For all three array configurations holds that a steering angle  $\theta$  of zero degrees has the lowest costs. This behaviour is expected because the signals arrive from the broadside direction.

Note that, based on Figure 3 (and Equation 11), the error-performance surface of Angular CMA has the appearance of

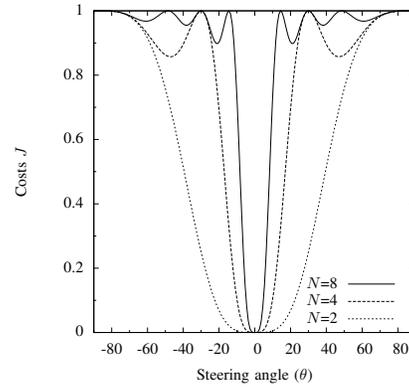


Figure 3. Error-performance surface of Angular CMA.

a vertically flipped beam pattern. An increase in the number of antenna elements results in a smaller angular range where convergence to the global minimum can be guaranteed. This region corresponds to the Inter-Null Beamwidth (INBW) of an array pattern. The INBW is defined as the difference between the nearest two nulls around a given angle [1]. If this given angle is the center of the main beam then the INBW can be expressed as follows:

$$\text{INBW} = 2 \sin^{-1}(\lambda/(dN)) \quad (12)$$

Thus, for an eight element array with  $\frac{\lambda}{2}$  element spacing Angular CMA has a convergence region width of  $2 \sin^{-1}(\frac{1}{4}) \approx 29^\circ$ .

DOA estimation algorithms, such as MUSIC, can be used to provide an initial steering angle for Angular CMA [4]. The accuracy of this estimate should be within the angular range of Angular CMA that provides convergence to the global minimum.

##### C. Derivation of the minimizer

The minimizer of Angular CMA can be found by expressing  $y$  in Equation 3 as  $\vec{w}(\theta)^H \vec{x} = (e^{\phi(\theta) \cdot \vec{n}})^H \vec{x}$ . Thereafter, the first derivative of  $J$  with respect to  $\theta$  is determined:

$$J(\theta) = E\{(|(e^{\phi(\theta) \cdot \vec{n}})^H \vec{x}|^2 - 1)^2\} \quad (13)$$

$$\frac{\partial}{\partial \theta} J = E\{2(|y|^2 - 1) \cdot \frac{\partial}{\partial \theta} ((e^{\phi \cdot \vec{n}})^H \vec{x} \vec{x}^H e^{\phi \cdot \vec{n}} - 1)\} \quad (14)$$

$$\frac{\partial}{\partial \theta} J = 2E\{(|y|^2 - 1) \cdot \frac{\partial}{\partial \theta} (\vec{x}^H e^{\phi \cdot \vec{n}} (e^{-\phi \cdot \vec{n}})^T \vec{x} - 1)\} \quad (15)$$

Continue by writing  $e^{\phi \cdot \vec{n}} (e^{-\phi \cdot \vec{n}})^T$  as matrix  $B$ :

$$\frac{\partial}{\partial \theta} J = 2E\{(|y|^2 - 1) \cdot \frac{\partial}{\partial \theta} (\vec{x}^H B \vec{x} - 1)\} \quad (16)$$

$$B = \begin{bmatrix} 1 & \dots & e^{\phi \cdot (n_0 - n_{N-1})} \\ \vdots & & \vdots \\ e^{\phi \cdot (n_{N-1} - n_0)} & \dots & 1 \end{bmatrix} \quad (17)$$

The derivative of  $J$  with respect to  $\theta$  is written in Equation 18. Note that  $\phi$  is dependent on  $\theta$ .

$$\frac{\partial}{\partial \theta} J = 2E\{(|y|^2 - 1) \cdot (\bar{x}^H B'(\theta) \bar{x})\} \quad (18)$$

Where  $B'(\theta) = \frac{\partial}{\partial \theta} B(\theta)$  can be written as:

$$\begin{bmatrix} 0 & \cdots & (n_0 - n_{N-1})\phi' e^{(n_0 - n_{N-1})\phi} \\ \vdots & & \vdots \\ (n_{N-1} - n_0)\phi' e^{(n_{N-1} - n_0)\phi} & \cdots & 0 \end{bmatrix} \quad (19)$$

Herein,  $\phi'(\theta)$  is  $\frac{\partial}{\partial \theta} \phi(\theta)$ :

$$\phi'(\theta) = \frac{j2\pi \cdot d \cos(\theta)}{\lambda} \quad (20)$$

The use of an instantaneous gradient estimate of Equation 18 yields the following algorithm:

$$\theta[k+1] = \theta[k] - \mu(|y[k]|^2 - 1) \cdot (\bar{x}[k]^H B'(\theta[k]) \bar{x}[k]) \quad (21)$$

Herein, the value 2 is absorbed in the convergence factor  $\mu$ .

## V. SIMULATION RESULTS

In this section, the convergence of Angular CMA is examined by looking at the cost and Mean Square Error (MSE) behaviour during adaptation. Simulations are performed for both Angular CMA and traditional CMA to provide means for comparison.

For simulation the adaptive beamformer architecture presented in Figure 1 is implemented. An eight element ULA with  $\frac{\lambda}{2}$  element spacing is assumed. The baseband samples are Quadrature Phase-Shift Keying (QPSK) modulated and have a Signal-to-Noise Ratio (SNR) of 16 dB. Note that QPSK signals possess the CM property, which is required by both Angular CMA and CMA.

### A. Learning curve

The performance of an adaptive algorithm is often studied by looking at the ‘learning curve’ of the cost function  $J$ . As mentioned in the theoretical analysis, an eight element array running Angular CMA should be able to converge to the correct array steering angle if the initial steering angle stays within the  $29^\circ$  wide convergence region. Validity of this statement is checked in simulation. This is done by setting the DOA angle to broadside ( $0^\circ$ ), while having the initial steering angle set to one of the extremes of the convergence region ( $0^\circ \pm \frac{29}{2} \approx \pm 14^\circ$ ). Figure 4 shows that Angular CMA rapidly converges to the correct steering angle and consequently minimizes the costs.

The previous simulation uses a convergence factor  $\mu$  of  $5 \cdot 10^{-2}$ , in contrast to  $5 \cdot 10^{-3}$  typically used for CMA. Angular CMA is still robust with a large convergence factor, because the gradient descent ensures global minimization if the steering angle is initially within the convergence region.

In an attempt to compare the convergence properties of Angular CMA and CMA, the cost behaviour of CMA is

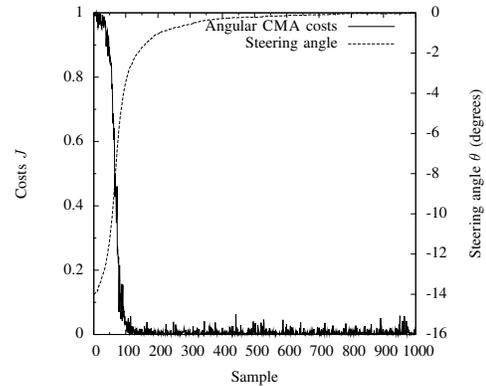


Figure 4. Convergence behaviour of Angular CMA ( $\mu = 5 \cdot 10^{-2}$ ).

plotted in Figure 5 using the same simulation parameters as in Figure 4. Clearly, CMA takes more samples to convergence than Angular CMA. The rapid convergence of Angular CMA is caused by its drastic search space reduction from  $N$  complex weights to one real value. A theoretical analysis on the convergence properties of Angular CMA and choosing the factor  $\mu$  is beyond the scope of this paper.

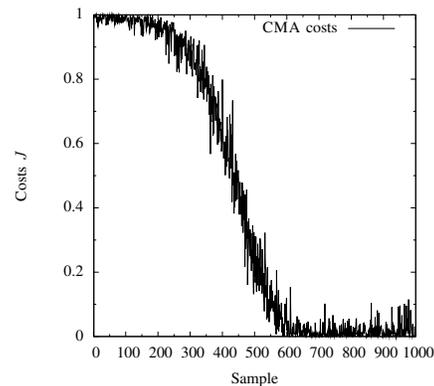


Figure 5. Convergence behaviour of CMA ( $\mu = 5 \cdot 10^{-2}$ ).

### B. Mean Square Error (MSE)

A common method to evaluate an equalizer is to determine the MSE of its output signal (predetection error). Both CMA and Angular CMA do not correct phase offsets in the beamformer output signal. This phase blindness should not affect the MSE of their output signals, therefore the following method is used to calculate the MSE:

$$\text{MSE} \triangleq \min_{\alpha} E\{|\bar{w}[k]^H \bar{x}[k] - e^{j\alpha} s[k]|^2\} \quad (22)$$

Herein,  $s[k]$  represents the (noiseless) transmitted signal at sample instant  $k$ . This unconventional method of MSE calculation is introduced by Treichler and Agee [9].

Estimates for the MSE of the CMA and Angular CMA outputs can be seen in Figure 6. The estimates are based on

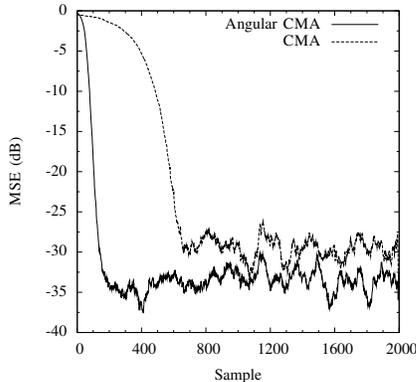


Figure 6. MSE comparison ( $\mu = 5 \cdot 10^{-2}$ ).

the same scenario as in the previous simulation. The results are smoothed by a moving average filter to reveal trends.

Figure 6 reveals that the MSE for both algorithms drop, as expected, on the same time scale as the costs in Figure 4 and Figure 5. Interestingly, the MSE level of Angular CMA after convergence is below that of CMA.

## VI. COMPLEXITY ANALYSIS

A short analysis on the complexity of Angular CMA is given to assess scalability and implementability. The computational complexity of Angular CMA is compared with that of CMA and MUSIC. MUSIC is included in the comparison because it provides DOA estimates of all impinging signals. Therefore, repeated execution of MUSIC could be used for tracking. However, this requires a method to identify the DOA of the desired signal out of all angles found by MUSIC.

The computational order of complexity of the algorithms is determined by counting the number of complex Multiply-Accumulate (MAC) operations. The complexity of Angular CMA is determined based on Equation 21. Calculation of matrix  $B'(\theta[k])$  is kept out of the complexity analysis since we expect that, by exploiting its Hermitian symmetry, this can be done very efficiently. Therefore, not taken calculation of  $B'(\theta[k])$  into account, the order of Angular CMA is dominated by the matrix-vector multiplication  $B'(\theta[k])\vec{x}[k]$  and is for that reason  $\Theta(N^2)$ . Herein,  $N$  is the number of antenna elements.

The stochastic gradient descent version of CMA, often referred to as Stochastic Gradient CMA (SG-CMA), is of order  $\Theta(N)$  [12]. MUSIC is computationally much more expensive with a complexity of order  $\Theta(N^3)$  [13]. Thus, the complexity of Angular CMA is in between that of CMA and MUSIC.

## VII. DISCUSSION

The reduction of the CMA search space from  $N$  complex weights to a single steering angle has certain important

implications. The main advantages and disadvantages of Angular CMA are discussed in this section.

Traditional CMA is attractive for application in array architectures because of its insensitivity to array imperfections [3], low complexity and because training sequences are not required. Since Angular CMA controls steering by distributing only a single steering parameter, it can not correct for small array imperfections.

Despite the fact that a single steering parameter cannot cope with array imperfections, it is advantageous in systems where the dynamics of the source can be expressed by a Linear Time-Invariant (LTI) system. When dealing with such a system the notion of a physical DOA angle can be input to a state estimation technique, such as the Kalman filter [7]. Erratic behaviour in the steering angle can then be corrected and most likely the update rate of Angular CMA can be further reduced.

CMA minimizes the effect of interferers by placing nulls at their respective DOA angles. Angular CMA is restricted to array responses that can be generated by a LPT. Therefore, it cannot place nulls as CMA can. Nonetheless, after correct convergence of Angular CMA, exact numbers on the suppression of other directions can be given. These numbers are found by evaluating the LPT array response based on the steering angle output from Angular CMA. In hierarchical arrays exact directivity properties for each array level ease design choices for the aggregate structure.

## VIII. CONCLUSION AND FUTURE WORK

Angular CMA, a modified version of CMA, is presented in this work. The algorithm calculates steering angle updates to keep track of the desired signal. Within the context of hierarchical arrays it provides a means for efficiently tracking and distributing steering parameters. The cost behaviour of Angular CMA is compared with that of CMA. Angular CMA provides faster convergence and a lower MSE floor. The complexity of the algorithm is of order  $\Theta(N^2)$ . Compared to the complexity of other methods for steering angle calculation (e.g., MUSIC), our method is favorable.

Angular CMA is investigated within the context of a simplified non-hierarchical beamformer. Further research should look at the applicability of the algorithm in a hierarchical setup in simulation and in practice. The algorithm cannot correct for small array imperfections. Whether these influence the convergence of Angular CMA should be examined. Our current work is focused at extending the angular convergence region by exploiting other optimization techniques and incorporating other cost functions to deal with phase blindness [14].

## ACKNOWLEDGMENTS

The authors would like to thank Marco Gerards for assistance on the mathematics.

## REFERENCES

- [1] B. Allen and M. Ghavami, *Adaptive Array Systems, Fundamentals and Applications*. John Wiley & Sons, 2005.
- [2] B. Smolders and G. Hampson, "Deterministic RF nulling in phased arrays for the next generation of radio telescopes," in *IEEE Transactions on Antennas and Propagation*, 1999.
- [3] R. Gooch and J. Lundell, "The CM array: An adaptive beamformer for constant modulus signals," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86.*, vol. 11, Apr 1986, pp. 2523–2526.
- [4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *Antennas and Propagation, IEEE Transactions on*, vol. 34, no. 3, pp. 276–280, Mar 1986.
- [5] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.
- [6] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 10, pp. 1553–1560, Oct. 1988.
- [7] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice Hall, 2001.
- [8] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *Communications, IEEE Transactions on*, vol. 28, no. 11, pp. 1867–1875, Nov. 1980.
- [9] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 31, no. 2, pp. 459–472, Apr 1983.
- [10] P. Nelson and S. Elliott, *Active Control of Sound*. Academic Press, 1993.
- [11] A. Rudge, *The Handbook of Antenna Design: Volume 1*, K. Milne, A. Olver, and P. Knight, Eds. Institution of Electrical Engineers, 1982.
- [12] V. Zarzoso and P. Comon, "Blind channel equalization with algebraic optimal step size," in *Proc. XIII Eur. Signal Process. Conf*, 2005.
- [13] M. Rubsamen and A. Gershman, "Direction-of-arrival estimation for nonuniform sensor arrays: From manifold separation to fourier domain music methods," *Signal Processing, IEEE Transactions on*, vol. 57, no. 2, pp. 588–599, 2009.
- [14] K. Blom, M. van de Burgwal, K. Rovers, A. Kokkeler, and G. Smit, "DVB-S signal tracking techniques for mobile phased arrays," in *Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd*, 2010, pp. 1–5.

# A Smart Predictive Link Layer Trigger Algorithm to Optimize Homogenous/Heterogeneous Networks WiFi Handover Decisions

Chunmei Liu  
Clearwire Corp.  
Herndon, VA, USA  
mayliu@alum.mit.edu

Christian Maciocco  
Intel Labs, Circuits and Systems Research  
Hillsboro, OR, USA  
christian.maciocco@intel.com

**Abstract**— In traditional wireless networks, link layer metrics used to trigger handover are mainly signal quality based, such as Received Signal Strength (RSS). While signal quality is one major reason for poor performance in traditional wireless networks, WiFi networks are special in that there is another major reason for poor performance, which is collision. While there exist some metrics to reflect collisions and trigger handover when necessary, such as channel load, this paper explains why these existing metrics, such as channel load, cannot capture the actual collision situation in the network and that one station is experiencing. Based on this observation, this paper proposes a new metric, called station collision probability, as an additional handover trigger metric, and develops a prediction algorithm for this metric. Specifically, for WiFi networks, station collision probability is the probability that a packet being transmitted by a station incurs a collision. A prediction algorithm is developed for station collision probability on unlicensed WiFi networks, which takes the number of collisions between two successful transmissions on the channel as the measurement and predicts the station collision probability by solving a developed equation. The algorithm does not require the station to send any traffic, and applies to real time decisions, including predictive handover decisions to initiate and prepare the handover to reduce latency and service interruption for the end-users. This paper focuses on defining an optimal collision estimation algorithm and the simulation results validates that the algorithm predicts station collision probability and adapts well to the change of network traffic. The predicted station collision probability can then be integrated with the existing signal quality based trigger metrics to trigger handover, which is beyond the scope of this paper and will be the next steps.

**Keywords**- predictive handover trigger; station collision probability; WiFi networks; intra- and inter-technology handover.

## I. INTRODUCTION

In wireless networks, handover is one of the key approaches to ensure user experience and network performance. From user perspective, a user may request handover when it cannot receive quality services from its current access point (AP). From network perspective, an AP may request a group of users to perform handover to balance load. This applies to both homogenous and heterogeneous networks.

Figure 1 illustrates a typical handover architecture based on IEEE 802.21 [1], which is a standard that focuses on

media-independent handover (MIH) method and procedures. The first step in this architecture is link layer trigger, which is also the focus of this paper. The objective of link layer trigger is to provide algorithms to trigger handover when link is not “good”, viewed from data link layer and below. The link layer trigger is then followed by network selection and handover execution.

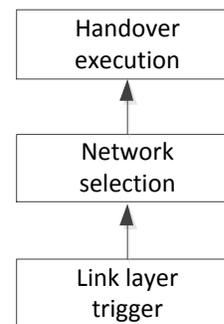


Figure 1. 802.21 based handover architecture.

Link layer trigger consists of the following three portions: 1) trigger metric selection; 2) metric estimation/prediction; 3) metric based trigger algorithm. This paper focuses on the first two: handover trigger metric selection and estimation/prediction. Essentially, this paper considers predictive link layer handover triggers whose purpose is to allow early handover initiation and preparation so that the handover latency and service interruption time can be reduced significantly.

In traditional wireless networks, signal quality related metrics, such as Received Signal Strength (RSS), are typically the metrics of choice used in conventional handover algorithms. Gregory [2] reviews conventional trigger algorithms based on RSS, where RSS used is averaged RSS. Alexe, Vijayan, Zhang, et al. [3][4][5] provide analysis on the handover performance based on these algorithms. Zonoozi et al. [6] further optimizes the parameters including the hysteresis level and signal averaging time.

Signal quality related metrics, such as RSS, captures how good the physical channel is, and gives limits on the max data rate the station can receive, which is directly related to

the station performance. In addition to signal quality, there is another significant aspect that directly impacts station performance, that is, *how much time a station can effectively utilize the physical channel*. The latter is especially important for WiFi networks due to its nature of contention-based MAC, where stations under one AP compete for channel access.

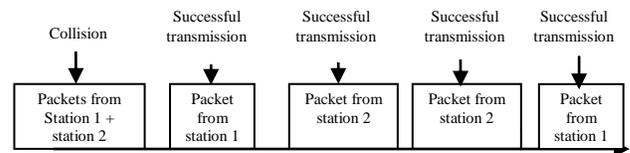
One example of the importance of the latter with respect to station's performance is that, in a WiFi network, even when a station is close to an access point (AP) and the signal quality is excellent, if the network is highly congested, the transmissions from this station will incur collisions coming from other competing stations with high probability. These collisions result in not only packet losses, but also additional backoff stages. All the collision slots and backoff slots are not actually utilized by the station, in that either the transmissions from the station incur collisions and fail (collision slots), or there is no transmission from the station at all (backoff slots). The only effective slots are those that contain successful transmissions. Apparently in this scenario, the time when the station can effectively utilize the physical channel is limited. As a result, although the signal quality is excellent and the channel is excellent, the station experiences significantly deteriorated performance such as low throughput and high delay. Accordingly, in such scenarios, the station should search for handover opportunities, either to another AP within the WiFi network, or to a base station (BS) of another network such as a WiMAX BS.

This paper focuses on the latter aspect, that is, *“how much time can a station effectively utilize the physical channel”*. Specifically, in addition to existing signal quality related metrics, this paper proposes to use station collision probability as a metric to trigger handover from WiFi networks. Station collision probability of a station is defined to be the probability that a packet being transmitted by the station incurs a collision. This probability is the same as the conditional collision probability in Bianchi [7]. Since this probability is to be used as a handover trigger metric, the term station collision probability is used for clarity and to avoid mathematical terms. According to the WiFi backoff mechanism specified in the standard 802.11 [8], once the station collision probability is known, the time that the station can effectively utilize the channel can be calculated from the backoff mechanism. Together with the signal quality, the station performance can then be estimated and a handover can be triggered when necessary.

Note that while currently there do exist some metrics that are somewhat related to MAC collisions and are being used to trigger handover from WiFi networks, such as channel load, they do not reflect collisions. One such an example is when there is only one station transmitting within an AP cell, but the station is saturated, meaning that it always has

packet to transmit. In this case the channel load is high, while there are 0 collisions and the station receives excellent service from the AP. Hence the station should stay with this cell instead of triggering a handover to some other networks.

Also note that since one collision may involve several packets, station collision probability is different from the channel collision probability, which is the probability of a collision among all the transmissions on the channel. Hence station collision probability cannot be estimated by simply computing the percentage of collisions over overall transmissions on the channel. Figure 2 shows an example which illustrates these two probabilities and their differences. It shows that packet collision probabilities for station 1 and 2 are 1/3, respectively, while channel collision probability is 1/5, which does not equal to the station probability of 1/3.



- Station collision probability of station 1 and 2 = 1/3: each station transmitted 3 packets, one of which incurs a collision.
- Channel collision probability = 1/5: the channel sees 5 transmissions, one of which is a collision.
- Station collision probability ≠ channel collision probability

Figure 2. Station collision probability vs channel collision probability.

In addition to propose station collision probability as a handover trigger, this paper develops a prediction algorithm for station collision probability. The algorithm measures number of collisions between two successful transmissions occurring on the channel, computes its average, and then obtains the station collision probability by solving one equation. Since the algorithm uses all channel data, which includes data from other stations instead of data from the considered station only, the algorithm does not require the station to send any traffic. In addition, channel data provides many more samples within a period of time than the station's own data. Hence the proposed algorithm does not require probing, adapts to network traffic changes well, and applies to real-time decisions, including handover decisions.

The predicted station collision probability can then be integrated with the existing RSS based link-layer predictive triggers, such as those described in Huaiyu, Choong, et al. [9][10], to optimize network selection and seamless handover over homogenous or heterogeneous networks. Note that one major difference between the triggers described in Huaiyu, Choong, et al. [9][10] and the triggers proposed in this paper is that the triggers in Huaiyu, Choong, et al. [9][10] are RSS based, while the triggers in the paper focuses on *“how much time can a station*

effectively utilize the physical channel". The integration of the triggers is beyond the scope of this paper and will be our future work.

This paper is outlined as follows: Section II derives the prediction algorithm for station collision probability. Section III validates the prediction algorithm using OPNET-based simulations. Section IV concludes the paper.

## II. PREDICTION ALGORITHM FOR STATION COLLISION PROBABILITY

### A. Analytical Model

The development of the prediction algorithm for station collision probability is based on the analysis framework proposed in Bianchi [7]. In Bianchi [7], by assuming ideal channel condition, the author considered saturated network and analyzes DCF (Distributed Coordination Function) performance. It is assumed that at each transmission attempt, each packet collides with constant and independent probability. A single station is then modeled as a bi-dimensional Markov chain, with the backoff stage and counter as states, and transition probabilities as functions of assumed constant collision probability. The Markov chain is solved to obtain collision probabilities, saturated throughput, and other parameters. The results show good consistency of the analysis with simulation results.

Two major results obtained in Bianchi [7] are the relationship among the number of saturated stations  $n$ , the probability that a station transmits in a randomly chosen slot  $\tau$ , and station collision probability  $p$ , as follows:

$$\tau = \frac{2(1-2p)}{(1-2p)(W+1) + pW(1-(2p)^m)}, \quad (1)$$

$$p = 1 - (1-\tau)^{n-1}, \quad (2)$$

where  $m$  is the value such that  $CW_{\max} = 2^m CW_{\min}$ ,  $W = CW_{\min}$ ,  $CW_{\min}$  and  $CW_{\max}$  are the minimum and maximum contention window, respectively.

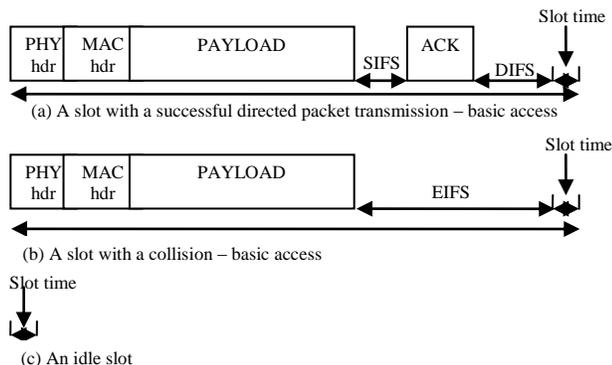


Figure 3. Generic slots in basic access mechanism.

In order to be consistent with the 802.11 standard [8], this paper slightly modifies the format of generic slot used in Bianchi [7] as follows: in Bianchi [7], for basic access, a generic slot that includes a collision consists of PHY and

MAC header plus payload plus DIFS. According to the 802.11 standard [8], a station shall use EIFS, instead of DIFS, before transmission, when it determines that the medium is idle following a collision. Hence the generic slot in this paper is defined as those in Figure 3.

Fortunately, the basic unit of the Markov model in Bianchi [7] is a generic slot, and the Markov model does not change with the format of the generic slot. Hence the analysis in Bianchi [7] can be reused under the new definition of generic slots, and equation (1) and (2) above, which are derived from the Markov model, still holds under the new definition.

### B. Parameter Relationships

Now, let's consider the number of collision slots between two successful transmissions and its expectation, denoted by  $n_c$ , and  $E[n_c]$ , respectively.  $E[n_c]$  will be used later in the proposed prediction algorithm and the predictive trigger.

Let  $P_{tr}$  denote the probability that there are at least one transmission in the considered slot time, and  $P_s$  denote the probability that a transmission occurring on the channel is successful. Then the probabilities that a random slot is an idle slot, a slot with successful transmission, and a slot with collision, are  $1 - P_{tr}$ ,  $P_s P_{tr}$ , and  $(1 - P_s) P_{tr}$ , respectively. Hence, the expectations of  $n_c$  can be represented as follows:

$$E[n_c] = \frac{(1 - P_s) P_{tr}}{P_s P_{tr}} = \frac{1}{P_s} - 1. \quad (3)$$

Moreover, from their definitions,  $P_{tr}$  and  $P_s$  can be represented in terms of  $n$ ,  $\tau$ , and station collision probability  $p$ , as follows:

$$P_{tr} = 1 - (1 - \tau)^n, \quad (4)$$

$$P_s = \frac{n\tau(1-\tau)^{n-1}}{P_{tr}} = \frac{n\tau(1-\tau)^{n-1}}{1 - (1-\tau)^n}. \quad (5)$$

By combining equation (2)-(5) and after some manipulations, parameters  $p$ ,  $n$ , and  $\tau$  satisfy the below equation:

$$1 - p - \frac{1}{1 - \tau + n\tau(E[n_c] + 1)} = 0. \quad (6)$$

Furthermore, from equation (2),  $n$  can be represented by  $p$  and  $\tau$  as follows:

$$n = 1 + \frac{\ln(1-p)}{\ln(1-\tau)}. \quad (7)$$

Equations (1), (6) and (7) show that, once we have  $E[n_c]$ ,  $p$ ,  $n$ , and  $\tau$  can be solved by these three equations. This leads to the below prediction algorithm.

### C. Prediction Algorithm

Define function  $f(p)$  as below:

$$f(p) = 1 - p - \frac{1}{1 - \tau + n\tau(E[n_c] + 1)}. \quad (8)$$

where  $\tau$  is given in terms of  $p$  as in equation (1), and  $n$  is given in terms of  $p$  and  $\tau$  as in equation (7).

**Prediction algorithm:**

Choose an arbitrarily small number  $\varepsilon > 0$ . The prediction algorithm for station collision probability is as follows:

Step 1: Upon each successful transmission, collect  $n_c$  and update its average as  $E[n_c]$ .

Step 2: Find  $p \in [0, 1 - \varepsilon]$  such that  $f(p) = 0$  and update  $P_{\text{prediction}} = p$ .

Step 3: Go back to step 1.

**Lemma 1:** for  $E[n_c] \geq 0$  and an arbitrarily small number  $\varepsilon > 0$ , there exists a  $p \in [0, 1 - \varepsilon]$  such that  $f(p) = 0$ .

This lemma can be proved by showing that:

- $f(p)$  is continuous and decreases monotonically with  $p$ ;
- $f(0) \geq 0$ ;
- $f(1 - \varepsilon) < 0$ .

For a), it is straightforward to show that  $f(p)$  is continuous, as below: equation (1) can be alternatively written as below:

$$\tau = \frac{2}{W + 1 + pW \sum_{i=0}^{m-1} (2p)^i} \quad (9)$$

Hence  $\tau$  decreases with  $p$ ,  $\tau \in [\frac{2}{2^m W - 1}, \frac{2}{W + 1}]$ , and  $\tau \leq 1$ . Equation (7) shows that  $n \geq 1$  and by definition,  $E[n_c] + 1 \geq 1$ . Hence  $f(p)$  is continuous by its definition in equation (8).

Monotone can be shown by considering equation (1), (7) and (8) together.

For b), by plugging  $p = 0$  into (1), (7) and (8),  $f(p)$  becomes:

$$f(0) = 1 - \frac{1}{1 + \frac{2}{W + 1} E[n_c]} \quad (10)$$

Since  $W > 0$  and  $E[n_c] \geq 0$ ,  $f(0) \geq 0$ .

For c), since  $\tau \in [\frac{2}{2^m W - 1}, \frac{2}{W + 1}]$  as shown previously, from equation (7),

$$n = 1 + \frac{\ln(1 - p)}{\ln(1 - \tau)} \leq 1 + \frac{\ln(1 - p)}{\ln(1 - \frac{2}{2^m W - 1})}. \quad (11)$$

Hence from equation (8),

$$\begin{aligned} f(1 - \varepsilon) &\leq 1 - (1 - \varepsilon) - \frac{1}{1 - \frac{2}{W + 1} + n \frac{2}{W + 1} (E[n_c] + 1)} \\ &\leq \varepsilon - \frac{1}{1 - \frac{2}{W + 1} + \left(1 + \frac{\ln(1 - p)}{\ln(1 - \frac{2}{2^m W - 1})}\right) \frac{2}{W + 1} (E[n_c] + 1)} \\ &\leq \varepsilon - \frac{1}{c_0 - c_1 \ln \varepsilon} \end{aligned}$$

where  $c_0$  and  $c_1$  are constants and  $c_0 = 2 - \frac{2E[n_c]}{W + 1}$ ,

$$c_1 = -\frac{E[n_c] + 1}{\ln(1 - \frac{2}{2^m W - 1})} \frac{2}{W + 1} \geq 0.$$

Further processing the above equation shows that

$$f(1 - \varepsilon) \leq \frac{c_0 \varepsilon - c_1 \varepsilon \ln \varepsilon - 1}{c_0 - c_1 \ln \varepsilon} \leq 0, \quad (12)$$

where the last inequality holds since the numerator is less than 0 for arbitrarily small  $\varepsilon$ , and the denominator comes from  $1 - \frac{2}{W + 1} + n \frac{2}{W + 1} (E[n_c] + 1)$ , which is greater than 0 as  $n \geq 1$  and  $E[n_c] + 1 \geq 1$  as shown previously.

**Lemma 2:** The computation complexity of the proposed algorithm is upper bounded by  $\log_2 \frac{1}{\delta}$ , where  $\delta$  is the given tolerance.

The computation complexity of the algorithm depends on the number of iterations needed for solving  $f(p) = 0$  in step 2. If the bisection method is employed, the complexity is  $\log_2 \frac{1}{\delta}$ . Hence the computation complexity of the estimation algorithm is upper bounded by  $\log_2 \frac{1}{\delta}$ .

Note that this lemma says that if the tolerance is  $10^{-2}$ , then at most 7 iterations are enough. This shows the low computation complexity of the proposed algorithm.

### III. SIMULATION RESULTS

The proposed algorithm is simulated by OPNET. The scenario simulated is shown in Figure 4, where station 1 to 9 has saturated traffic with destination being station 0, and the physical channel is set to be perfect to exclude the effects of channel losses. Station 1 to 9 start packet transmissions at a random time between 0s and 5s after the start of the simulation, and prediction of station collision probability starts at 5s.

For station 5, Figure 5 plots the comparison between predicted station collision probability and actual percentage of packets that incurred collisions, where the blue curve is

the actual percentage and the red curve is the predicted station collision probability. The transient period of the blue curve at the beginning of the simulation is due to the random start time of all stations' traffic.

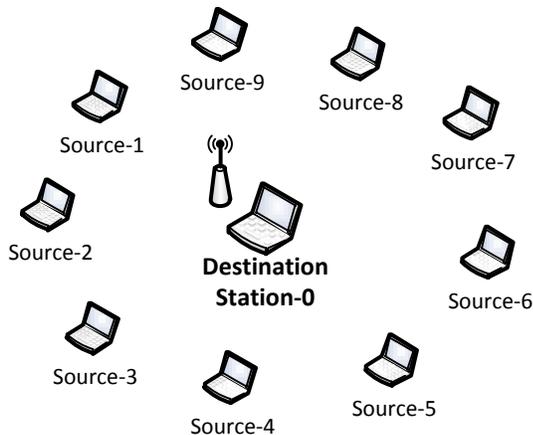
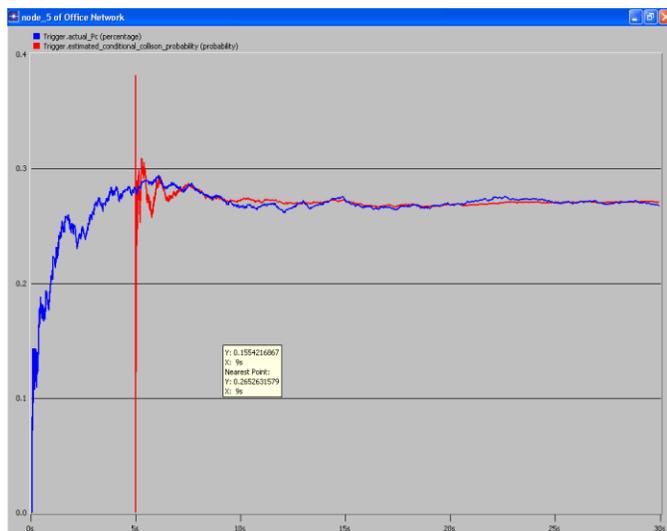


Figure 4. OPNET simulation scenario with ten stations.



Red curve - predicted station collision probability  
 Blue curve - actual percentage of packets from station 5 that incurred collisions.

Figure 5. Comparison between actual percentage of collisions and predicted station collision probability.

For other stations, the comparison between predicted station collision probability and actual percentage of collided packets is similar, which is not shown here due to limited space. Figure 5 clearly shows that the estimated station collision probability converges to the actual percentage of collisions very fast. In addition, the prediction starts during the transient period, and the predicted station collision probability reaches its "steady state" value faster than the actual percentage of collided packets. Hence this algorithm indeed predicts future station collision probability and adapts to the change of network traffic well.

Detail data analysis shows that the estimation errors differ from station to station. The largest estimation error is

7.5%, and the smallest is 1.2%, which validates the accuracy of the proposed prediction algorithm.

#### IV. CONCLUSION

For wireless stations, it is key to detect and react rapidly to link condition changes as they directly affect the station connectivity and application performance.

In this paper, in addition to the commonly used signal quality based triggers, we proposed a novel smart predictive handover trigger algorithm based on mobile station collision probability once an issue has been detected with the current network. A prediction algorithm is developed to predict station collision probability, which does not require the station to send any traffic, has low computation complexity, and applies to real time decisions. Simulation results show that the predicted value matches well with the actual value. The predicted station collision probability hence provides the basis for a predictive handover trigger that based on not only signal quality, but also potential collisions one station may experience, which captures the actual performance the station may expect.

As a next step we are integrating the described station collision probability with RSS to investigate how to select the most appropriate ones under various conditions and the resulting handover performance for WiFi wireless station.

#### REFERENCES

- [1] IEEE P802.21/D01.00, March 2006, "Draft IEEE Standard for Local and Metropolitan Area Networks: Media Independent Handover Services".
- [2] Gregory P. Pollhi, "Trends in Handover Design". IEEE Communications Magazine, pp. 82-90, March 1996.
- [3] Alexe E. Leu and Brian L. Mark, "A Discrete-Time Approach to Analyze Hard Handoff Performance in Cellular Networks". IEEE Transactions on Wireless Communications, pp. 1721-1733, Sept. 2004.
- [4] R. Vijayan and J. M. Holtzman, "A model for analyzing handoff algorithms," IEEE Trans. Veh. Technol., vol. 42, pp. 351-356, Aug. 1993.
- [5] N. Zhang and J. M. Holtzman, "Analysis of handoff algorithms using both absolute and relative measurement," in Proc. Veh. Technol. Conf., 1994, pp. 82-86.
- [6] M. Zonoozi, P. Dassanayake, M. Faulkner, "Optimum Hysteresis Level, Signal Averaging Time and Handoff Delay". IEEE 47th VTC, May 1997, USA, pp. 310-313.
- [7] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE J. Select. Areas Commun., vol. 18, no. 3, pp. 535-547, 2000
- [8] IEEE P802.11, Jun. 2007, "IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications".
- [9] Huaiyu Liu, Christian Maciocco, Vijay Kesavan, and Andy Low, "Energy Efficient Network Selection and Seamless Handovers in Mixed Networks". IEEE World of Wireless Mobile and Multimedia Networks, 2009
- [10] K.N. Choong, Vijay Kesavan, S.L. Ng, Francisco de Carvalho, Andy Low and Christian Maciocco, "SIP-based IEEE 802.21 Media Independent Handover - A BT Intel Collaboration". IEEE Springer, BT Technical Journal Vol 25 N0 2, pp. 219-230, April 2007.

## Exploiting Movement synchronization to Increase End-to-end file Sharing efficiency for Delay Sensitive Streams in Vehicular P2P Devices

Constandinos X. Mavromoustakis

Department of Computer Science,  
University of Nicosia  
46 Makedonitissas Avenue, P.O.Box 24005  
1700 Nicosia, Cyprus  
mavromoustakis.c@unic.ac.cy

**Abstract**—A significant aspect of wireless systems is the intermittent-connectivity experienced by nodes, where sudden network partitioning problems rarely allow a connected path between a source node and its destination. Replication of any requested object and redundancy face the requests' failures whereas they create severe duplications and aggravate the capacity of the end-to-end path. This work quantifies the parameters that affect the end-to-end efficient transmission by taking into consideration the synchronization between moving peers in order to assign the requested resources in the end-to-end path. Synchronization and assignment of the moving Mobile Infostation (MI) peer to a certain vehicle is done with the introduced Message Ferry (MF) mobile Peer in a unidirectional way. A resource assignment cooperation engine is being developed with respect to the cooperation model and end-to-end capacity using passive message ferries in order to efficiently enable delay sensitive streaming. Simulation results have shown that the scheme offers high throughput and reliability and a robust solution for sharing resources of any capacity in dynamically changing mobile peer-to-peer wireless environments.

*Keywords*-synchronized mobility scheme; partially synchronized mobility scheme; file sharing scheme; end-to-end efficiency; evaluation through simulation.

### I. INTRODUCTION

The development of new applications on-the-move, demands the exploration of new dynamically adjusting approaches that enable reliability in an end-to-end manner. Many constraints exist in such networks like resource availability whereas the topological scheme followed in these infrastructures should be combined with the availability of the requested resources and the time-access for sharing resources of each device with the synchronized motion within a specified time duration  $t$  as in real time Vehicle-to-Vehicle (V2V) communication, in a cooperative manner. A Vehicular Ad-Hoc Network (VANET) is a technology that uses moving cars as devices/nodes in a dynamically changing network to establish a mobile network connectivity. In this paper a reliable file sharing scheme for vehicular Mobile Peer-to-Peer (MP2P) devices is proposed taking the advantages of moving devices within a specified roadmap with different pathways like in real time vehicular networks. This work exploits the movements of the devices

and the passive device synchronization to increase end-to-end file sharing efficiency through vehicular users and mobile Infostations [1]. Through geographical roadmaps landscapes where mobile Infostations are set and initialized, the passive synchronization enables through the replication policy to create a replicated object in order to enable reliable file sharing. Role-based Mobile Infostations (MIs) are selected based on their velocity, residual energy, remaining capacity etc and are assigned according to the passive Message Ferry peer. This scheme proved its scalability in node's density since it does not require the knowledge of network at any single host. Additionally it does not require spatial distributions to efficiently spread information while enables reliability in supported mobility without the scheduled 'rendezvous', whereas it effectively passes the requested replicas to designated users.

The organization of the paper is as follows: Section II discusses the related work that has been done on similar schemes which use similar approaches for establishing and maintaining end-to-end file sharing efficiency. Section III then introduces the proposed model on the wireless mobility with the exploitation of passive movement synchronization to increase end-to-end file sharing reliability and a stochastic measure to estimate the end-to-end capacity within the path where the requested replicas were created. Section IV shows the experimental performance evaluation of the proposed scheme and the comparisons done under different convergent parameterized conditions. Particular focus was paid on the impact of certain movements made by Vehicular-Peer-to-Peer (VP2P) devices where multi-client applications dynamically demand resources directly from certain nodal vehicles. For this consideration a stochastic model is introduced for the end-to-end capacity measurements and the dynamic caching activity of the requested objects onto opportunistic neighboring devices.

### II. RECENT SCHEMES AND WORK DONE

Mobility in autonomic communication is an essential parameter and along with the user's demands they pose the vision of what self-behaving flexibility should encompass in next-generation self-tuning behavior [1]. The capacity of the nodes which are traversed in the requested path, can be reduced significantly particularly if we are dealing with

delay sensitive traffic or bursty traffic [1] whereas the underlying end-to-end supporting mechanism should be aware of the dynamic movements in a Peer-to-Peer manner. Obviously, if the transmission-range of a node increases, then the interference it causes will increase and probably the number of nodes which will have copy/copies of the packets that should be forwarded, will increase. Toumpis and Goldsmith [2] define and study capacity regions for wireless ad hoc networks with an arbitrary number of nodes and topology. These regions describe the set of achievable rate combinations between all source-destination pairs in the network under various transmission strategies for EC content sharing and power control. In this work we consider the capacity but in an end-to-end path-request manner and take into consideration the variations caused by the dynamic movements of the devices/vehicles. Most existing architectures (including Grace [3], Widens [4], MobileMan [5]) rely on local information and local devices' views, without considering the global networking context or views which may be very useful for wireless networks in optimizing load balancing, routing, energy management, and even some self-behaving properties like self-organization. This work's contribution is that it associates the synchronized movements and connectivity aspects among vehicles as well as the connectivity resistance and synchronization, whereas the proposed and developed scheme increases the end-to-end file sharing efficiency for delay sensitive streams in vehicle MP2P devices. The scheme extends the advantages offered by the Hybrid Mobile Infostation System (HyMIS) architecture proposed by Mavromoustakis and Karatza, in [6], where the Primary Infostation (PI) is not static but can move according to the pathway(s) of the roadmaps. HyMIS adopts the basic concept of pure Infostation system in terms of capacity service node but it avoids flooding the network with unnecessary flow of information. This capacity node plays a role of control storage node as Haas and Small mention in [7]. Taking the advantages of the proxy caching work done by Liu and Xu in [8] this work proposes an exploitation of the mobility characteristics of each user by selecting the MI peer to be dynamically selected according to characteristics such as the residual capacity of the device based on the push-based activities by other nodes. Heavy emphasis of this work has been put on push-based dissemination explored in [9] by Little and Agarwal, and in [10] by Lochert et al, and analytical dissemination through vehicle-to-vehicle propagation proposed by Wu, Fujimoto and Riley [11] as well as on some recent findings on practical systems as in [12], [13] by Lee et al, and Mahajan et al respectively, for pull-based diffusion activities. The scheme is proposing an index-based mechanism which will enable the selection of the MI in a cluster  $L$ . The following section explores the passive synchronized mobility model in the end-to-end path and presents an analytical model for the end-to-end capacity estimations.

### III. SYNCHRONIZED MOBILITY MODEL IN THE END-TO-END PATH AND CAPACITY CONSIDERATIONS

#### A. Communicating scheme in Vehicle-to-Vehicle communications

The interactions with roadside equipment can likewise be characterized fairly accurately, whereas most vehicles are restricted in their range of motion, for example by being constrained to follow a paved highway. Automobile high speed information interexchange access would transform the vehicle's on-board computer to an essential productivity tool, making virtually any web technology-using pure Infostations) available in the car. However a significant aspect of wireless systems is the sudden partitioning of the connectivity, namely intermittent-connectivity experienced by nodes, where sudden network partitioning problems prevent the exchange of any requested information. Requested object replication and replicas redundancy face the requests' failures whereas they create severe duplications and aggravate the capacity of the end-to-end path. Figure 1 shows the proposed VP2P push and pull procedure in a path using the Passive Opportunistic Synchronized Approach (POSA) as follows: We have enabled a HyMIS configuration where the primary Infostation is not static (PI) but can move as the pathway allows called Mobile Infostation (MI). MIs enable recoverability for any requested object in the end-to-end path and it maintains the sharing reliability. As vehicles are moving from one direction to the other the  $i$ -th vehicle (MI) can pull requested resources to  $i-1, i-2, i-3, i-k$ , where  $k$  is the number of peer vehicle in the end-to-end path requesting resource  $R_i$ . Figure 1 also shows the proposed vehicular MP2P *push* and *pull* procedure where the  $i$ -th vehicle is assigned as MI and can pull requested resources to  $i-1, i-2, i-3, i-k$ , whereas the vehicle which the MI follows can then push any of these resources to the  $i+1, i+2, i+MI_k$  vehicle (dash lines denote the push procedure which takes place and solid lines denote the pull procedure). Both procedures take place until the next and preceding MI is reached while  $i$ -th node is sharing resources, respectively. These notations can also be seen in a more clear form in the figure's 2 pseudocode, which shows a single step for the vehicle's MI transition.

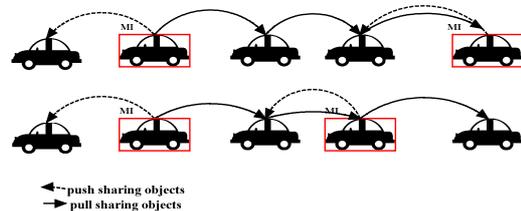


Figure 1. The push and pull configuration for Vehicular MP2P devices while moving in predetermined paths.

```

Set communication Path(A,B, N)
{
  If (MI criteria meet==TRUE)
    Set MI in the Path(A,B, N);
  else
    form Path(A,B, M)  $\forall M_i \in N$ 

  for (MI=i;i<k;i++)  $\forall MI_k \notin MI_i$ 
    pull_requestedObj(Ob_id, Cap, Peers,
    estimated_delay,reputation_degree);

  for (MI=j;j>k;j--)  $\forall MI_k \notin MI_j$ 
    push_requestedObj(Ob_id, Cap, Peers, estimated_delay);
}
    
```

Figure 2. Pseudocode for a single round trip step for the vehicle's MI transition in order to enable object replication placement scheme between synchronized peers.

### B. Multi-hop mobility model and user's capacity in the end-to-end path

Resources availability problems can be also faced using a local summary of the global system-or clustered information of the subsystem-by using the property –well-known in distributed systems, of the generically referred concept as aggregation by Renesse et al in [14]. MP2P systems require to guarantee the availability any requested resources as well as to enforce appropriate access control policies. In our application scenario we assume that a common lookup application is being used in order to enable nodes to interexchange locally the requested information objects. As a starting measure we estimate the synchronized cooperative movements of each vehicle by measuring the motion performed while measuring at the same time the reserved capacity by each vehicle. Since vehicles are moving in an organized and –sometimes- predictable way, the pull and push model aggravates the capacity of each device, as in a MP2P environment. Through the proposed sharing scheme for Vehicle-to-Vehicle communications as well as the additional parameters that are being considered, like the evaluated end-to-end relay epoch/latency, the mobility pattern and the time frame for the allowed promiscuous caching introduced in [15] by Mavromoustakis, the proposed model enables efficient capacity manipulation in the end-to-end relay region and efficient data manipulation in the intercluster communication.

By adopting the modified scheme of Mavromoustakis and Karatza [6] and enable the role of MI to be adjusted into the vehicular devices, the PI and MI are now implemented by a certain frontal vehicle and the connectivity, where only unidirectional sharing and connectivity occurs.

When mobility is considered, the design of efficient rendezvous data dissemination protocols for enabling efficient manipulation and availability of resources is complex, and the existing solutions do not consider the random probabilistic movements of devices while disseminating data. In order to measure the direction of the movement of each device. Each device is associated with a random variable which represents the direction movement. For the movement this work considers a probabilistic Random Walk in a predefined pathway represented as a Graph (G) where this G enables as a random

variable the weights of these random movement. A device can perform random movements according to the topological graph  $G = (V,E)$  where it comprises of a pair of sets V (or  $V(G)$ ) and E (or  $E(G)$ ) called vertices (or nodes) and edges (or arcs), respectively, where the edges join different pairs of vertices. This work considers a connected graph with  $n$  nodes labeled  $\{1, 2, \dots, n\}$  in a cluster  $L^n$  with weight  $w_{ij} \geq 0$  on the edge  $(i, j)$ . If edge  $(i, j)$  does not exist, we set  $w_{ij} = 0$ . Each node moves from its current location to a new location by randomly (probabilistically) choosing an arbitrary direction and speed from a given range. Such a move is performed either for a constant time for a constant distance traveled. Then new speed and direction are chosen. In the probabilistic Mobility model is described as a memoryless mobility pattern because it retains no knowledge concerning its past locations and speed values. In this work a Probabilistic Version of the Random Walk Mobility Model is used as in [21] by Ibe. In this model the last step made by the random walk influences the next one based on the stationarity and the correlations between the movements. Under the condition that a node has moved to the right the probability that it continues to move in this direction is then higher than to stop movement. This leads to a walk that leaves the starting point much faster than the original random walk model. Given that the device/vehicle is currently at node  $i$ , the next node  $j$  is chosen from among the neighbors of  $i$  with probability:

$$p_{ij}^L = \frac{w_{ij}}{\sum_k w_{ik}} \quad (1)$$

where in (1) above the  $p_{ij}$  is proportional to the weight of the edge  $(i, j)$ .

### C. Cooperation model and end-to-end capacity using passive message ferries for delay sensitive streaming

In order to define which requested objects should be outsourced onto preceding m-peers a ranking model has been applied as follows: To find the rank of an object  $a_1 a_2 \dots a_m$ , one should find the number of objects preceding it. It can be found by the following function:

function rank( $a_1, a_2, \dots, a_m$ )

rank ← 1 ;

**for**  $i \leftarrow 1$  **to**  $m$  **do**

**for** each  $k < a_i$

rank ← rank +  $N(a_1, a_2, \dots, a_{i-1}, k)$

Then the new sequence will cache onto other nodes in the path the first  $k$ -requested objects where  $k$  is defined as a function of the remaining capacity onto each device as:

$$|k| = \inf\left(\frac{\sum_{N=i}^N (1 - \rho_N)}{N}\right) \quad (1.1)$$

where  $\rho_N$  is the utilized capacity and  $N$  is the number of hops in the end-to-end path. Nodes in the path are moving according to the 2-D plane mobility model  $L \subset \Lambda, \Lambda \subset \mathcal{R}^2$ . A moving square (the  $\{\Lambda_1, \Lambda_2, \Lambda_3, \dots\}$  bounded area) is

divided into multiple sub-squares, called cells as in [1], and time is divided into slots of equal duration. At each time slot a node is in and can be only in one cell. The initial position of a node is uniformly chosen from all cells. At the beginning of each time slot, the node jumps from its current cell to one of its adjacent cells with equal probability. Two mobile nodes can communicate with each other whenever they are within a distance of  $d$ , the transmission range of the mobile node. In order not to have an optimistic assumption a low density population network is assumed with regards to the number of traversing nodes per  $\Lambda_i$ . We assume that no conspiracy policy exists where nodes somehow conspire together not to meet each other forever and move at  $d > D$  and in parallel.

The *index* of each node is being transferred using the message ferries that are passively passing from any other pathway within the distance of communication range of each device. Figure 3 shows this approach where the message ferries are crossing any other pathway and at the same time being in the distance transmission range of each device that they pass from.

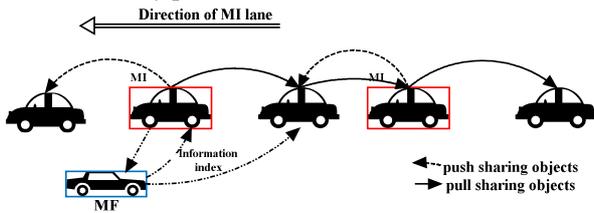


Figure 3. Passive message ferries where any other device can play the role of the messenger regarding the information index.

Taking into account the delay characteristics, let  $N$  be the number of source peers in the network ( $N$  different end-to-end paths) and  $C_i(t)$  be the service capacity of source peer  $i$  at time slot  $t$ . An end-to-end download can be then depicted as a function of time as derived from Chiu and Young Eun in [16] and the  $w_{ij}^L$  of the end-to-end path in the cluster  $L$  as:

$$T = \min \left\{ s_{ij} > 0 \mid \sum_{t=1}^s C(t) \geq F \right\} \quad (2)$$

where  $F$  is the file capacity defined as  $\{f_1, f_2, f_3, f_4, \dots, f_n\}$  equi-divided file chunks and  $s$  a given end-to-end bounded allowed delay for this file to be downloaded from any numbers of peers in the end-to-end path. The obtained eq. (2) derived from Wald's equation introduced by Ross in [17] can therefore be expressed as:

$$F = E \left\{ \sum_{t=1}^T C^L(t) \right\} = E \{ C^L(t) \} E \{ T \} \quad (2.1)$$

where we can easily extract the slotted amount of file chunks that are shared in the end-to-end path. The  $A(\bar{c})$  is the minimum average capacity offered by each link in the path as:

$$A(\bar{c}) = \frac{1}{N} \sum_1^N \inf(C_{ij}(t)) \quad (3)$$

where  $A(\bar{c})$  is the requested and available arithmetic mean for the capacity in the path. The average capacity offered by the end-to-end path considering all the links in the path of the requested file  $F$ , can be denoted as  $A(\bar{c}) / E \{ C_{ij}^L(t) \}$ , the average download time is:

$$E \{ T_{F_{ij}} \} = \frac{F_c}{A(\bar{c})} = \frac{N \cdot F_c}{\sum_1^N C_{ij}(t)} \quad \forall w_{ij} \in L \quad (4)$$

while it stands that for  $C_{ij}(t) = \min(\inf(C_{ij}(t)) | E \{ T_{F_{ij}} \})$ .

Let  $t_\lambda = \max(\Theta_{MI,j})$  be the contact rate estimation and  $\Theta_{MI,j}$  is the estimated contact time between MI and a moving node  $j$ , then it stands that a vehicle remains as a MI in the path if the following is satisfied:

$$t_{\lambda_{ij}} \geq \frac{A(\bar{c})}{BW_{ij}} \quad \text{where } t_{\lambda_{ij}} \text{ is the contact rate in the path}$$

between  $i,j$  and  $BW_{ij}$  is the associated bandwidth in the path between  $i,j$ . The estimation of  $t_{\lambda_{ij}}$  is essential since it can determine the time that a mobile node can remain as a MI.

#### D. Considering contact interactions for collaborative streaming

In this section we propose a number of social interaction parameters which take place in collaboration with the file chunk outsourcing of the previous section. The metrics are community-oriented and are considering the number of created clusters  $C_N(t)$  in a specified Relay region of a certain transmitter—and a number of receivers (1, N] under the relay node pair ( $u,w$  |MI $_i$ )-as a modified definition of [18]- as follows the:

$$C_N(t) = \frac{2|h_N(t)|}{|I_{C(N)}(t)| \cdot (|I_{C(N)}(t)| - 1)}, \text{ iff } P_{u \rightarrow w \rightarrow (x,y)} > W_N(t) \quad (5)$$

where  $W$  is the Community streaming factor and is defined as the number of existing communities in the intercluster communicational links at a given time instant. The  $h_N(t)$  is the number of hops in the existing clusters and the  $I_{C(N)}(t)$  is the number of interconnected nodes  $N$  in the cluster  $C_N(t)$ .  $W$  can be defined according to the download frequency of the file chunks in the intercommunity as follows:

$$W_N(t) = \frac{DldRate \cdot \# \text{ sharing Chunks}}{\text{Total} \# \text{ dlds}(t) \cdot \# \text{ inactive Chunks}} \quad (5.1)$$

where in (1.4) the download rate is considered in contrast

with the number of chunks being shared in a specified instant time  $t$ .

IV. PERFORMANCE EVALUATION, EXPERIMENTAL RESULTS AND DISCUSSION

A. Dedicated Short Range Communications (DSRC)

To emulate the scenario described earlier, the need of a possible realistic environment must be achieved. DSRC was used for the evaluation of the proposed scenario which is two-way short- to medium-range wireless communication channels specifically designed for automotive use and utilizes a corresponding set of protocols and standards [19]. Considered to be short to medium range communication technology it operates in the 5.9 GHz range. The Standards Committee E17.51 endorses a variation of the IEEE 802.11a MAC for the DSRC link. DSRC supports vehicle speed up to 120 mph, nominal transmission range of 300m (up to 1000 m), and default data rate of 6 Mbps (up to 27 Mbps). This will enable operations related to the improvement of traffic flow, highway safety, and other Intelligent Transport System (ITS) applications in a variety of application environments called DSRC/WAVE (Wireless Access in a Vehicular Environment). In the evaluation of the proposed scheme we evaluated the Peer-to-Peer/Ad hoc mode (vehicle-vehicle) scenario and took into account the signal strength parameters and the minimized ping delays between the nodes in the end-to-end path according to the  $d_p = \text{Min} \sum_{i=1}^n D_i$ , where  $D$  is

the delay from a node  $i$  to node  $j$ , and  $d_p$  is the minimized evaluated delay in the end-to-end available path. Moreover, considering the need of bandwidth for the wireless devices, it is necessary to apply efficient routing algorithms to create, maintain and repair paths, with least possible overhead production. The proposed scenario uses the Zone Routing Protocol (ZRP) [20]. The number of nodes varies depending on the mobility degree and the distance variations of each user within a connectivity scope. The user's transition probability arises from a specified location where certain information is pending to be received by this user.

B. Simulation results of the proposed scenario and discussion

In this section, we present the results extracted after conducting the discrete time performance evaluation through simulation of the proposed scenario. The simulation used a two-dimensional network, consisting of 100 nodes dynamically changing the topology on a non-periodic basis (asynchronously as real time mobile users do). For each node, it stands that after random time each node moves at a random walk to one of the possible destinations (north, east, west, south) in an organized vehicular way. Each link (frequency channel) has max speed reaching  $10\text{Mb per sec}$ . The propagation path loss is the two-ray model without fading. The network traffic is modeled by generating constant bit rate (CBR) flows. Each source node transmits one 512-bytes (~4Kbits) packet. Packets generated at every time step by following Pareto distribution, destined for a random destination uniformly selected. Nodes have at

any time measures of the information destined for each node (for a given time interval) sent by any node.

Figure 4 shows the network dimensions with the data and capacity exchanged through the created clusters. Figure 4 shows that even when the files that are being exchanged are greater than the network dimensions, the proposed scheme effectively handles the end-to-end transmissions and enables the complete download whereas for this evaluation two measures were taken into consideration: the data exchanged within the cluster  $i$  and the data exchanged with other clusters.

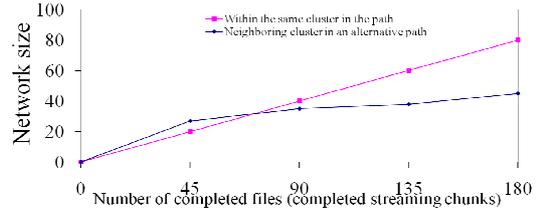


Figure 4. Dimensions of the network with the data and capacity exchanged through the clusters formed.

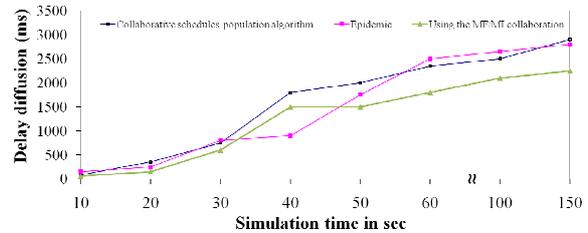


Figure 5. The delay of the diffusion outsourcing process with the simulation time compared with Epidemic and collaborative schedules schemes.

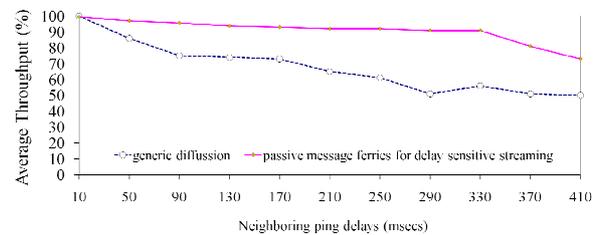


Figure 6. The average throughput with the neighboring ping delays.

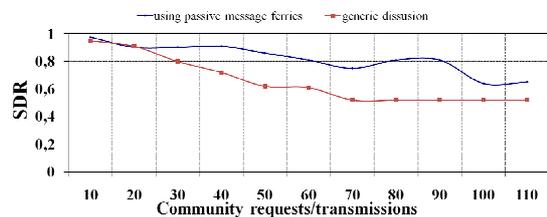


Figure 7. The SDR with the community/cluster requests using passive message ferries and the generic diffusion methods.

Figure 5 shows the delay of the diffusion outsourcing process with the simulation time compared with two different in implementation schemes: the epidemic and collaborative schedules schemes. It is easily spotted that fig. 8 shows the supremacy of the proposed scheme for this specific scenario in vehicular P2P systems whereas it shows

the effectiveness with the significant robustness in the delay diffusion process-which is further minimized.

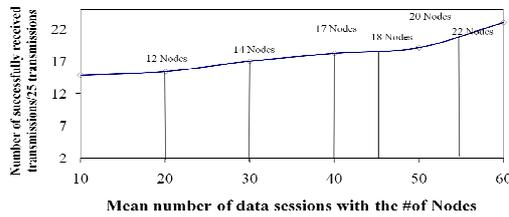


Figure 8. The number of successfully received transmissions over of total /25 transmissions with the mean number of sessions created in the system.

Figure 6 shows the average throughput with the neighboring ping delays comparing the proposed passive message ferries scheme and the generic diffusion cluster scheme. Figure 7 shows the SDR with the community/cluster requests using passive message ferries and the generic diffusion method, whereas in figure 8 the number of

V. CONCLUSION AND FURTHER RESEARCH

In this work, we have proposed a resource assignment approach while synchronized in-motion nodes are exchanging resources with bounded end-to-end delay. The method encompasses the assignment of the moving-so called-Mobile Infostation (MI) peer to a certain vehicle whereas this is done with the introduced Message Ferry (MF) mobile Peer in a unidirectional way. Passive message ferries are utilized as a resource index for the end-to-end path in order to efficiently enable delay sensitive streaming. Simulation results have shown that the scheme offers high throughput and significant end-to-end reliable exchange of resources whereas it offers high SDR for completed files.

Current and future research directions include the modeling of the mobility pattern of the peers by using approaches like the fractional Brownian motion taking into account the global requests and different network partitioning parameters as well as evaluating an extended version of the proposed scheme in real time.

REFERENCES

[1] C. Mavromoustakis and H. Karatza, "A Gossip-based optimistic replication for efficient delay-sensitive streaming using an interactive middleware support system", *IEEE Systems Journal*, IEEE USA, Vol. 4, no. 2, June 2010, pp. 253-264.

[2] S. Toupmpis and A. Goldsmith, "Capacity regions for wireless ad hoc networks", *IEEE Transactions on Wireless Communications*, Vol. 2, No. 4, July 2003, pp. 736-748.

[3] Daniel G.Sachs et al., "GRACE: A Hierarchical Adaptation Framework for Saving Energy", *Computer Science*, University of Illinois Technical Report UIUCDCS-R-2004-2409, February 2004.

[4] Dzmityr Kliazovich and Fabrizio Granelli, "A Cross-layer Scheme for TCP Performance Improvement in Wireless LANs", *Globecom 2004*, IEEE Communications Society, pp. 841-844.

[5] Marco Conti, Gaia Maselli, Giovanni Turi, and Sylvia Giordano "Cross layering in mobile Ad Hoc Network Design", *IEEE Computer Society*, February 2004, pp. 48-51.

[6] C. Mavromoustakis and H. Karatza, "Segmented File Sharing with Recursive Epidemic Placement Policy for Reliability in Mobile Peer-to-Peer Devices". *Proceedings of the 13th Annual Meeting of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*

successfully received transmissions over of total of 25 transmissions in the path/clustered end-to-end transmission is shown, with the mean number of sessions created in the system. Finally, figure 9 depicts that neighboring feedback can enable better streaming stability in a multistreaming end-to-end path.

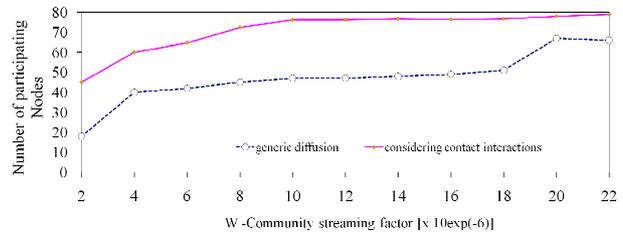


Figure 9. Number of participating nodes with the W.

(MASCOTS), Georgia Tech, Atlanta, Georgia, September 26-29, 2005, pp 371-380.

[7] Z. Haas and T. Small, "Evaluating the Capacity of Resource-Constrained DTNs," *International Wireless Communications and Mobile Computing Conference*, pp. 545-550, July 2006.

[8] J. Liu and J. Xu, "Proxy caching for media streaming over the Internet," *IEEE Commun. Mag.*, vol. 42, no. 8, pp. 88-94, Aug. 2004.

[9] T.D.C. Little and A. Agarwal, "An Information Propagation Scheme for VANETs," *Proc. 8th Intl. IEEE Conf. on Intelligent Transportation Systems (ITSC)*, Vienna Austria, Sept. 2005, pp. 361-368.

[10] C. Lochert, B. Scheuermann, M. Caliskan and M. Mauve, The feasibility of information dissemination in vehicular ad-hoc networks. in *Proceedings of the 4th Annual Conference on Wireless On-demand Network Systems and Services* (Jan. 2007), pp. 92-99.

[11] H. Wu, R. Fujimoto, and G. Riley, Analytical models for information propagation in vehicle-to-vehicle networks. *IEEE Vehicular Technology Conference*, 2004. VTC2004-Fall. 2004, Sept. 2004, Vol. 6, pp. 4548-4552.

[12] Lee, K., et al. First experience with cartorrent in a real vehicular ad hoc network testbed. In *Proc. of MOVE (2007)*.

[13] Mahajan, R., Zahorjan, J., Zill, B. Understanding wifi-based connectivity from moving vehicle. In *Proc. of IMC (2007)*.

[14] R. Van Renesse, K.P. Birman, W. Vogels, Astrolabe: a robust and scalable technology for distributed system monitoring, management, and data mining, *ACM Trans. Comput. Sys.* 21 (2).

[15] C. Mavromoustakis "Synchronized Cooperative Schedules for collaborative resource availability using population-based algorithm", submitted to the *Simulation Practice and Theory (SIMPRA) Journal*, Elsevier, Volume 19, Issue 2, February 2011, pp. 762-776.

[16] Yuh-Ming Chiu and Do Young Eun, "Minimizing File Download Time in Stochastic Peer-to-Peer Networks," *IEEE/ACM Transactions on Networking*, Vol. 16, No. 2, pages 253-266, April 2008.

[17] S. M. Ross, *Stochastic Processes*, 2nd ed. New York: John Wiley & Son, 1996.

[18] M. Fiore and J. Harri. "The networking shape of vehicular mobility". In *Proceedings ACM MobiHoc*, pp. 261-272, 2008.

[19] Dedicated short-range communications (DSRC) for wireless communication channels found on [http://en.wikipedia.org/wiki/Dedicated\\_short-range\\_communications](http://en.wikipedia.org/wiki/Dedicated_short-range_communications), last accessed on 18 Oct. 2010.

[20] Z. Haas and M. Pearlman, "The Zone Routing Protocol (ZRP) for Ad Hoc Networks", *Internet Draft*, draft-ietf-manet-zone-zrp-02.txt, June 1999.

[21] Oliver Ibe, *Markov Processes for Stochastic Modeling*, ISBN-10: 0123744512, Academic Press (September 16, 2008), pp.512.

## Vertical Fast Handoff in Integrated WLAN and UMTS Networks

Vahid Solouk<sup>1</sup>, Borhanuddin Mohd Ali<sup>2</sup>, and K. Daniel Wong<sup>3</sup>

1. Department of Electrical Engineering, UUT, Urmia, IRAN

e-mail: vsolouk@comsoc.org

2. Department of Computer System and Communication Engineering, UPM, Serdang, MALAYSIA

e-mail: borhan@ieee.org

3. 3G Wireless & Technology Sdn Bhd, Petaling jaya, MALAYSIA

e-mail: daniel\_wong@ieee.org

**Abstract**— Mobility management and on top of that, vertical handoffs remains as one of the most challenging obstacles in 4G evolution. In this paper, we present Vertical Fast Handoff protocol as a solution to the mobility issues in integrated WLAN-UMTS networks which utilizes Early Binding Update technique to achieve reasonable performance. It contains several key factors including new network modules and procedures. In order to evaluate the performance, an analytical model is presented that includes metrics describing handoff and packet delivery delays, and signaling overhead. Based on the assessments, it is shown that the proposed method exhibits tolerable performance in terms of delays as well as signaling overhead.

**Keywords**—vertical handoff; early binding update; packet delivery; latency; signaling cost

### I. INTRODUCTION

The coexistence of various communication systems as recommended for the next generation of mobile systems requires mobility solutions for users with seamless inter-technology roaming capabilities; this means that a seamless inter-system handoff is required. Currently, the noticeable ambition toward the convergence of access technologies foreseen by many research bodies has resulted in several approaches for achieving seamless vertical handoffs. The most noticeable discussions in the literature are currently on whether the integration of two standards namely, Media Independent Handover (MIH) proposed by IEEE802.21 and IETF mobility working group (mipshop) will lead to mobility solutions for future mobile networks. Despite the initial wrap-ups of the mentioned standardization bodies independently, only a promise of minimized data interruption during vertical handoffs is made certain by either the integration of the standards or other solutions or other proposals in the area.

Naturally, every inter-system roaming which leads to vertical handoff requires that both link and IP layer handoffs take place, since both network points of attachment as well as the device interface are subject to change. Several initiatives have been made to finally design and implement each of these communications layers. For vertical handoff in a heterogeneous wireless network, the integration and interworking of these two layers with a properly designed timing can directly impact on the performance parameters and subsequently lead to seamless handoffs. In this paper, we

propose Vertical Fast Handoff (VFHO) as a new method which is applied at both IP and link layers. VFHO utilizes some features of Fast Handoff for Mobile IPv6 (FMIPv6) in a different manner and manipulate the timing in IP layer including Early Binding Update (EBU) with the Correspondent Node (CN). We then present an analytical model to evaluate the performance of our method. The rest of this paper is organized as follows:

The next section reviews some of the efforts made in the field, followed by a full description of the protocol design in Section 3, while Section 4 describes the protocol in further detail. In Section 5, we present an analytical model for the performance metrics including handoff and packet delivery delay, as well as signaling overhead in form of cost functions. Finally, the paper is concluded in Section 6.

### II. HANDOFF IN HETEROGENEOUS NETWORKS

Many handoff protocols promise seamless mobility, focusing mainly on the handoff operation latency, packet loss during the handoff, or similar metrics. However, the issue of seamless mobility becomes more fragile when inter-system or vertical handoff is the case.

IEEE802.21 MIH [1, 2] supports various types of layer-3 mobility management protocols, specifically Mobile IP (MIP), MIPv6 and Session Initiation Protocol (SIP) [3-5]. As this standard focuses mainly on solving media independency problem, it operates closer to link layer than on the mobility management protocols of layer-3. Hence, integration with layer-3 protocols to optimize vertical handoff has been the interest of several proposals [2, 3, 6-11]. For instance, the primitives in MIH to support handoff is far from adequate, hence several works addressing this issue have been proposed (i.e., in [6]) where a new primitive, namely MIH-PrefixInfo including the prospective Access Router (AR) info was linked to L2 events, and based on modified event triggers, a similar mechanism to FMIPv6 for handoff has been proposed. Although this work originally addressed the issue of anticipation and ping-pong effect in FMIPv6, the method for AR discovery was not indicated and neither was information gathering from the neighborhood. Besides, the proposed handoff mechanism results in more deployment complexities in AR.

Access Router Information Protocol (ARIP) [12] is another proposal based on IETF SEAMOBLY working group project [13] defined as Candidate Access Router Discovery

(CARD). The information on neighboring ARs (ARIP) needed for MN is provided at MN's local or home AR and then sent to the MN. However, the protocol suggests no method on how to collect ARIP information from the neighboring ARs and how the procedure should be initialized. Additionally, maintaining such processes for AR requires more network resources and more AR functionalities while the rest of handoff operations still need to be performed by Mobile Node (MN). Moreover, the protocol still requires AR deployment which is a technology obstacle.

Few other proposals focus on improving the IEEE802.21 proposed Media Independent Information Services (MIIS) [3]. MIIS information primitives are utilized in [9, 11]; this is done by selecting a higher layer mechanism of mobility management, which is a SIP-based mechanism, to obtain information of neighboring networks from different access technologies. The method was tested with an MN with two neighboring subnets. These approaches suggest that the MIH information is obtained through several query/response messages to estimate the network.

### III. VFHO DESIGN

VFHO conceptually differs from other handoff methods in the way of service disruption and packet transmission period; this is due to proper interaction of link and IP layers and hence less disruptive mobility and handoff. It includes a procedure to collect and process user and network traffic information from higher layers; this is necessary to select an appropriate network for the next point of attachment. Furthermore, VFHO resolves the issue of packet delivery delay which arises from reroute and retransmissions between the old and new points of attachments. Table 1 lists the new messages and service primitives introduced through the proposal. The message flow diagram of the proposed

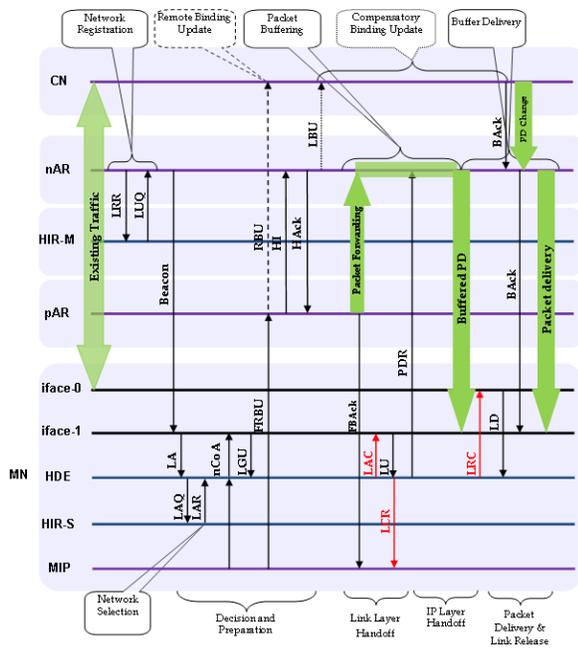


Figure 1. Message Flow Diagram of VFHO

approach is depicted in Fig. 1. The network registration is performed once an AR is switched on and the active ARs send an LUR message to Home Information Register in Master mode (HIR-M) in intervals to preserve their status at HIR-M. If no LUR is received, HIR-M inquires the respective AR using LUQ message and unless it receives a reply from the AR, the status changes to inactive. If the AR does not reply to two consecutive LUQs, the record is deleted from the database. In the following, the process is described through some operational phases which are identified in Figure 1. We assume a mobile user maintaining an ongoing connection with UMTS network approaches an indoor destination with WLAN coverage and switches on the WLAN interface. The base station ID received through beacon is reported to the Handover Decision Engine (HDE) and to HIR-S for information of the discovered AR. After the target network selection, the HIR-S informs HDE using LAR. As the new network is detected, a Link-Going-Up (LGU) event is sent to HDE to start Fast Binding Update (FBU).

#### A. Early Binding Update with CN

Majority of handoff studies, specifically of fast handoff suggest that the Binding Update (BU) procedure be started after the packet delivery from new AR (nAR) to MN. To reduce packet delay, BU should be initiated at the start of packet forwarding. The BU message is formed in nAR using the nCoA and forwarded to the CN's IP. The proposed early BU can be performed in two cases, both prior to link switch. For the first case, the BU message is appended to FBU and sent to previous AR (pAR) which in turn, processes and extracts the BU and sends it to CN. Since the BU in this case is sent through the old network, it is recognized in CN as the Remote BU (RBU). In the second case, it is assumed that the pAR has no signaling message ready to perform BU hence, as soon as a bi-directional tunnel between the two ARs is established, the BU message is forwarded to nAR and thereby, to CN as a Local BU (LBU).

#### B. Link Activation and IP Layer Handoff

Now, HDE can initiate the nCoA activation on WLAN interface and the MIP layer sends a message to nAR to inform that the MN is ready to receive packets. As the main part of IP layer handoff, nCoA has already been configured and validated therefore, the immediate action after the link state changes is to inform the nAR. The event indicated in this process is LU which also triggers HDE to assist issuing the message to nAR. HDE informs the MIP layer using an information message on link switch status which is called Link Change Report (LCR). The packet delivery phase starts immediately after IP layer handoff completion. However, packet forwarding through the old network continues until the CN confirms the BU by sending Binding Acknowledgement (BA). At this time, the HDE commands the release of old link to UMTS interface using Link Release Command (LRC) and the process is completed. While most of handoff methods consider the start of packet

delivery from the buffer as completion point of handoff process, this proposed model strives to satisfy QoS requirements for various traffic classes by decreasing the packet delay due to handoff.

#### IV. PERFORMANCE ANALYSIS

The analytical model to assess the proposed handoff protocol is illustrated in Fig. 2. This model was inspired from [14] and leads to obtain a general cost function to describe the metrics.

Table 1. New Message and Primitive Structure for VFHO

Message/Primitive Name	Service Type	Parameters
Link Available (LA)	Event	MN WLAN MAC, Link Type, nAR MAC, Activity Flag
Local Area Query (LAQ)	Command	pAR ID, nAR ID, Usability Code
Local Area Report (LAR)	Information	MN MAC, nAR MAC, BSS ID, Status, Priority Code
Remote AR Query (RAQ)	Command	pAR ID, nAR ID, nAR Prefix
Remote AR Report (RAR)	Information	nAR Prefix, AR Type, Priority Code, Neighbors (Prefix, AR Type, Priority Code)
Local Registration Request (LRR)	Information	BSS ID, AR Prefix, Link Type, Available BW, Cost of Service, Offering Service Codes, Reg. Flag
Local Update Request (LUR)	Information	BSS ID, AR Prefix, Link Type, Available BW, Cost of Service, Offering Service Codes, Upd. Flag
Local Update Query (LUQ)	Command	Link ID, AR (SGSN) Prefix, AR (WLAN) Prefix, Lifetime
Link Release Command (LRC)	Command	Old Link ID, MN UMTS MAC, pAR MAC, Reason Code
Link Activate Command (LAC)	Command	New Link ID, MN MAC, nAR MAC, Priority Code
Link Change Report (LCR)	Information	nAR MAC, MN WLAN MAC, Result Code

##### A. Handoff Latency

Handoff delay is comprised of several elements including Link and IP layers handoff latencies, and packet delivery delay. It can be concluded that the period of packet buffering and binding update is equal to the time taken to perform link and IP handoffs, and the packet forwarding period. However, unlike the other handoff protocols, the time taking processes are incorporated in VFHO. It can be inferred that the expected handoff latency of VFHO depends highly on transmission periods between MN and nAR as well as process delays in the MN as these processes are scheduled to start early and by special events as described in protocol description. Hence,  $T_{LH}$  is only a portion of the actual link layer handoff time,  $T_{IH}$  excludes nCoA configuration as another long process, and  $T_{BU}$  is a short time as it has started prior to link layer handoff. The total handoff latency therefore, is:

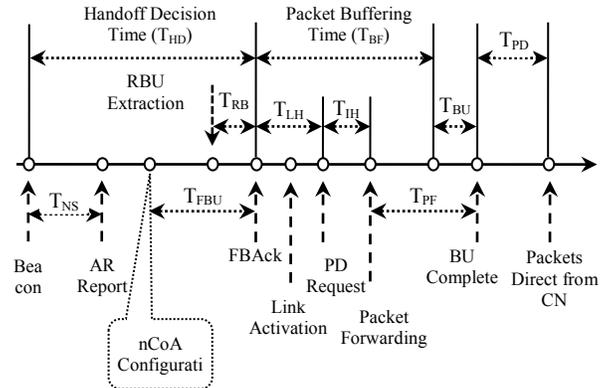


Figure 2. Timing Diagram for Transmissions and Processes of VFHO

$$T_{VFHO} = T_{HD} + T_{BF} + T_{BU} + T_{PD} \quad (1)$$

where,  $T_{BF} + T_{BU} = T_{LH} + T_{IH} + T_{PF}$ . For identical distances, wireless components, TMN-nAR and TMN-pAR can be expressed as TMN-AR. The handoff delay can be written as:

$$T_{VFHO} = T_{LH} + T_{PD} + 2T_{MN-AR} \quad (2)$$

One of the major delay contributors in IP-based handoffs is the procedure of Duplicate Address Detection (DAD) or TDAD which is performed within binding update procedure. The duration of this procedure is topology dependent and is reported in many IP-based network infrastructures to take between 0.5 and 1 s. As the BU procedure starts some time prior to link switch, DAD is given some time to complete before the IP layer handoff. Hence, the overall handoff latency is independent of TDAD. For the sake of calculations, we assume TDAD = 600ms in Equation 2.

The handoff delay for various wireless link delays and MN speeds are shown in Fig. 3-a. The wireless link delay was varied from 10 to 500ms with different steps. Although 10m/s is a high speed for an MN to move, the handoff latency could be maintained as low as 600ms when the wireless link delay reaches 75ms. However, for lower speeds (i.e. up to fast walking speed of 5m/s), the handoff delay is around 300ms and reaches 400ms when wireless link shows a delay of around 130ms.

##### B. Packet Delivery

We analyze packet delivery from two main aspects, the cost of delivering data packets and the cost of signaling. We propose an analytical model similar to what was introduced in [15] to determine the packet delivery cost from data transmission aspect which is used to obtain the end-to-end latency during the total handoff process. The packet delivery cost consists of two main elements namely, transmission and process costs. We assume  $\alpha$  and  $\beta$  as normalized weighting factors that influence the two cost elements of packet transmission and processing. Hence, the packet

delivery cost and the ratio of average size for data and signaling packets can be obtained as:

$$C_{PD}^D = \alpha C_T + \beta C_P \quad (3)$$

$$\phi = \frac{\vartheta_P}{\vartheta_P + \vartheta_S} \quad (4)$$

where  $\vartheta_P$  and  $\vartheta_S$  denote the data and signaling packet sizes, respectively. The transmission cost of data packets is a portion of total transmission cost by the coefficient  $\phi$  and can be written as:

$$C_T^P = \phi C_T \quad (5)$$

$$C_T = \lambda_p (T_{BU} + T_{BF}) \quad (6)$$

$$C_T^P = \phi \cdot \lambda_p (T_{BU} + T_{BF}) \quad (7)$$

where,  $\lambda_p$  is packet arrival rate (number of packets per time unit) and  $\phi$  is the ratio of data packets to the overall data.

Fig. 3-b illustrates packet delivery delay versus data packet size when packet arrival rate changes. When the data packets form the maximum of 50% of the total packets, it can be seen that the packet delivery cost shows small variations with a maximum of 38 at the rate of 25 packets per second. As data packets increase to above 70%, the cost becomes more sensitive to the arrival rate showing variations of about 35 to 50. Although this shows that the packet delivery is highly dependent on the size and arrival rate of the data packets, even the highest delivery cost hardly causes packet disruption as it is still comparable to the overall signaling cost of around 300 (discussed in next section).

### C. Signaling Cost

Signaling cost is defined as the total cost of signaling traffic overhead which in turn, is the total number of control messages exchanged between MN and network components (AR or CN). To determine total signaling cost, the main cost equation is extended to signaling costs for the four stages of Handoff Decision, Link Layer Handoff, IP Layer Handoff, and Packet Delivery, and can be expressed as:

$$C_{Sig} = \sum_i C_i = C_{HD} + C_{LH} + C_{IH} + C_{PD} \quad (8)$$

Table 2. Parameters for Analytical Model

$\alpha$	B	$D_r$	$D_l$	$\lambda_s$	w
0.1	0.2	6	4	1	2

We use random-walk mobility model which is generally confined to a limited geographical area and speeds [16]. For random movements over a certain period, the probabilities of the user leaving and staying in a local area are  $p$  and  $q = 1-p$ , respectively. The user position is defined as state  $k$  Markov chain. The two transition probabilities  $\alpha_{k,k+1}$  and

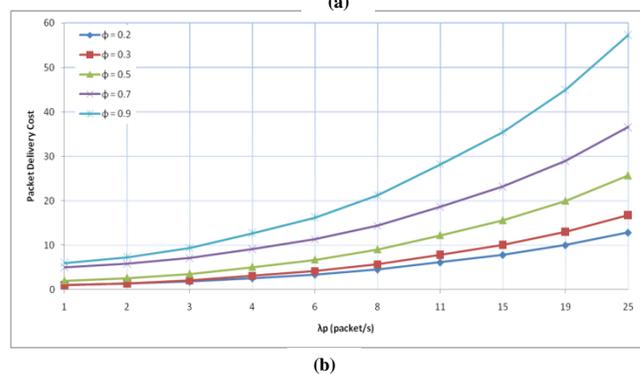
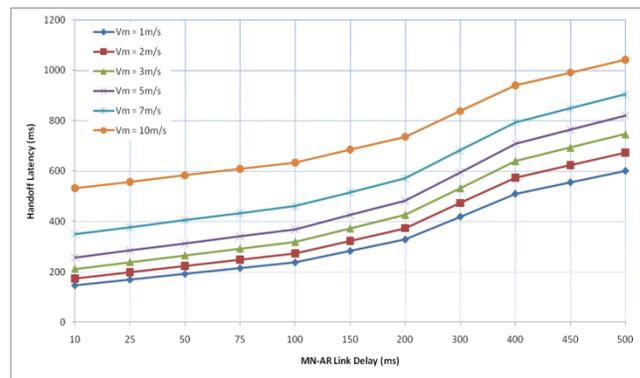


Figure 3. a) VFHO Latency vs. Wireless Delay, b) Packet Delivery vs. Arrival Rate

$\beta_{k,k-1}$  are defined as probabilities that the user approaches or retreats with one random step unit in a hexagonal macro-cell with  $k$  surrounding hexagonal micro-cells [17].

$$\alpha_{k,k+1} = \begin{cases} (1-p), & k=0 \\ (1-p) \left( \frac{1}{3} + \frac{1}{6k} \right), & K \geq k \geq 1 \end{cases} \quad (9)$$

$$\beta_{k,k-1} = (1-p) \left( \frac{1}{3} + \frac{1}{6k} \right), \quad K \geq k \geq 1 \quad (10)$$

Using Equations 8 and 9, the steady-state probability of state  $k$  within the local area with  $K$  local areas,  $p_{k,K}$ , can be obtained in terms of the steady-state probability,  $p_{0,K}$  with the conditions  $\sum_{k=0}^K p_{k,K} = 1$  can be written as:

$$p_{0,K} = \frac{1}{1 + \sum_{k=1}^K \prod_{i=0}^{k-1} \frac{\alpha_{i,i+1}}{\beta_{i+1,i}}} \quad (11)$$

The impact of wireless link on overall signaling cost is described based Session-to-Mobility Ratio (SMR) [17], which is defined as the ratio of session arrival rate ( $\lambda_s$ ) to session crossing rate ( $R_s$ ) in a random-walk mobility model [18].

The probability that the user moves through a random-walk mobility model from a local area ( $l$ ) to a routing area ( $r$ ),  $P_T(r,l)$  can be expressed by:

$$P_T(r,l) = \sum_{k=0}^K P_K(k) \cdot P_T(K|k) \cdot \alpha_{K,K+1} \quad (12)$$

where  $P_K(k)$  is the probability of an incoming session during the time that the user stays in state  $k$  for a given number of states  $K$  and  $P_T(K|k)$  is the probability that a session initiated in state  $k$ , continues in state  $K$ . These probabilities can be obtained from:

$$P_K(k) = \frac{\bar{t}(k) \cdot p_{k,K}}{\sum_{k=0}^K \bar{t}(k) \cdot p_{k,K}} \quad (13)$$

where,  $\bar{t}(k)$  is the mean time the user stays in state,  $k$ . finally, the cost of each operation can be expressed as follows.

$$C_{HD} = \lambda_S [P_T(r, l) \cdot (D_r + D_l) + D_l(1 - P_T(r, l))] \quad (14)$$

$$C_{LH} = \lambda_S [P_T(r, l) \cdot (D_r + D_l)] \quad (15)$$

$$C_{IH} = \lambda_S [D_r \cdot P_T(r, l)] \quad (16)$$

$$C_{PD} = \lambda_S [D_r \cdot P_T(r, l)] \quad (17)$$

Where,  $D_r$  and  $D_l$  denote two units of signaling cost through wireless link in routing and local areas, respectively, and  $\lambda_S$  is session arrival rate in packets per time unit. We assume typical parameter values that were reported in various studies with similar analytical models [17, 19] as listed in Table 3. The signaling cost in wireless link is defined as the product of the distance between the two nodes and transmission cost in wireless link ( $w$ ).

Fig. 4 illustrates the resulting total signaling cost under

various circumstances. The total signaling cost is shown based on the number of local areas within a routing area as in Fig. 4-a, and based on the number of MNs in the routing area in Fig. 4-b. As shown through the figures, the probability of MN leaving the local area is an important factor in the total cost, and greater values cause increase in the handoff decision element of the total cost. Additionally, increasing both number of WLAN ARs and MNs also causes higher signaling costs.

Fig. 4-c shows the variation of total signaling cost versus SMR through a random-walk mobility model within a routing area. The total cost was determined for SMR values of 0.1, 1, 10, and 100. For the minimum value of SMR, 0.1, the total signaling cost shows very high value. However, as the SMR increases to larger than 1 which implies session arrival rate is higher than mobility rate, the signaling cost decreases but the impact of  $k$  factor in the total signaling becomes more chromatic. This is because the packet delivery cost is the dominant factor when the session arrival rate is higher than the mobility ratio.

When the value of cell crossing rate is fixed, the increase of SMR should result in the increase of session arrival rate and thereby, the total cost. This is because the link switch cost is more dominant than packet delivery cost over the total cost. However, the size of routing area as depicted in Fig. 4-d is almost as significant as cell crossing rate in the resulting

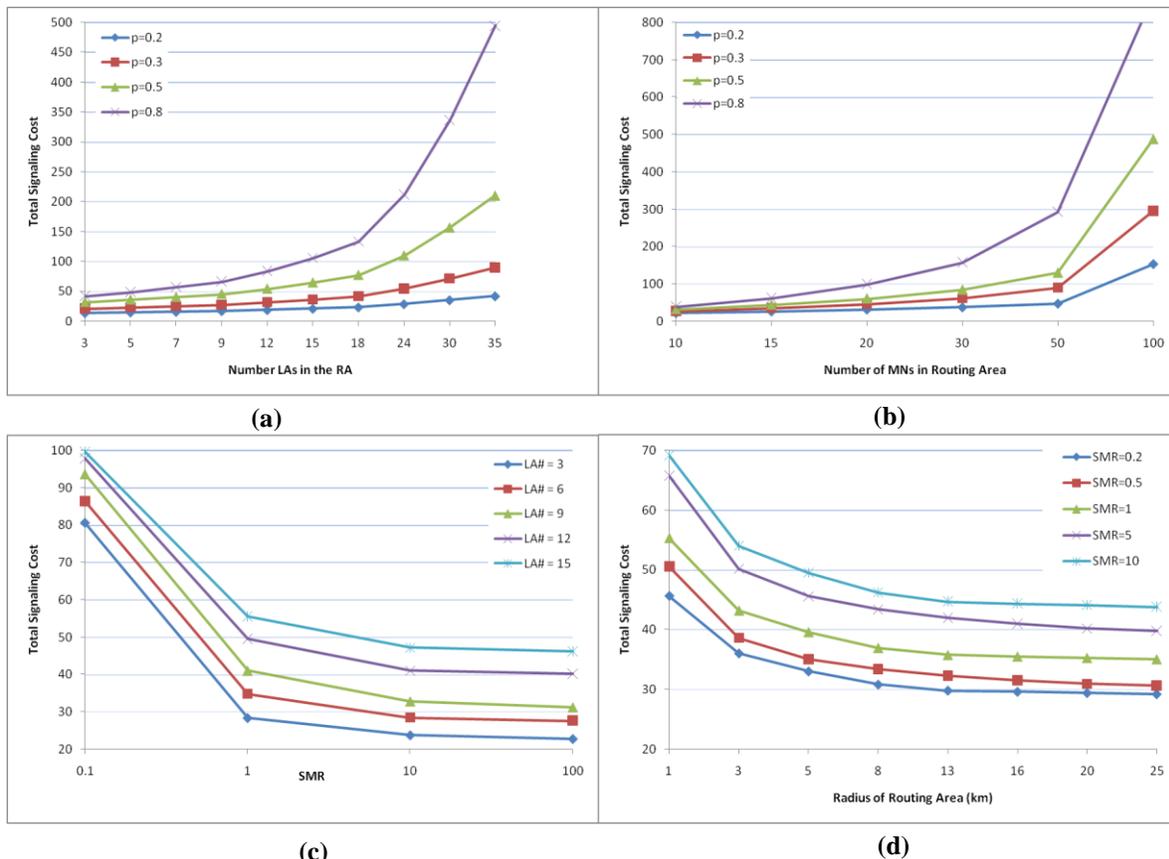


Figure 4. Total Signaling Cost Variations vs. a) Number of LAs, b) MN Density, c) SMR, and d) RA Size

total cost. For instance, the radius of 25km can cause a total signaling cost of as low as radius of 1km when SMR is decreased by 50 times. As a result, higher SMR values incur higher packet delivery cost, while we can still control link layer switching by adjusting the size of routing area.

#### V. CONCLUSION

This paper proposes a seamless vertical handoff protocol in heterogeneous wireless network of WLAN and UMTS technologies. The proposed handoff protocol, VFHO, is a combination of link and IP layers operations, which handles media heterogeneity in between these two layers as well as information of application layer. The introduced approach utilizes some techniques such as EBU to guarantee the continuation of packets in heterogeneous networks which has been barely the concern of the existing literature. Hence, through the distinct definition of handoff latency proposed here, as well as costs of signaling, the proposed method performs more affordable than the existing methods upon being built up under identical circumstances. VFHO was analyzed mathematically to examine packet delivery delay and signaling overhead in terms of cost functions. The proposed framework and vertical handoff method show robust performance in terms of tolerable signaling overhead as well as handoff and packet delays.

#### REFERENCES

- [1] "IEEE Standard for Local and metropolitan area networks- Part 21: Media Independent Handover," *IEEE Std 802.21-2008*, pp. c1-301, 2009.
- [2] G. Lampropoulos, *et al.*, "Media-Independent Handover for Seamless Service Provision in Heterogeneous Networks," *IEEE Communication Magazine*, vol. 46, pp. 64-71, 2008.
- [3] V. Gupta, "IEEE802.21 Standard and Metropolitan Area Network: Media Independent Handover Services," *IEEE*, vol. Draft, March 2006.
- [4] R. S. Koodli and C. E. Perkins, *Mobile Inter-Networking with IPv6: Concepts, Principals, and Practices*. New Jersey: John Wiley & Sons, Inc., 2007.
- [5] K. D. Wong, *Wireless Internet Telecommunications*. Norwood: Artech House Mobile Communications Series, 2005.
- [6] Y. Y. An, *et al.*, "Reduction of Handover Latency Using MIH Services in MIPv6," in *20th International Conference on Advanced Information Networking and Applications*, 2006, pp. 229 - 234.
- [7] L. Dimopoulou, *et al.*, "Analysis and evaluation of layer 2 assisted fast mobile IPv6 handovers in a WLAN environment," in *Computers and Communications, 2005. ISCC 2005. Proceedings. 10th IEEE Symposium on*, 2005, pp. 327-332.
- [8] W. Tong, *et al.*, "Cost-Aware Handover Decision Algorithm for Cooperative Cellular Relaying Networks," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, 2008, pp. 2446-2450.
- [9] A. Dutta, *et al.*, "Seamless Handover across Heterogeneous Networks - An IEEE802.21 Centric Approach," Patent (WO/2006/125085), Available: <http://www.wipo.int/pctdb/en/wo.jsp?WO=2006125085>, Last visited: Sept. 2010.
- [10] M. Kassar, *et al.*, "An overview of vertical handover decision strategies in heterogeneous wireless networks," *Computer Communications*, vol. 31, pp. 2607-2620, 2008.
- [11] Q. B. Mussabbir, *et al.*, "Optimized FMIPv6 Using IEEE 802.21 MIH Services in Vehicular Networks," *Vehicular Technology, IEEE Transactions on*, vol. 56, pp. 3397-3407, 2007.
- [12] D.-H. Kwon, *et al.*, "Access router information protocol with FMIPv6 for efficient handovers and their implementations," in *IEEE Global Telecommunications Conference, 2005. GLOBECOM '05*, 2005, p. 6 pp. 3819-3825.
- [13] M. Liebsch, *et al.* (2005). *Candidate Access Router Discovery (CARD)*. Available: <http://www.ietf.org/rfc/rfc4066.txt?number=4066>, Last visited: Dec. 2010.
- [14] E. Stevens-Navarro and V. W. S. Wong, "Comparison between Vertical Handoff Decision Algorithms for Heterogeneous Wireless Networks," in *Vehicular Technology Conference, 2006. VTC 2006-Spring. IEEE 63rd*, 2006, pp. 947-951.
- [15] G. Chuanxiong, *et al.*, "A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 22, pp. 834-848, 2004.
- [16] G. Bolch, *et al.*, *Queueing Networks and Markov Chains, Modeling and Performance Evaluation with Computer Science Applications*, 2 ed. New Jersey: Wiley & Sons, Inc., 2006.
- [17] D. H. Kim, *et al.*, "Dynamic location management scheme for location registers in wireless ATM networks," *Electronics Letters*, vol. 36, pp. 1650-1651, 2000.
- [18] J. S. M. Ho and I. F. Akyildiz, "Mobile user location update and paging under delay constraints," *Wirel. Netw.*, vol. 1, pp. 413-425, 1995.
- [19] X. Zhang, *et al.*, "P-MIP: Paging Extensions for Mobile IP," *Mobile Networks and Applications*, vol. 7, pp. 127-141, 2002.

# A Hop-Oriented Position Estimation Algorithm for Wireless Sensor Networks

Hsing-Lung Chen, Li-Chung Lin

Department of Electronic Engineering  
National Taiwan Univ. of Sci. and Tech.  
Taipei, Taiwan

hlchen@mail.ntust.edu.tw, M9302111@mail.ntust.edu.tw

Shu-Hua Hu

Department of Comp. Sci. and Info. Eng.  
Jinwen Univ. of Sci. and Tech.  
New Taipei, Taiwan  
shhu@just.edu.tw

**Abstract**—A wireless sensor network consists of a group of sensor nodes that broadcast the sensed data to the base station hop by hop via radio frequency. It is useful only if the sensed data are associated with the locations of the sensor nodes. Therefore, location estimation of sensor nodes has become an important issue. Ideally, each sensor node can obtain its location by being equipped with a GPS device. However, this approach costs too much, contradicting the objective of low cost sensor nodes. Hence, it is reasonable that a few sensor nodes only are equipped with a GPS device, and the others estimate their own locations by way of the collected information. The purpose of this paper is to propose a hop-oriented position estimation algorithm (HOPE). Four beacon nodes, each equipped with GPS, broadcast the hop count information, and normal nodes rebroadcast the information after receiving it. Finally, normal nodes employ the received hop count information to estimate their own locations with simple calculations. The simulation results show that the proposed algorithm has better accuracy for location estimation than other proposed methods.

**Keywords**- sensor nodes; GPS; broadcast; hop count; beacon nodes; normal nodes

## I. INTRODUCTION

Recent advances in MEMs (MicroElectroMechanical Systems) technology and wireless communications technology have provided micro-electronic devices with the capabilities of accurate sensing, communication and computation; such devices are called sensor nodes [1]. Because the sensor node hardware is designed with the objectives of low cost, small size and low power consumption, there are great restrictions on sensor nodes' power, memory and computation ability.

A wireless sensor network (WSN) consists of a group of sensor nodes that broadcast the sensed data to the base station hop by hop via radio frequency. These sensor nodes can be randomly deployed in a region by the hundreds or thousands. However, such a network is useful only if the sensed data are associated with the locations of the sensor nodes. Therefore, how to get the precise location information with a reasonable amount of effort is an important issue in WSN design.

In WSNs, some research assumes that sensor nodes maintain their own location information. This can be realized if the WSNs are situated in a small region, or if the sensor nodes can be placed regularly. In recent years, because of the development of the Global Positioning System (GPS) [2], it is assumed that all sensor nodes can obtain location information by GPS. However, the price of GPS is more

expensive than the sensor node itself. Hence, it is unreasonable that each sensor node be equipped with GPS. A more reasonable approach is to assume that a few sensor nodes (named beacon nodes) are equipped with GPS, and the others (named normal nodes) estimate their own locations by way of the collected information. This paper introduces a hop-oriented position estimation algorithm (HOPE). HOPE only employs four beacon nodes to broadcast the hop count information. Normal nodes can estimate their own locations more accurately with simple calculations by way of the collected hop count information.

This paper is organized as follows. In Section 2, we introduce some of the related works in location estimation research. Section 3 describes our Hop-Oriented Position Estimation algorithm. Section 4 presents the simulation results. Section 5 concludes this paper.

## II. RELATED WORK

Current location estimation algorithms for WSNs can generally be categorized as range-based and range-free approaches.

The range-based approaches estimate the locations of sensor nodes by the information of distance and angle between two different sensor nodes. For example, the Time of Arrival (TOA) approach measures the relative distance between two nodes by the signal propagation time [3]. However, it is hard to synchronize the time of all sensor nodes, resulting in poor accuracy of estimation. Another approach, Time Difference of Arrival (TDOA) [4], [5], [6], is similar to TOA. However, for TDOA, the sensor nodes are equipped with both ultrasound and RF hardware to solve the problem of time synchronization, resulting in good accuracy of estimation. Angle of Arrival (AOA) employs directional antennae or digital compasses [7] to measure the relative angle between two sensor nodes. The Received Signal Strength Indicator (RSSI) [4], [8], [9], [10], [11], [12] approach translates the received signal strength to the estimated distance between the receiving node and the transmitting node based on the radio propagation theory. Nevertheless, the received signal strength is often affected by the physical-layer problems inherent in RF systems, such as multi-path fading, unstable signal propagation and background noise. The range-based approaches often need additional equipment to improve the accuracy of location estimation. Hence, these approaches are not suitable for resource-limited WSNs.

Because the cost of range-based approaches is higher, many range-free approaches are proposed. For example, DV-Hop [7], [13] is based on the concept of distance vector

routing (DV-routing). DV-Hop refers to location-aware nodes as anchors. Each anchor broadcasts its position throughout the network. All nodes should maintain the minimum hop count to each anchor in the table. Each anchor can calculate the average distance per hop by hop count and the distance, and then broadcast it to the neighboring normal nodes. Normal nodes can estimate their location based on the minimum hop count table, the average distance per hop, and the anchors' locations. The GRIPHON [14] scheme also utilizes hop count information for location estimation. Four nodes with known locations called markers are placed at the four corners of the network region. Like anchors, markers broadcast their positions to their neighbors. The other nodes can obtain the minimum hop count of each marker. The hop count vector ( $hcv$ ) of node  $k$  is represented as  $H_k = (h_1, h_2, \dots, h_i)$ , where  $h_i$  is the minimum hop count from node  $k$  to marker  $i$ . The network region is further subdivided into small grid zones. Then the control center has already calculated a mean  $hcv$  for each grid zone. At the control center, the  $hcv$  of each node is compared with the mean  $hcv$ s of all grid zones to determine in which grid zone the node is residing. The sensor node's location is estimated as the central point of its residing grid zone. The approaches employing hop count information to do location estimation require a great deal of communication, and only work well in dense networks.

The Convex Position Estimation (CPE) [15] algorithm assumes that if a normal node can receive the broadcasts of neighboring beacon nodes, it must reside in the overlapping region of the communication range of these beacon nodes. CPE defines the estimative rectangle (ER) as the smallest rectangle covering the overlapping convex, and regards the center of the rectangle as the estimative position of the normal node. However, it needs a great deal of computation to obtain the overlapping convex. Hence, CPE needs a central controller to estimate the position of each normal node and to flood the estimative position back to each normal node. Moreover, CPE has poor scalability due to the heavy traffic load.

Like CPE, the Distributed Location Estimation (DLE) [16] algorithm employs the ER to estimate locations of normal nodes. However, DLE simplifies the computation by replacing complicated functions with simple operations. Therefore, the computation can be distributed to every normal node, instead of just being performed by the central controller. In order to improve the accuracy of the estimative location, DLE introduces the concept of the farther neighboring beacon nodes, whose communication range may overlap the ER and does not cover the normal node. Then normal nodes can adjust their ERs by excluding the communication range of the farther neighboring beacon node, from the original ER. Thus, the accuracy of location estimation can be improved.

### III. HOP-ORIENTED POSITION ESTIMATION ALGORITHM

The preliminary version of this paper is in [17]. In this section, we describe the HOPE algorithm for location estimation in WSNs. Our HOPE algorithm makes the following assumptions:

- Each sensor node has a unique ID.

- Sensor nodes are deployed randomly.
- Beacon nodes obtain their location by GPS.
- Normal nodes need to estimate their own locations.
- Beacon nodes have six kinds of communication range ( $1\beta, 2\beta, 3\beta, 4\beta, 5\beta$  and  $6\beta$ ), where  $\beta$  is the unit of one hop.
- Normal nodes have three kinds of communication range ( $4\beta, 5\beta$  and  $6\beta$ ).

Our proposed HOPE algorithm not only achieves more accurate location estimation but also employs less beacon nodes than other approaches. There are two phases in the HOPE algorithm. We will describe these in the following section.

#### A. Data Collection Phase

In the HOPE algorithm, each normal node estimates its own location by the number of hop counts to every beacon node. To make the estimation more accurate, we assume that one beacon node is placed in each of the four corner areas at random, as shown in Fig. 1.

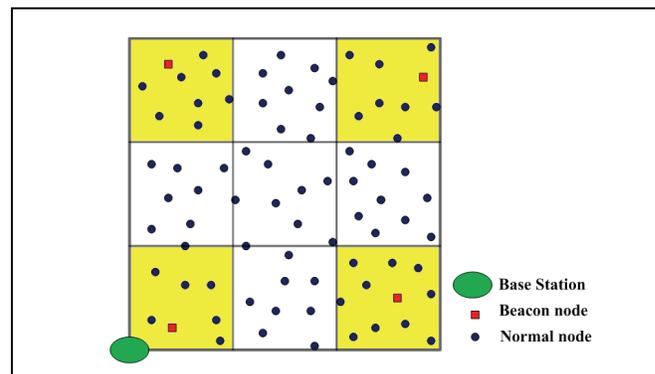


Figure 1. Example of network region.

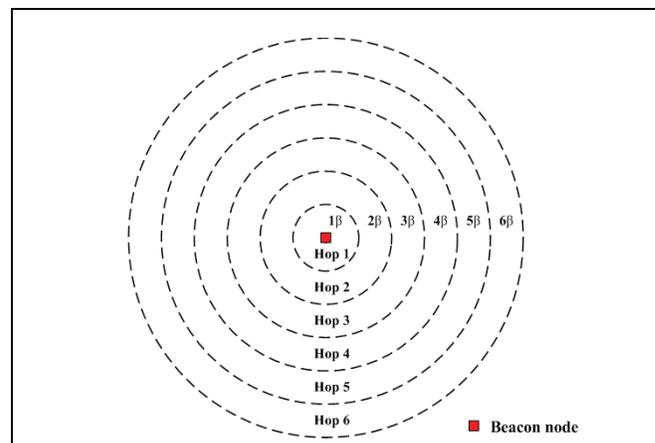


Figure 2. The scenario of beacon node broadcasting.

Because it is difficult to synchronize the time of all sensor nodes, HOPE lets the base station be the master to control the sensor nodes. The base station will notify a beacon node that it can start broadcasting hop count information. After a fixed time interval, the base station will

notify the next beacon node to start broadcasting. For instance, beacon *A* first broadcasts information with hop count 1 to its neighbors by using the transmission range  $1\beta$ . Each node, when it receives the hop count information 1, will record the hop count 1. Then beacon *A* will change the transmission range from  $2\beta$  to  $6\beta$  successively and broadcast information with hop counts from 2 to 6, respectively. Rings 1 to 6 are formed. Each node records the coordinates of beacon *A* and the smallest hop count to beacon *A*. The scenario of beacon node broadcasting is shown in Fig. 2.

In order to reduce broadcasting collisions, the concept of a “broadcast slot” is proposed. We let nodes start forwarding the hop count information according to their hop counts. The beacon node will broadcast the information with hop counts from 1 to 6 in slot 0. The nodes with received hop count 1 will broadcast the information with hop counts from 5 ( $1+4$ ) to 7 ( $1+6$ ) by using the transmission range from  $4\beta$  to  $6\beta$ , respectively, in slot 1. Nodes with the received hop count  $k$  will broadcast the information with hop counts from  $(k+4)$  to  $(k+6)$  in slot  $k$ . Each node will keep the smallest hop count to the beacon. The nodes in ring  $j$  will receive the messages with hop count  $j$  from the nodes in rings  $j-4, j-5$  and  $j-6$ . This redundancy can reduce the effect of message collision and non-uniform distribution of sensor nodes. An example of broadcast slots is shown in Fig. 3. The maximum number of slots can be estimated with the regional size and  $\beta$ .

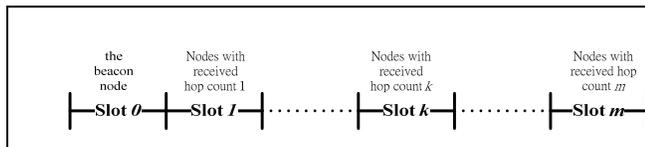


Figure 3. An example of broadcast slots.

The broadcast data contains the hop counts of this message and sender, the coordinates of the beacon node, and the remaining time. The remaining time is defined as the time interval between the sending time and the ending of the current slot. When a node receives the broadcast data, it can calculate the waiting time of its broadcasting slot, which is equal to the remaining time plus the remaining slots. Then the node starts broadcasting information randomly in its broadcasting slot. We use relative time to solve the problem of time synchronization. Nodes will broadcast data in their corresponding broadcasting slots at random to reduce the occurrence of broadcasting collisions.

After four beacon nodes (*A, B, C, D*) finish broadcasting, each node will have the smallest hop counts away from four beacon nodes, ( $h(A), h(B), h(C), h(D)$ ), called hop count vector (*hcv*).

The nodes within the same ring will have the same received hop count. Considering the coverage of a beacon node’s broadcast, it will form six concentric circles. We hope that further coverage of broadcastings approach the concentric circles. Therefore, other nodes broadcast information using three kinds of transmission range ( $4\beta, 5\beta, 6\beta$ ). After the concentric circles for the four beacon nodes are overlaid, the region will be divided into several disjointed areas, called vector blocks. Nodes in the same vector block

have the same *hcv*. Fig. 4 shows an ideal vector block situation.

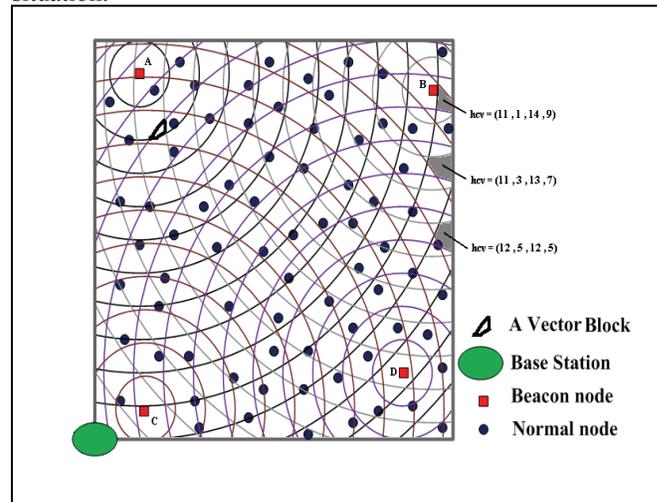


Figure 4. An ideal vector block situation.

### B. Position Estimation Phase

After each node has obtained a hop count vector, the base station will broadcast the message of location estimation. Each node can estimate its own location by using the *hcv* and the coordinates of the beacon nodes. Each vector block, which is formed with the overlapping area of four corresponding rings, has a unique *hcv*. Each node estimates its location as follows:

1. A normal node first chooses two beacon nodes. The intersection of their corresponding rings will be two quasi-parallelgrams. The centers of the two quasi-parallelgrams can be derived by intersecting the central lines of two corresponding rings. One quasi-parallelgram closely related to the hop count vector is selected as the target quasi-parallelgram.
2. Then it chooses one of the two remaining beacon nodes. The intersection region will be formed by overlapping this corresponding ring and the target quasi-parallelgram. In the center line of this corresponding ring, find a point closest to the center of the target quasi-parallelgram. The middle point between this point and the center of the target quasi-parallelgram is regarded as the center of this intersection region.
3. Repeat the last procedure for the last beacon node. The center of the vector block is regarded as the estimated location of this normal node.

### C. Quasi-parallelgrams

During the process of finding the target quasi-parallelgram, several problems occur. If we choose any two beacon nodes, the intersection of the two corresponding rings may not be two disjointed quasi-parallelgrams. It is difficult to calculate the center of the target quasi-parallelgram. The angle with respect to two selected beacon nodes is defined as the angle from the normal node to the two selected beacon

nodes, which can be estimated by the hop count vector. The two beacon nodes whose corresponding angles are closest to a right angle are selected, because the corresponding quasi-parallelgram is the closest to a rectangle.

Another problem is that a normal node will select an incorrect target quasi-parallelgram. If one of two selected beacon nodes is closest to the normal node, we may select an incorrect target quasi-parallelgram because two quasi-parallelgrams are too close to distinguish by inspecting the hop count vector. In selecting two beacon nodes, that which is closest to the normal node is not considered. Consider the example in Fig. 5. Beacon node *A* is not considered because it is closest to the normal node. There are three combinations left to be considered: *BC*, *BD* and *CD*. Beacon nodes *B* and *C* are selected because the angle with respect to beacon nodes *B* and *C* is closest to a right angle. The quasi-parallelgram close to beacon node *A* is selected as the target.

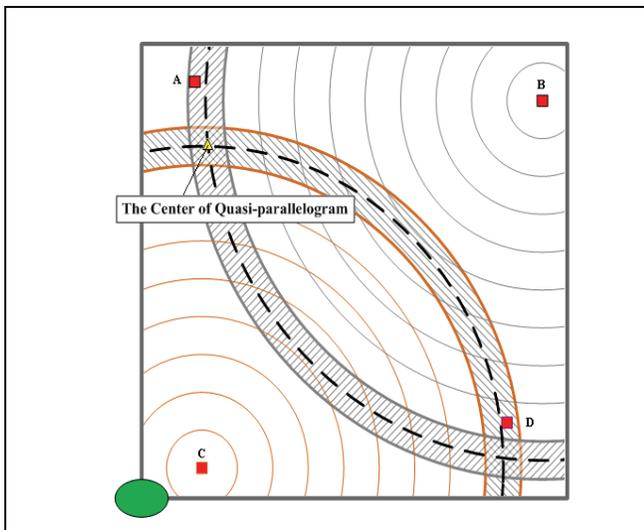


Figure 5. An example of obtaining the center of the target quasi-parallelgram.

**D. Intersection Region**

Recall that a vector block is the intersection of four corresponding rings. The target point of the central line of a ring is defined as a point on the central line closest to the center of the target quasi-parallelgram. At first, we choose one beacon node with the shorter distance from its corresponding target point to the target quasi-parallelgram. Thus, we can obtain a bigger intersection region by intersecting the target quasi-parallelgram and the corresponding ring.

The center of the intersection region can be estimated as the middle point between the center of the target quasi-parallelgram and the corresponding target point, as shown in Fig. 6. Finally, the vector block is the intersection of the intersection region and the last corresponding ring. The center of the vector block can be estimated as the middle point between the center of the intersection region and the corresponding target point, as shown in Fig. 7.

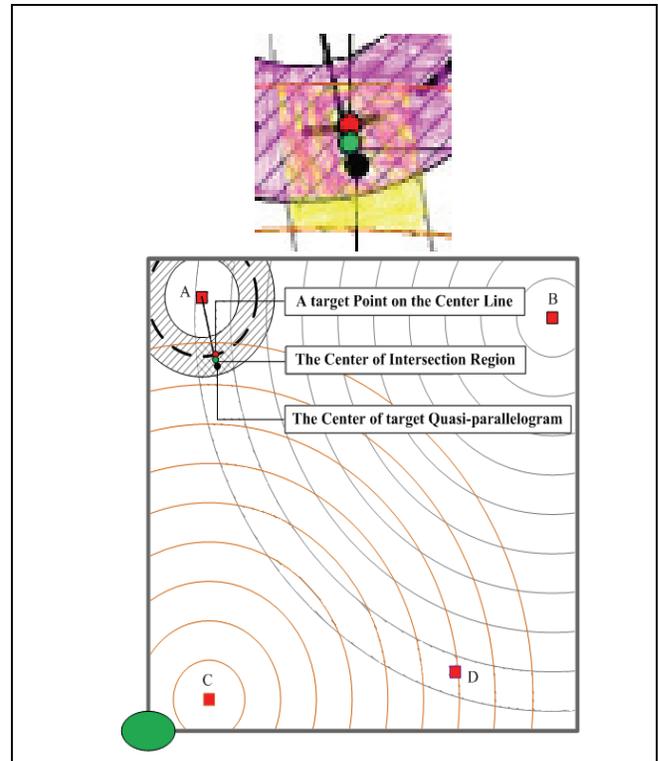


Figure 6. An example of obtaining the center of the intersection region.

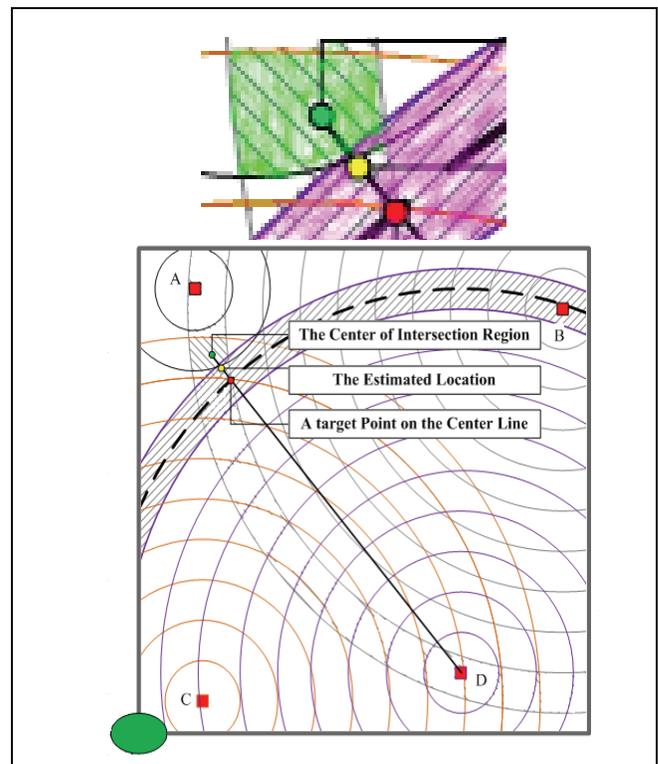


Figure 7. An example of obtaining the estimated location.

IV. SIMULATION AND RESULTS

In order to demonstrate the accuracy and feasibility of the HOPE algorithm, we will conduct the simulations in two kinds of situations.

At first, we compare the Distributed Location Estimation algorithm with HOPE. We assume that there are no collisions, and energy consumption is not considered. The simulation area is a square region of side length 1,000m. The total number of nodes is tuned between 50 and 250. To show the accuracy, we compute the mean error which is the average distance between the estimated and actual locations. Then HOPE uses Network Simulator version 2 (NS-2) to further consider the collisions. The 802.11 MAC protocol with the proposed broadcast slot mechanism is employed. To show the feasibility, the mean errors of the direct computing version and the NS-2 version are compared.

A. Impact of the Density of Sensor Nodes

Because the DLE algorithm needs a proportion of beacon nodes, the ratio of beacon nodes to the total nodes is set to 40%. Fig. 8 shows the impact of the density of sensor nodes on the mean error. When the node density is low (less than 150 nodes), DLE has poor performance. Although DLE has good performance when the node density is high, the number of beacon nodes will be increased. Under any node density, HOPE has a smaller mean error than DLE. To show the stability of both algorithms, we define the mean error range of DLE and HOPE. Fig. 9 shows that the mean error range of HOPE is smaller than that of DLE. Therefore, it is shown that our HOPE algorithm is more stable than the DLE algorithm.

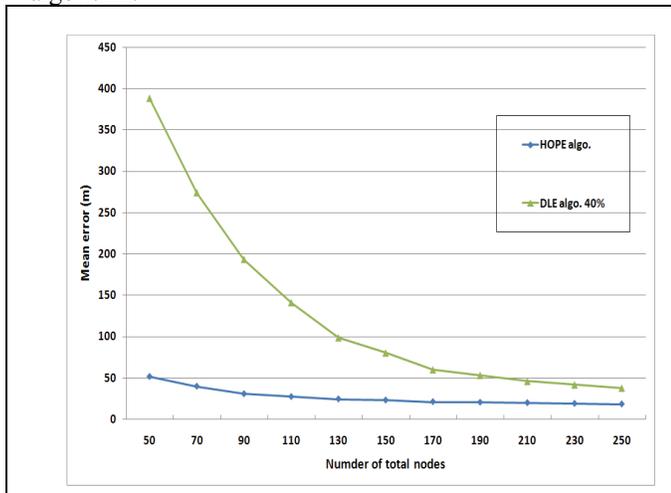


Figure 8. Node density vs. mean error.

B. Impact of the Time Synchronization and Collisions

In HOPE, each node obtains needed information by way of broadcasting messages. Collisions will happen in actual networks. In order to prove the feasibility of the algorithm, we conduct the simulations in NS-2, and compare it with the collision-free version. Fig. 10 shows the comparison of the collision-free version and the NS-2 version in terms of mean error. In low node density, the NS-2 version has more errors

than the collision-free version because some nodes cannot receive the correct hop count information due to collisions. In high node density, the NS-2 version has almost the same mean error as the collision-free version, because most of the nodes can receive the correct hop count information from at least one node without collisions. It is shown that the broadcast slot mechanism can avoid most collisions successfully.

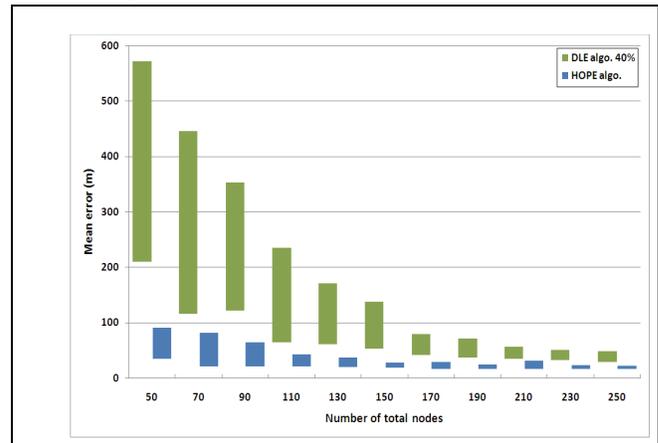


Figure 9. The experiment result of mean error range.

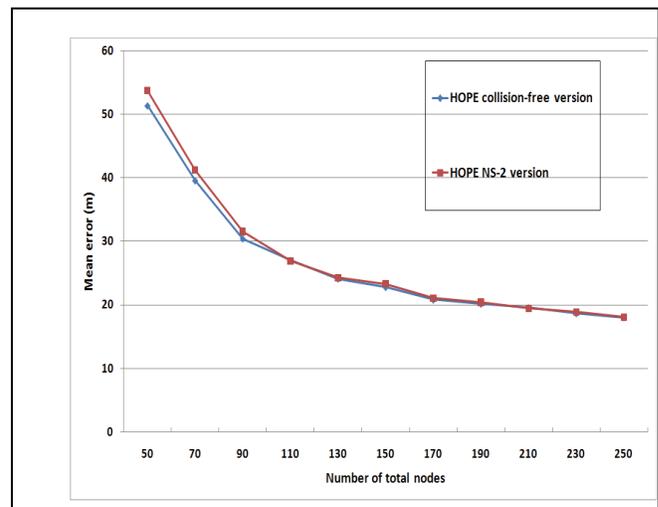


Figure 10. The comparison of the collision-free version and the NS-2 version.

C. Enhancing the Accuracy of the Algorithm

Based on the concept of vector blocks, it could be inferred that if the vector block can become smaller, normal nodes in the vector block will be closer to the center of the vector block, resulting in a decrease in the mean error. The vector block can be diminished by decreasing the  $\beta$  value. Fig. 11 shows the mean error in low node density by tuning the  $\beta$  value between 20 and 50, and using a 95% confidence interval. From this figure, it is found that if node density is less than some value with respect to  $\beta$ , mean error is increased abruptly, because some nodes cannot receive the correct hop count information.

Fig. 12 shows the mean error in high node density and using a 95% confidence interval. It is found that if the node density is sufficient, the mean error will be smaller by reducing the  $\beta$  value.

From Figs. 11 and 12, we know that when the density is less than 400 (node/km<sup>2</sup>), the use of  $\beta = 50$  is better. When the density is between 400 and 700, we should use  $\beta = 40$ . When the density is between 700 and 1,000, we can employ  $\beta = 30$ . When the density is from 1,000 to 5,000, we can use  $\beta = 20$ . If the density is higher than 5,000, we recommend the use of  $\beta = 10$ . Users can thus adjust the  $\beta$  value to obtain a better mean error.

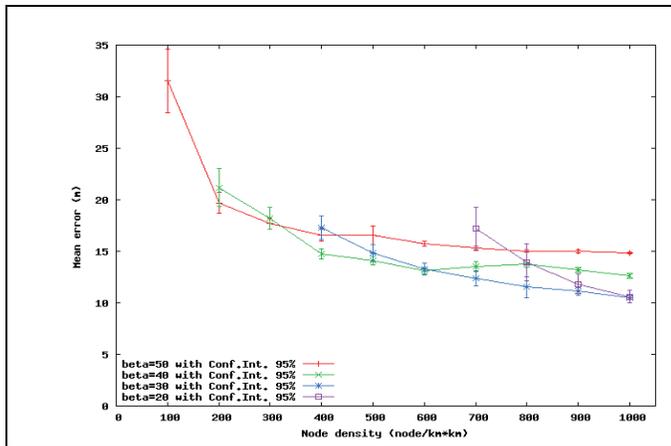


Figure 11. Adjusting  $\beta$  value in low node density.

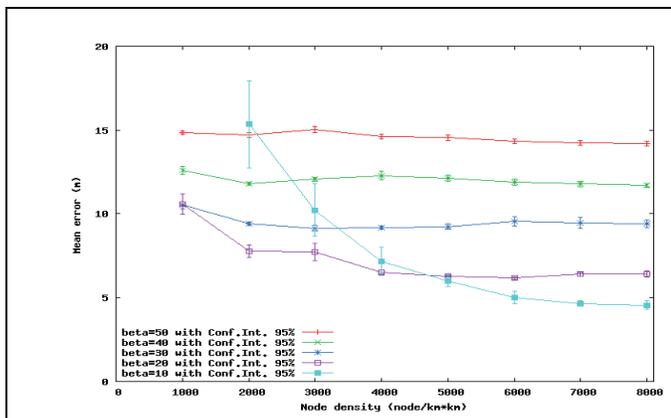


Figure 12. Adjusting  $\beta$  value in high node density.

V. CONCLUSIONS

WSNs are useful only if the sensed data are associated with the locations of the sensor nodes. It is reasonable that a few sensor nodes are equipped with a GPS device and that the others estimate their own locations by way of collected information. In order to reduce costs while maintaining estimation accuracy, the HOPE algorithm is proposed.

The simulation results show that the HOPE algorithm not only uses fewer beacon nodes, but also has higher accuracy as a result of tuning the  $\beta$  value. In simulations of the NS-2

version, it is found that a broadcast slot mechanism can reduce collisions significantly, resulting in high stability of the HOPE algorithm. Implementing the HOPE algorithm in real WSNs is worthy of further investigation.

ACKNOWLEDGMENT

This work was supported, in part, by the National Science Council, Republic of China, under Grant No. 99-2218-E-011-001.

REFERENCES

- [1] J. Hill, R. Szewczyk, A. Woo, S. Hollar, D. Culler, and K. Pister, "System architecture directions for networked sensors," *ASPLOS 2000*, Cambridge, USA, November 2000, pp. 93-104.
- [2] Garmin international, "What is GPS," <http://www.garmin.com/aboutGPS>.
- [3] S. Capkun, M. Hamdi, and J.-P. Hubaux, "GPS-free positioning in mobile ad-hoc networks," *Cluster Computing*, Vol. 5, No. 2, pp 157-167, 2002.
- [4] J. Hightower and G. Borriello, "Location Systems for Ubiquitous Computing," *IEEE Computer*, vol. 34, No. 8, pp. 57-66, August 2001.
- [5] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The Cricket Location-Support System," *ACM/IEEE MobiCom 2000*, pp. 32-43, Boston, MA, August 2000.
- [6] S. Ray, R. Ungrangsi, F. D. Pellegrini, A. Trachtenberg, and D. Starobinski, "Robust Location Detection in Emergency Sensor Networks," *INFOCOM 2003*, Vol. 2, pp. 1044-1053, March-April 2003.
- [7] D. Niculescu and B. Nath, "DV Based Positioning in Ad Hoc Networks," *Telecommunication Systems*, Vol. 22, No. 1-4, pp. 267-280, 2003.
- [8] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less Low Cost Outdoor Localization For Very Small Devices," *IEEE Personal Communications*, Vol. 7, No. 5, pp. 28-34, October 2000.
- [9] F. Mondinelli and Z. M. Kovacs-Vajna, "Self Localizing Sensor Network Architectures," *IEEE Transactions on Instrumentation and Measurement*, Vol. 53, No. 2, pp. 277-283, 2004.
- [10] N. Patwari, R. J. O'Dea, and Y. Wang, "Relative Location in Wireless Networks," *IEEE VTC 2001*, Vol.2, pp. 1149-1153, Rhodes, Greece, May 2001.
- [11] P. Bahl and V. N. Padmanabhan, "RADAR: An In-Building RF-based User Location and Tracking System," *INFOCOM 2000*, Vol. 2, pp. 775-784, March 2000.
- [12] J. Hightower, R. Want, and G. Borriello, "SpotON: An Indoor 3D Location Sensing Technology Based on RF Signal Strength," *UW CSE 2000-02-02*, University of Washington, Seattle, WA, February 2000.
- [13] D. Niculescu and B. Nath, "Ad Hoc Positioning System (APS)," *IEEE Globecom 2001*, Vol. 1, pp. 2926-2931, November 2001.
- [14] J. G. Lim and S. V. Rao, "A grid-based location estimation scheme using hop counts for multi-hop wireless sensor networks," *2004 International Workshop on Wireless Ad-Hoc Networks*, pp. 330-334, May 2004.
- [15] L. Doherty, K. S. J. Pister, and L. El Ghaoui, "Convex Position Estimation in Wireless Sensor Networks," *INFOCOM 2001*, vol. 3, pp. 1655-1663, April 2001.
- [16] J. P. Sheu, J. M. Li, and C. S. Hsu, "A Distributed Location Estimating Algorithm for Wireless Sensor Networks," *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing 2006*, Vol. 1, pp. 218-225, June 2006.
- [17] L. C. Lin, "A Hop-Oriented Position Estimation Algorithm for Wireless Sensor Networks," *Master Thesis*, National Taiwan University of Science and Technology, Taiwan, June 2006.

# Investigation of the Correlation Effect on the Performance of V-BLAST and OSTBC MIMO Systems

Ali Jemmali

Departement of Electrical Engineering  
Ecole Polytechnique de Montreal  
Email: ali.jemmali@polymtl.ca

Jean Conan

Departement of Electrical Engineering  
Ecole Polytechnique de Montreal  
Email: j.conan@polymtl.ca

**Abstract**—In this paper, we study the effect of the spatial correlation between multiple antennas on the performance of the MIMO systems in flat fading channels. As performances metrics, the Bit Error Rate (BER) and Symbol Error Rate (SER) are investigated for Orthogonal Space Time Block Coding (OSTBC) and Vertical Bell Laboratories Layered Space-Time code (V-BLAST) architecture respectively. It is assumed that all antennas have the same radiation pattern and the spatial correlation is independent of the position of an antenna in the array. In this investigation, it is assumed that the correlation matrix of the MIMO channel is the Kronecker product of the receive and transmit correlation matrices. The desired correlated channel is obtained by the Cholesky decomposition of the correlation matrix.

**Keywords**—Multi-antenna MIMO system, OSTBC, V-BLAST, Bit Error Rate

## I. INTRODUCTION

Multiple-input multiple-output (MIMO) systems are those that have multiple antenna at both transmitter and receiver. They were first investigated by computer simulation in the 1980s, and later papers explored them analytically. Since that time, interest in MIMO systems has exploded. Multiple antennas can be used for increasing capacity or for increasing diversity. To improve the diversity in a MIMO system, the Orthogonal Space Time Block Coding are used [10]. On the other hand, to improve the capacity, independent data streams are transmitted simultaneously from multiple antennas [2]. The Vertical Bell Laboratories Layered Space-Time code (V-BLAST) is an architecture for realizing very high data rates over the rich scattering wireless channel [3]. The detection algorithms for V-BLAST architecture are traditionally based on the Order Successive Interference Cancellation (OSIC) [4]. The V-BLAST OSIC algorithm is an iterative algorithm where the the strongest signal is decoded first, then the effect of this strongest signal is cancelled from each of the received signals [8]. The correlation at the receiver and the transmitter side can considerably reduce the performance of the MIMO systems. It is important to quantify the effect of the correlation to properly

design MIMO systems [12]. In [5], the BER expressions of optimum combining and maximal ratio combining in the presence of one co-channel interferer was derived. In [6], the effect of transmission design and spatial correlation on the symbol error rate (SER) was analyzed for MIMO communication systems. While most previous work considered the analysis for OSTBC and V-BLAST with separately correlation model, this work considers the effect of the channel correlation using the jointly-correlated model. With this approach an analytical solution can be derived for the BER and SER performances. When all the antennas in the MIMO systems have the same radiation pattern, which is the case in almost the practical cases, and when the spatial correlation is independent of the position of an antenna in the array, it is shown in [1] that the (4x4) MIMO correlation matrix can be approximated by the Kronecker product of the transmitter correlation matrix and the receiver correlation matrix.

In this paper, the BER and the SER of the OSTBC and OSIC V-BLAST, respectively, are investigated with respect to the spatial correlation effect. The kronecker model of the correlated channel and the Cholesky decomposition of the cross-correlated matrix are considered for the study.

The remainder of this paper is organized as follows. In Section II, we present the system model and give a brief review of the detection algorithm of OSIC. A brief description of the OSTBC is also reviewed in this section. In Section III, we present the correlation model. The simulation results and discussion of results will be presented in Section IV. Finally, we conclude our paper in Section V.

## II. SYSTEM AND CHANNEL MODEL

We consider a narrow band MIMO wireless communication system with 2 transmit antennas and 2 receive antennas. It is assumed that the channel experiences quasi-static flat Rayleigh fading. We also assume that CSI is known at the receiver. For the OSTBC, at each time the two transmit antennas send

Alamouti coded blocks [10] and two receive antennas are used to receive the encoded transmitted block. Let  $H = [h_{i,j}]$  denote the 2x2 MIMO channel matrix where  $h_{i,j}$  is the fading coefficient between transmit  $j$  and receive antenna  $i$ ,  $j = 1, 2$ ,  $i = 1, 2$ .  $h_{i,j}$  is a sample of independent complex Gaussian random variables with zero mean and variance 1/2 per dimension. For decoding purposes, a 2x2 matrix  $\mathbf{H}$  is used to denote the channel between the two transmit and two receive antennas. The complex envelope of the received signal at the antenna array after matching filter is given by [2]:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathbf{y}$  represent a two dimensional received signal,  $\mathbf{n}$  is a 2 dimensional complex additive white Gaussian noise (AWGN) vector, of which each component is statistically independent and has zero mean and a variance  $\sigma^2/2$  per dimension,  $\mathbf{x}$  a two dimensional of transmitted signal and  $\mathbf{H}$  is a 2x2 MIMO fading channel [11]:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \quad (2)$$

To introduce the correlation effect in the system model, a correlated matrix should be generated according to [7]

$$\mathbf{vec}(\mathbf{H}_c) = R_{MIMO}^{1/2} \mathbf{vec}(\mathbf{H}) \quad (3)$$

where  $\mathbf{H}_c$  represent the correlated matrix of the MIMO channel and is given by [7]:

$$\mathbf{H}_c = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} \quad (4)$$

The correlation matrix of the MIMO channel,  $R_{MIMO}$ , is obtained by the Kronecker product of the transmit and receive correlation matrices as follow [11]:

$$R_{MIMO} = R_{MS} \otimes R_{BS} \quad (5)$$

Using the properties of the Kronecker product the model of (2) is simplified to the form [7]:

$$\mathbf{H}_c = R_{MS}^{1/2} \mathbf{H} (R_{BS}^{1/2})^T \quad (6)$$

The correlation matrices of the transmitter and the receiver are given by [11]:

$$R_{BS} = \begin{bmatrix} 1 & \rho_{12}^{BS} \\ \rho_{21}^{BS} & 1 \end{bmatrix} \quad (7)$$

$$R_{MS} = \begin{bmatrix} 1 & \rho_{12}^{MS} \\ \rho_{21}^{MS} & 1 \end{bmatrix} \quad (8)$$

To generate the desired correlated elements, the elements of (4) are arranged in vector form as:

$$\mathbf{vec}(\mathbf{H}_c) = \begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{bmatrix} \quad (9)$$

The correlated and no-correlated elements are then related by [11]:

$$\begin{bmatrix} \alpha_{11} \\ \alpha_{12} \\ \alpha_{21} \\ \alpha_{22} \end{bmatrix} = C \begin{bmatrix} h_{11} \\ h_{12} \\ h_{21} \\ h_{22} \end{bmatrix} \quad (10)$$

where  $C$  is the result of the Cholesky decomposition of the  $R_{MIMO}$  [11]:

$$R_{MIMO} = CC^T \quad (11)$$

### III. V-BLAST OSIC DETECTION ALGORITHM

The simple transmission and detection mechanism of the V-BLAST and its ability to achieve a high data rate performances have made the technic a popular solution for signal processing. In this section, we briefly describe the V-BLAST OSIC algorithm. Letting  $\mathbf{x} = [x_1, x_2]^T$  represent the vector of the transmitted symbols, then the corresponding received vector is given by [9]:

$$\mathbf{y}_1 = \mathbf{H}_c \mathbf{x} + \mathbf{n} \quad (12)$$

In the above equation,  $\mathbf{x}$  and  $\mathbf{y}_1$  are two vectors of transmitted and received signals respectively. To be able to detect the different stream send by the transmitter, an iterative process in which stream by stream detection is necessary. In other words, the detection process should be performed on one stream at time. This process is based on the linear combinatorial nulling

on the received vector and an operation of cancelling the obtained symbol operation is followed [9]. The process should have the following steps:

In a first step, the received vector,  $\mathbf{y}_1$ , should be multiplied by a nulling vector,  $\mathbf{w}_{k1}$  to obtain the quantity  $y_{k1}$  [9]:

$$y_{k1} = \mathbf{w}_{k1}^T \mathbf{y}_1 \quad (13)$$

In the next step,  $y_{k1}$  is sliced using the quantization,  $Q(\cdot)$ , operation appropriate to the the constellation in use and the  $a_{k1}$  is obtained [9]:

$$a_{k1} = Q(y_{k1}) \quad (14)$$

Finally, a modified received vector  $\mathbf{y}_2$  is obtained by cancelling  $a_{k1}$  from the received vector  $\mathbf{y}_1$ , resulting [9]:

$$\mathbf{y}_2 = \mathbf{y}_1 - a_{k1}(\mathbf{H}_c)_{k1} \quad (15)$$

where  $(\mathbf{H}_c)_{k1}$  represents the  $k_1$ th column of  $\mathbf{H}_c$ . The previous steps are then performed on the modified received vector  $\mathbf{y}_2$  to detect the following symbol. The process should be stopped until all symbols are detected.

#### IV. SIMULATION RESULTS

In this section, we illustrate the results of the effect of spatial correlation of MIMO channel on the performances of V-BLAST and OSTBC MIMO architecture. The Monte-Carlo simulations are used to obtain the results. In V-BLAST architecture, a frame of 100 symbols is transmitted over the MIMO channel. The channel is represented by a 2x2 matrix with correlated elements. The performances of the non linear detection are obtained for 5000 iterations. The detected symbols at the output of the receiver are compared with the ones transmitted from the transmitter side so that the SER is calculated. The results of the simulations are represented in the Fig. 1. In this figure, the curve which correspond to zero correlation is identical to the litterature results [2]. From theses results it is shown that the correlation between sub channels degrade considerably the performances of V-BLAST architecture. In fact, for a fixed SNR of 15 dB, the SER of  $\rho = 0.8$  is much important than the one of  $\rho = 0.7$ . For a fixed value of correlation, say  $\rho = 0.4$ , the SER decrease when the value of the signal to noise ratio is increased. We also notice that for  $\rho = 0.999$ , the SER is almost constant

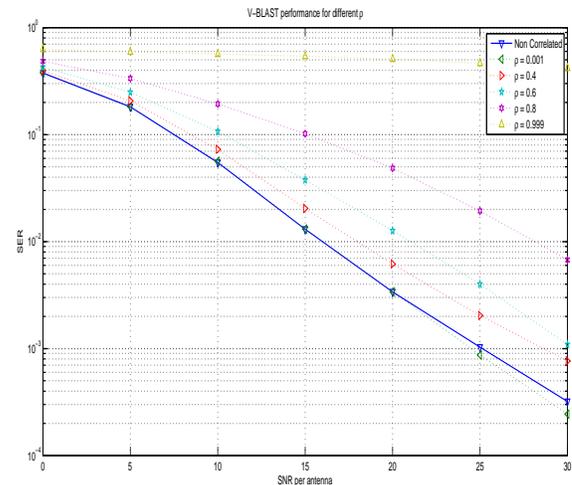


Fig. 1. SER Performances of V-BLAST OSIC algorithm with respect to different spatial correlation  $\rho$

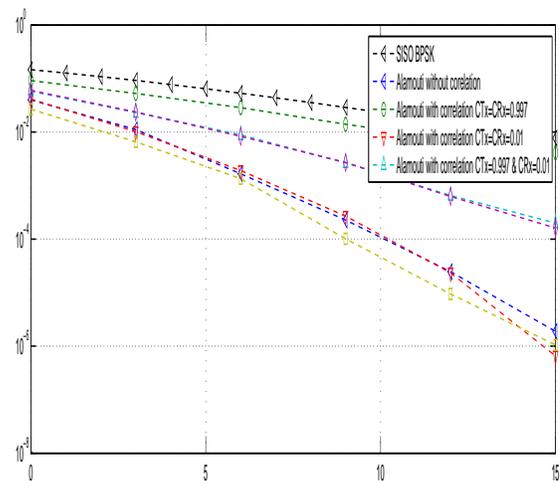


Fig. 2. BER Performances of OSTBC using Alamouti code with respect to different spatial correlation  $\rho$

and doesn't depend on the SNR. This can be explained by the fact that with strong correlation, the algorithm become less efficient and hence the importance of the spatial correlation on the hole performances. For the performances of the OSTBC, the results are presented in the Fig. 2. The one rate Alamouti code and BPSK modulation with real constellation were used for the simulations [10]. In the simulation, a random 2x2 MIMO channel Matrix is generated. The results show also that the strong correlation decrease considerably the BER performances of the OSTBC.

## V. CONCLUSION

In this paper, we investigated the effect of the spatial correlation on the performances of two MIMO architectures known as the V-BLAST and OSTBC. The Kronecker model for the MIMO channel combining with Cholesky decomposition were used to the simulation. The BER and the SER are used as performances metrics for the study. The results shows clearly that the correlation decrease considerably the performances of the MIMO systems.

## REFERENCES

- [1] A. Hotinen, O. Tirkkonen, and R. Wichman " Multi-antenna transceiver techniques for 3 G and beyond, West Sussex, England; Hoboken, NJ: J. Wiley, 2003"
- [2] G. J. Foschini and M. G. Gans " On limits of wireless communications in a fading environment when using multiple antennas" *IEEE Wireless Personal Communications*, pp. 311–335, Mars 1998.
- [3] G. J. Foschini " Layered space time architecture for wireless communication in a fading environment using multi-element antennas" *IEEE Bell Labs. Tech. J. vol. 1, no. 2. Communications*, pp. 41–59, 1996.
- [4] Z. Hufei, W. Chen, B. Li, and F. Gao " An Improved Square Root Algorithm for V-BLAST Based on Efficient Inverse Cholesky Factorization" *IEEE Trans. on Wireless Communications*, vol. 10, no. 1, pp. 43–48, January 2011.
- [5] J. Cui, A.U.H. Sheikh, and D.D. Falconer " BER analysis of optimum combining and maximal ratio combining with channel correlation for dual antenna systems" *IEEE Vehicular Technology Conference, IEEE 47th*, vol. 1, no. 1, pp. 43–48, May 1997.
- [6] E. Bjornson, E. Jorswieck, and B. Ottersten "Impact of Spatial Correlation and Precoding Design in OSTBC MIMO Systems," *Wireless Communications, IEEE Transactions on*, vol. 9, no. 11, pp.3578-3589, November 2010
- [7] H. Ozcelik, M. Herden, W. Weichselberger, J. Wallace, and E. Bonek " Deficiencies of Kronecker MIMO radio channel model" *IEEE Electronics Letters.*, vol. 39, no. 16, pp. 1209–1210, August 2003.
- [8] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela " V-BLAST an architecture for realizing very high data rate over the rich-scattering wireless channel" *URSI International Symposium on Signals, Systems and Electronics.*, 1998.
- [9] G.D. Golden, C.J. Foschini, R.A. Valenzuela, and P.W. Wolniansky " Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture" *IEEE Electronics Letters*, vol. 35, no. 1, pp. 14–16, Jan. 1999.
- [10] S. M. Alamouti "A Simple Transmit Diversity Technique for Wireless Communications", *IEEE Journal on Select Areas in Communications*, vol. 16, no. 8, pp. 1451-1458, October 1998.
- [11] Kermoal, J. P., Schumacher, L., Pedersen, K. I., Mogensen, P. E., and Frederiksen, F "A stochastic MIMO radio channel model with experimental validation", *Selected Areas in Communications, IEEE Journal on*, vol 20, No 6, August 2002, pages 1211-1226.
- [12] Zhemi Xu, Sana Sfar, and Rick S. Blum "Receive Antenna Selection for Closely Spaced Antennas with Mutual Coupling", *IEEE Transaction on Wireless Communications*, vol. 9, no. 2, pp. 652-661, February 2010.

# Simple Detection Schemes for the Alamouti Code Assisted V-BLAST (ACAV) System

David Wee-Gin Lim

*Department of Electrical and Electronic Engineering  
The University of Nottingham Malaysia Campus  
43500 Semenyih, Selangor, Malaysia.  
Email: lim.wee-gin@nottingham.edu.my*

**Abstract**—The Alamouti Code Assisted V-BLAST (ACAV) is a promising hybrid MIMO transmission scheme that is adopted by IEEE 802.11n-2009 (WiFi) recently. It combines spatial multiplexing (SM) and Alamouti space-time block codes (STBC) so that if the Alamouti symbols were detected first, which are generally more reliable due to STBC, subsequent interference cancelation (IC) stages will suffer less from error propagation. The optimal IC detector for the ACAV can be undesirably complex but we found a procedure to reduce the complexity and processing time without compromising its optimality. We further developed a simpler suboptimal detector that is suitable for rapidly varying channels. Simulation results show both ACAV detectors can outperform the detectors of pure SM scheme of the same data rate.

**Keywords**-Hybrid transmission scheme; ACAV; STBC-VBLAST; MIMO detector; MIMO detection scheme

## I. INTRODUCTION

At the turn of the millennium, the multiple-input and multiple-output (MIMO) system has emerged as a promising technology for high data-rate broadband wireless communications. Traditional MIMO schemes either achieve diversity gain to increase the link reliability against fading, e.g., Alamouti space-time block code (STBC) [1], [2], or achieve spatial multiplexing (SM) gain to increase spectral efficiency and data throughput, assuming there is sufficient signal scattering and antenna spacing, e.g., Vertical Bell Labs Layered Space-Time (V-BLAST) [3]. More recently, hybrid MIMO transmission schemes (HMTS) [4] have been proposed to achieve a finer trade off between pure diversity gain and pure SM gain so that parts of the data are space-time coded across some antennas and others are spatially multiplexed. A number of the HMTS have been included in the recent WiFi (IEEE 802.11n-2009 [5]) and WiMAX (IEEE 802.16-2009 [6]) standards.

Among the HMTS that are described in [4], this paper addresses a special family of HMTS called the Alamouti-Code-Assisted-VBLAST (ACAV) – a term coined by Zhang *et al* [7]. Two of its  $M_t$  transmit antennas are assigned to transmit  $2 \times 2$  Alamouti STBC data streams while the remaining  $M_t - 2$  antennas are used to transmit independent V-BLAST SM data streams. The ACAV is equivalent to HMTS G2+1 for 3 antenna and G2+1+1 for 4 antenna in

Table I  
ANTENNA MAPPING FOR THE SPACE/TIME CODE-RATE 3 ACAV USING FOUR TRANSMIT ANTENNAS [5]

Rate-3 (6 symbols in 2 symbol periods)				
Time Slot	Ant 1	Ant 2	Ant 3	Ant 4
1	$a_1(k)$	$a_2(k)$	$a_3(k)$	$a_4(k)$
2	$-a_2^*(k)$	$a_1^*(k)$	$a_5(k)$	$a_6(k)$

Table II  
ANTENNA MAPPING FOR SPACE/TIME CODE-RATE 1, 2 AND 4 TRANSMIT SCHEMES OF IEEE 802.16E-2005 [8]

Rate-1				
Time Slot	Ant 1	Ant 2	Ant 3	Ant 4
1	$a_1(k)$	$a_2(k)$	-	-
2	$-a_2^*(k)$	$a_1^*(k)$	-	-
3	-	-	$a_3(k)$	$a_4(k)$
4	-	-	$-a_4^*(k)$	$a_3^*(k)$

Rate-2				
Time Slot	Ant 1	Ant 2	Ant 3	Ant 4
1	$a_1(k)$	$a_2(k)$	$a_3(k)$	$a_4(k)$
2	$-a_2^*(k)$	$a_1^*(k)$	$-a_4^*(k)$	$a_3^*(k)$
3	$a_5(k)$	$a_6(k)$	$a_7(k)$	$a_8(k)$
4	$-a_6^*(k)$	$a_5^*(k)$	$-a_8^*(k)$	$a_7^*(k)$

Rate-4				
Time Slot	Ant 1	Ant 2	Ant 3	Ant 4
1	$a_1(k)$	$a_2(k)$	$a_3(k)$	$a_4(k)$

[4]. The 4 antenna ACAV delivers 3 spatial substreams [5] since 6 data symbols are transmitted over 2 symbol periods (see Table I). Alternatively, the ACAV is said to achieve a space/time code-rate of 3 [8]. In fact the older IEEE 802.16e-2005 standard [8] only outlined HMTS of rates-1, -2 and -4 for a  $4 \times 4$  MIMO system (see Table II), so the rate-3 ACAV bridges the gap in space/time code-rate.

The IEEE standards do not define the type of data detectors to be used for each HMTS. In the literature, two types of ACAV detectors have been reported and they can be loosely categorized as single stage and dual stage detectors. The single stage detector includes the likes of the linear detector (Lin) and the ordered-successive-interference-cancelation (OSIC)<sup>1</sup> non-linear detector, e.g., [9]. These detectors operate on an *inflated* channel matrix of dimension  $2M_r \times 2(M_t - 1)$ , rather than the original  $M_r \times M_t$  matrix,

<sup>1</sup>Even though the OSIC involves several sequential stages or iterations, we still consider the procedure as a single stage.

in order to effectively exploit the diversity gain of the STBC which spans two symbol periods. This increased matrix size imposes approximately exponential burden on detector complexity for every extra matrix column. The inflated channel matrix can be avoided to reduce complexity but not without sacrificing the bit-error-rate (BER) performance. This can be done through the use of a dual stage detector which first detects the Alamouti symbols (henceforth called A-symbols, and the V-BLAST symbols as V-symbols) using a sort of MIMO spatial filter of dimension  $M_t \times M_r$ , and then cancel its interference, before detecting the remaining V-symbols in the second stage, for e.g., [4].

In Section II, the system model is outlined and the architecture of the ACAV briefly reviewed. Section III outlines our proposed reduced complexity single-stage ACAV detector. Section IV describes another reduced complexity dual-stage detector. Section V presents complexity and simulation results. The paper is finally concluded in Section VI.

## II. SYSTEM MODEL & THE ACAV ARCHITECTURE

In this paper, we consider the  $3 \times 3$  and  $4 \times 4$  MIMO systems in accordance to IEEE 802.16e-2005 [8] and IEEE 802.11n-2009 [5]. The antenna mappings for a block of 6 ACAV symbols are given in Table I. The source symbols are drawn independently from the same alphabet set and grouped into “space-time blocks” of  $2(M_t - 1)$  symbols. The first two symbols  $a_1(k)$  and  $a_2(k)$  are Alamouti encoded and the rest are spatially multiplexed (V-BLAST). The transmission patterns of rate-1, -2 and -4 schemes of IEEE 802.16e-2005 are provided in Table II for comparison.

At the channel output, the received vector is governed by the following channel input-output relationship

$$\mathbf{r}(k) = \mathbf{H}\mathbf{a}(k) + \mathbf{n}(k) \quad (1)$$

where  $k$  denotes a space-time block sample occupying two symbol periods,  $\mathbf{r}(k) \triangleq [\mathbf{r}_1(k), \mathbf{r}_2(k)] \in \mathbb{C}^{M_r \times 2}$ ,  $\mathbf{H} \triangleq [\mathbf{H}_A, \mathbf{H}_V] = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_{M_t}] \in \mathbb{C}^{M_r \times M_t}$  is the Rayleigh flat fading channel (N.B.:  $\mathbf{H}_A = [\mathbf{h}_1, \mathbf{h}_2] \in \mathbb{C}^{M_r \times 2}$  denotes the Alamouti subchannels,  $\mathbf{H}_V = [\mathbf{h}_3, \dots, \mathbf{h}_{M_t}] \in \mathbb{C}^{M_r \times (M_t - 2)}$  denotes the V-BLAST subchannels),  $\mathbf{a}(k) \in \mathbb{C}^{M_t \times 2}$  denotes the source signals, whose energy is  $\sigma_a^2$ , in the manner shown in Table I and  $\mathbf{n}(k) \in \mathbb{C}^{M_r \times 2}$  is the additive white Gaussian noise (AWGN) matrix with a complex variance of  $\sigma_n^2$ . One way of data detection is to process the received vector  $\mathbf{r}(k)$  using a MIMO spatial filter to extract the A-symbols because they are usually more reliable, and subsequently perform interference cancelation to detect the remaining V-symbols [4]. We refer to this method as the dual-stage detector which we pursue in Section IV. To achieve better BER suppression, we should consider the inflated channel matrix,  $\tilde{\mathbf{H}}$ , defined below, which takes into account all  $2(M_t - 1)$  symbols so that the transmit diversity can be more effectively exploited [9], so that detection is

carried out in a single stage:

$$\tilde{\mathbf{r}}(k) = \tilde{\mathbf{H}}\tilde{\mathbf{a}}(k) + \tilde{\mathbf{n}}(k) \Rightarrow \begin{bmatrix} \mathbf{r}_1(k) \\ \mathbf{r}_2^*(k) \end{bmatrix} = \begin{bmatrix} \mathbf{H}_A & \mathbf{H}_V & \mathbf{0}_2 \\ \mathcal{A}(\mathbf{H}_A) & \mathbf{0}_2 & \mathbf{H}_V^* \end{bmatrix} \tilde{\mathbf{a}}(k) + \begin{bmatrix} \mathbf{n}_1(k) \\ \mathbf{n}_2^*(k) \end{bmatrix} \quad (2)$$

where the “Alamouti” operator  $\mathcal{A}([\mathbf{h}_1, \mathbf{h}_2]) \triangleq [\mathbf{h}_1^*, -\mathbf{h}_1]$ ,  $\mathbf{0}_2$  is a zero matrix of dimension  $M_r \times 2$  and  $\tilde{\mathbf{a}}(k) = [a_1(k), a_2(k), a_3(k), a_4(k), a_5^*(k), a_6^*(k)]^T$ . The special structure inherent in  $\tilde{\mathbf{H}}$  is the motivation behind the design of the low-complexity pair-wise OSIC detector in the following Section III.

## III. REDUCED COMPLEXITY SINGLE STAGE DETECTOR: PAIR-WISE OSIC (PWO)

In the following we propose a modified OSIC scheme which decodes twice as fast, uses less computations, but attains identical BER performance as the original OSIC for ACAV [3].

### A. Data Detection in Pairs

Let's consider only the zero-forcing (ZF) detector (extension to the MMSE detector is straight forward). Let  $\tilde{\mathbf{G}}$  denote the Moore-Penrose pseudoinverse of  $\tilde{\mathbf{H}}$ . Subject to the condition that the matrix  $\tilde{\mathbf{G}}$  must have an even number of rows, the pseudoinverse has this specific form:

$$\tilde{\mathbf{G}} = \tilde{\mathbf{H}}^\dagger = \begin{bmatrix} \mathbf{G}_A & \mathcal{A}(\mathbf{G}_A) \\ \mathbf{G}_{V1} & -\mathbf{G}_{V2}^* \\ \mathbf{G}_{V2} & \mathbf{G}_{V1}^* \end{bmatrix} \quad (3)$$

where  $\mathbf{G}_A, \mathbf{G}_{V1}, \mathbf{G}_{V2}$  are all  $2 \times M_r$  matrices. Expanding (3), it can be shown that the row-norms of certain subchannels  $\tilde{\mathbf{G}}$  are identical, since all elements of one row have counterparts of same magnitude in one other row. They are the 1st and 2nd row, 3rd and 5th row, and finally the 4th and 6th row. Since the row-norms are identical for the two subchannels within the pair, that means both symbols that belong to the equal-norm subchannels can be detected simultaneously *without any loss in optimality* at all.

### B. Optimal Sorting with Detection Speed Doubled

According to [3], the optimal sorting is based on selecting first the subchannel which has the minimum norm. The A-symbols are most robust to noise and therefore with high probability, their corresponding subchannels in  $\tilde{\mathbf{G}}$  have minimum norms. This is because their corresponding columns in  $\tilde{\mathbf{H}}$  (c.f. Eq. (2)) consist of only non-zero elements, i.e.,  $[\mathbf{H}_A \ \mathcal{A}(\mathbf{H}_A)]^T$ , while the columns of the other V-subchannels are occupied half of the time by 0, i.e.,  $[\mathbf{H}_V \ \mathbf{0}_2]^T$  and  $[\mathbf{0}_2 \ \mathbf{H}_V^*]^T$ . The robustness of the A-symbols is a consequence of STBC of  $a_1(k)$  and  $a_2(k)$ . However, optimality is not guaranteed if the A-symbols were detected and canceled first because one of the V-subchannels may have a higher SNR than the A-subchannel for a

Table III  
PAIR-WISE ORDERED SEQUENTIAL DETECTION

---

Let the  $j$ -th row of  $\tilde{\mathbf{G}}$  be denoted as  $(\tilde{\mathbf{G}})_j$ .  
 Let  $k_i(1)$  and  $k_i(2)$  denote the indices of the 1st and 2nd symbols,  
 respectively, of the symbol-pair with equal SNR, at the  $i$ -th iteration/  
 OSIC stage.  
 $\mathbb{Q}(\cdot)$  is the nearest neighbour hard quantization operator.  
 $\tilde{\mathbf{H}}^{(i)}$  is the deflated channel matrix of  $\tilde{\mathbf{H}}$  at the  $i$ -th iteration.

---

**Initialization:**  
 $i = 1$   
 $\tilde{\mathbf{r}}_1(k) = \tilde{\mathbf{r}}(k)$   
 $\tilde{\mathbf{G}}_1 = (\tilde{\mathbf{H}})^\dagger$

**Recursion:**

$$\mathbf{w}_i^{zf} = \begin{bmatrix} (\tilde{\mathbf{G}}_i)_{k_i(1)} \\ (\tilde{\mathbf{G}}_i)_{k_i(2)} \end{bmatrix}$$

$$\begin{bmatrix} \tilde{z}_{k_i(1)}(k) \\ \tilde{z}_{k_i(2)}(k) \end{bmatrix} = \mathbf{w}_i^{zf} \tilde{\mathbf{r}}_i(k)$$

$$\begin{bmatrix} \hat{a}_{k_i(1)}(k) \\ \hat{a}_{k_i(2)}(k) \end{bmatrix} = \begin{bmatrix} \mathbb{Q}(\tilde{z}_{k_i(1)}(k)) \\ \mathbb{Q}(\tilde{z}_{k_i(2)}(k)) \end{bmatrix}$$

$$\tilde{\mathbf{r}}_{i+1}(k) = \tilde{\mathbf{r}}_i(k) - [\mathbf{h}_{k_i(1)}, \mathbf{h}_{k_i(2)}] \begin{bmatrix} \hat{a}_{k_i(1)}(k) \\ \hat{a}_{k_i(2)}(k) \end{bmatrix}$$

$$\tilde{\mathbf{G}}_{i+1} = (\tilde{\mathbf{H}}^{(i)})^\dagger$$

$$i = i + 1$$


---

particular channel realization. Thus, to retain optimality, all the row norms of  $\tilde{\mathbf{G}}$  are still calculated just in case the A-subchannel is not the strongest. Subsequently the symbols, be it A- or V-, are detected and canceled according to the OSIC algorithm of [3]. This interference cancellation (IC) process is performed on the symbol-pair simultaneously so the number of sequential IC stages is halved. Table III summarizes this pair-wise OSIC (PWO) algorithm.

#### IV. REDUCED COMPLEXITY DUAL STAGE DETECTOR: GROUP-LINEAR (GL) AND GROUP-OSIC (GO)

Both the PWO algorithm in Section III and the OSIC of [9] deal with an inflated channel matrix  $\tilde{\mathbf{H}}$ . Trading off BER performance slightly to enjoy lower pre-processing burden (e.g., the pseudo-inverse operation), a dual stage detector may be employed to detect the ACAV symbols by using only the  $M_r \times M_t$  channel matrix. In the first stage, the A-symbols are detected using a group receiver [10], followed by interference cancelation (IC) and the detection of the V-symbols using a linear or OSIC detector. Thus we name our dual-stage detectors as the Group-Linear (GL) and the Group-OSIC (GO) detectors. They are especially beneficial in rapidly time-varying channels when the pre-processing cost is significant as compared to the cost of payload processing (i.e., computation required to process every symbol of frame).

#### A. Stage-1: Zero-forcing (ZF) Group Receiver

The ZF group receiver partitions the original channel matrix into four quadrants as follows [11]:

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \quad (4)$$

where  $\mathbf{A} \in \mathbb{C}^{2 \times 2}$ ,  $\mathbf{B} \in \mathbb{C}^{2 \times (M_t - 2)}$ ,  $\mathbf{C} \in \mathbb{C}^{(M_r - 2) \times 2}$ , and  $\mathbf{D} \in \mathbb{C}^{(M_r - 2) \times (M_t - 2)}$ .  $\mathbf{A}$  always carries the top left entries of  $\mathbf{H}$  of dimension  $2 \times 2$ , i.e.,

$$\mathbf{A} \triangleq \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \quad (5)$$

where  $h_{ij}$  is the element in the  $i$ -th row and  $j$ -th column of  $\mathbf{H}$ . Subsequently, ZF group equalization is performed on  $\mathbf{r}(k)$ . The ZF group equalizer takes the following form:

$$\mathbf{W}^{zf} = \begin{bmatrix} \mathbf{B}^{-1} & -\mathbf{D}^{-1} \\ \mathbf{A}^{-1} & -\mathbf{C}^{-1} \end{bmatrix} \quad (6)$$

so that the combined channel-group-equalizer response,

$$\mathbf{W}^{zf} \mathbf{H} = \begin{bmatrix} \mathbf{B}^{-1} \mathbf{A} - \mathbf{D}^{-1} \mathbf{C} & \mathbf{0}_{2r} \\ \mathbf{0}_{c2} & \mathbf{A}^{-1} \mathbf{B} - \mathbf{C}^{-1} \mathbf{D} \end{bmatrix} \quad (7)$$

is a diagonal matrix that isolates the two A-substreams from the V-substreams.  $\mathbf{0}_{2r}$  and  $\mathbf{0}_{c2}$  are zero matrices of dimensions  $2 \times (M_t - 2)$  and  $(M_t - 2) \times 2$ , respectively. As a result, the virtual channel of the A-substreams due to the ZF group receiver is

$$\tilde{\mathbf{H}}_A \triangleq \mathbf{B}^{-1} \mathbf{A} - \mathbf{D}^{-1} \mathbf{C} \quad (8)$$

where  $\tilde{\mathbf{H}}_A \in \mathbb{C}^{(M_t - 2) \times 2}$ . The received signal after the group receiver can be expressed as

$$\mathbf{s}(k) \triangleq \begin{bmatrix} \mathbf{s}_A(k) \\ \mathbf{s}_V(k) \end{bmatrix} = \mathbf{W}^{zf} \mathbf{r}(k) = \mathbf{W}^{zf} \mathbf{H} \mathbf{a}(k) + \tilde{\mathbf{n}}(k), \quad (9)$$

where  $\mathbf{s}_A(k) \in \mathbb{C}^{2 \times 2}$ ,  $\mathbf{s}_V(k) \in \mathbb{C}^{(M_t - 2) \times 2}$  and  $\tilde{\mathbf{n}}(k) \in \mathbb{C}^{M_t \times 2}$  is the filtered noise matrix. We have deliberately divided  $\mathbf{s}(k)$  into two matrices because we only want to retrieve  $\mathbf{s}_A(k)$  that corresponds to the Alamouti stream, i.e.,

$$\mathbf{s}_A(k) = [\mathbf{B}^{-1}, -\mathbf{D}^{-1}] \mathbf{r}(k). \quad (10)$$

Using  $\mathbf{s}_A(k)$  from (10), the optimal maximum likelihood soft decisions of the A-symbols are obtained as follows

$$\begin{bmatrix} \tilde{z}_1(k) \\ \tilde{z}_2^*(k) \end{bmatrix} = \left[ \|\tilde{\mathbf{H}}_A\|_F^2 \right]^{-1} (\tilde{\mathbf{H}}_{A2})^H \tilde{\mathbf{s}}(k) \quad (11)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm,  $(\cdot)^H$  denotes the Hermitian operation (complex conjugate transpose),

$$\tilde{\mathbf{H}}_{A2} \triangleq \begin{bmatrix} \tilde{\mathbf{H}}_A \\ \mathcal{A}(\tilde{\mathbf{H}}_A) \end{bmatrix} \quad (12)$$

where  $\mathcal{A}(\mathbf{x}) \triangleq \mathcal{A} \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} b^* & -a^* \\ d^* & -c^* \end{bmatrix}$ , and  $\tilde{\mathbf{s}}(k) \triangleq [s_{11}(k), s_{12}^*(k)]^T$  if  $M_t = M_r = 3$  where  $\mathbf{s}_A(k) = [s_{11}(k), s_{12}(k)]$ , or  $\tilde{\mathbf{s}}(k) \triangleq [s_{11}(k), s_{21}(k), s_{12}^*(k), s_{22}^*(k)]^T$

if  $M_t = M_r = 4$  where  $\mathbf{s}_A(k) \equiv \begin{bmatrix} s_{11}(k) & s_{12}(k) \\ s_{21}(k) & s_{22}(k) \end{bmatrix}$ . Subsequently, a nearest neighbour quantizer  $\mathbb{Q}(\cdot)$  is used to estimate the A-symbols:  $[\hat{a}_1(k), \hat{a}_2(k)]^T = [\mathbb{Q}(\hat{z}_1(k)), \mathbb{Q}(\hat{z}_2(k))]^T$ .

After the A-symbols have been detected, they will be canceled from the received signals,  $\mathbf{r}(k)$ , as follows:

$$\mathbf{t}(k) \triangleq [\mathbf{t}_1(k) \ \mathbf{t}_2(k)] = \mathbf{r}(k) - [\mathbf{h}_1 \ \mathbf{h}_2] \begin{bmatrix} \hat{a}_1(k) & -\hat{a}_2^*(k) \\ \hat{a}_2(k) & \hat{a}_1^*(k) \end{bmatrix}. \quad (13)$$

$\mathbf{t}_1(k) \in \mathbb{C}^{M_t \times 1}$  and  $\mathbf{t}_2(k) \in \mathbb{C}^{M_t \times 1}$  are the column vectors that correspond to the V-symbols of the first and second time slot in block  $k$ , respectively. If the decisions  $\hat{a}_1(k)$  and  $\hat{a}_2(k)$  are correct, then  $\mathbf{t}(k)$  will enjoy a large diversity gain since the the V-symbols are effectively being transmitted from  $M_t - 2$  antennas but are received by two more receive antennas! Moreover, the signals from the first time slot of space-time block  $k$  are now independent of those signals from the second time slot, so the detection of V-symbols is carried out with a much smaller deflated channel matrix *one time slot at a time*.

### B. Stage-2: Linear detection or OSIC

In stage 2, we can use either the linear or the OSIC algorithm. The effective channel affecting the remaining V-symbols is  $\mathbf{H}_V = [\mathbf{h}_3]$  for the  $3 \times 3$  ACAV, or  $\mathbf{H}_V = [\mathbf{h}_3 \ \mathbf{h}_4]$  for the  $4 \times 4$  ACAV. Both the linear detection algorithm and the first step of the OSIC algorithm [3] make use of either the same ZF filter  $\mathbf{G}_V^{(\text{ZF})}$  or the MMSE filter  $\mathbf{G}_V^{(\text{MMSE})}$ . These filters are the unbiased and biased pseudo-inverses of  $\mathbf{H}_V$ , respectively. Since  $\mathbf{H}_V$  is a ‘‘tall’’ or ‘‘slim’’ matrix with dimension  $M_r \times (M_t - 2)$ , computing the ZF or the MMSE filter is significantly simpler than it is with the ACAV channel matrix  $\tilde{\mathbf{H}}$  of dimension  $2M_r \times 2(M_t - 1)$  as described in [9]. In the pursuit of computational simplicity, the Greville algorithm [12] and the Sherman-Morrison algorithm [13] are adopted to compute  $\mathbf{G}_V^{(\text{ZF})}$  and  $\mathbf{G}_V^{(\text{MMSE})}$ , respectively. In this way, we can evaluate the exact computation required for the GL and GO algorithms.

1) *ZF Detectors for  $3 \times 3$  ACAV*: The V-symbols from the first and the second time slots in block  $k$ , i.e.,  $a_3(k)$  and  $a_4(k)$ , are detected as follows:

$$\mathbf{G}_V^{(\text{ZF})} \triangleq \mathbf{g}_3 = \mathbf{h}_3^\dagger = \frac{\mathbf{h}_3^H}{\|\mathbf{h}_3\|^2} \quad (14)$$

$$\hat{a}_3(k) = \mathbb{Q}(\mathbf{g}_3 \mathbf{t}_1(k)), \quad \hat{a}_4^*(k) = \mathbb{Q}(\mathbf{g}_3 \mathbf{t}_2(k)) \quad (15)$$

where  $\dagger$  denotes the Moore-Penrose pseudo-inverse operator.

2) *ZF Detectors for  $4 \times 4$  ACAV*: The pseudo-inverse of  $\mathbf{H}_V = [\mathbf{h}_3 \ \mathbf{h}_4]$  can be computed using the Greville algorithm [12] as follows:

$$\mathbf{G}_V^{(\text{ZF})} \triangleq \mathbf{H}_V^\dagger = \begin{bmatrix} \mathbf{g}_3 \\ \mathbf{g}_4 \end{bmatrix} = \begin{bmatrix} \mathbf{h}_3^\dagger - d(\mathbf{h}_4 - d\mathbf{h}_3)^\dagger \\ (\mathbf{h}_4 - d\mathbf{h}_3)^\dagger \end{bmatrix} \quad (16)$$

where  $d = \mathbf{h}_3^\dagger \mathbf{h}_4$  and  $d$  is a scalar. The linear detector decodes the V-symbols as follows:

$$\begin{aligned} \hat{a}_3(k) &= \mathbb{Q}(\mathbf{g}_3 \mathbf{t}_1(k)), & \hat{a}_4(k) &= \mathbb{Q}(\mathbf{g}_4 \mathbf{t}_1(k)) \\ \hat{a}_5^*(k) &= \mathbb{Q}(\mathbf{g}_3 \mathbf{t}_2(k)), & \hat{a}_6^*(k) &= \mathbb{Q}(\mathbf{g}_4 \mathbf{t}_2(k)). \end{aligned} \quad (17)$$

The OSIC detector extracts the V-symbols in sequential steps as follows. Assuming  $\|\mathbf{g}_3\|^2 < \|\mathbf{g}_4\|^2$ , then

$$\hat{a}_3(k) = \mathbb{Q}(\mathbf{g}_3 \mathbf{t}_1(k)) \quad (18)$$

$$\mathbf{u}_1(k) = \mathbf{t}_1(k) - \hat{a}_3(k) \mathbf{h}_3 \quad (19)$$

$$\hat{a}_4(k) = \mathbb{Q}(\mathbf{h}_4^\dagger \mathbf{u}_1(k)). \quad (20)$$

If, on the other hand, the SNR is stronger at the 4<sup>th</sup> channel, i.e.,  $\|\mathbf{g}_4\|^2 < \|\mathbf{g}_3\|^2$ , then

$$\hat{a}_4(k) = \mathbb{Q}(\mathbf{g}_4 \mathbf{t}_1(k)) \quad (21)$$

$$\mathbf{u}_1(k) = \mathbf{t}_1(k) - \hat{a}_4(k) \mathbf{h}_4 \quad (22)$$

$$\hat{a}_3(k) = \mathbb{Q}(\mathbf{h}_3^\dagger \mathbf{u}_1(k)). \quad (23)$$

Independent of the first time slot,  $\hat{a}_5^*(k)$  and  $\hat{a}_6^*(k)$  of the second time slot are obtained in the same fashion as (18)–(23) above.

3) *MMSE Detectors for  $3 \times 3$  ACAV*: The MMSE filter of  $\mathbf{H}_V$  is

$$\mathbf{G}_V^{(\text{MMSE})} \triangleq \mathbf{g}_3 = \frac{\mathbf{h}_3^H}{\|\mathbf{h}_3\|^2 + \alpha} \quad (24)$$

where  $\alpha$  is the ‘‘regularization’’ constant chosen as the ratio of the noise variance to the signal power (of individual transmit antenna) [13]. Subsequently,  $\hat{a}_3(k)$  and  $\hat{a}_4^*(k)$  are detected in the same manner as (15).

4) *MMSE Detectors for  $4 \times 4$  ACAV*: The Sherman-Morrison recursion algorithm [13] is used to compute the MMSE filter

$$\begin{aligned} \mathbf{G}_V^{(\text{MMSE})} &\triangleq \begin{bmatrix} \mathbf{g}_3 \\ \mathbf{g}_4 \end{bmatrix} = \mathbf{H}_V^H [\mathbf{H}_V^H \mathbf{H}_V + \alpha \mathbf{I}_{2 \times 2}]^{-1} \\ &= \mathbf{H}_V^H \mathbf{Q}, \end{aligned} \quad (25)$$

where  $\mathbf{I}_{2 \times 2}$  is the  $2 \times 2$  square identity matrix. The Sherman-Morrison algorithm computes  $\mathbf{Q}$  in a simple, recursive manner. It starts from the zero-th iteration:

$$\mathbf{Q}_{[0]} = (1/\alpha) \mathbf{I}_{2 \times 2} \quad (26)$$

and then computes the recursion from  $n = 1$  to  $n = M_r = 4$ :

$$\mathbf{Q}_{[n]} = \mathbf{Q}_{[n-1]} - \frac{\mathbf{Q}_{[n-1]} \mathbf{H}_{V,n} (\mathbf{H}_{V,n}^H \mathbf{Q}_{[n-1]})^H}{1 + (\mathbf{H}_{V,n}^H \mathbf{Q}_{[n-1]} \mathbf{H}_{V,n})} \quad (27)$$

where  $\mathbf{H}_{V,n}$  is the  $n$ -th column of the ‘‘tall’’ matrix  $\mathbf{H}_V$ . The V-symbols are detected in the same manner as (17) for linear detection and (18)–(20) for OSIC with the exception of the pseudo-inverses in (20) and (23) are replaced with the MMSE filter computed in the same fashion as (24).

Table IV  
 COMPLEXITY ANALYSIS FOR  $4 \times 4$  RATE-3 MIMO DETECTORS

Detector ( $N \times M$ )	ZF		MMSE	
	/Frame (Pre- processing) ( $M, \mathbf{A}$ )	/6-Symbols (Payload) ( $M, \mathbf{A}$ )	/Frame (Pre- processing) ( $M, \mathbf{A}$ )	/6-Symbols (Payload) ( $M, \mathbf{A}$ )
Lin( $8 \times 6$ )	(672, 593.5)	(48, 42)	(840, 606)	(48, 42)
OSIC( $8 \times 6$ )	(792, 705.5)	(48, 80)	(955, 649)	(229, 202)
PWO( $8 \times 6$ )	(716, 639)	(48, 80)	(859, 578)	(229, 202)
GL( $4 \times 4$ )	(32, 17.5)	(56, 46)	(76, 55.5)	(56, 46)
GO( $4 \times 4$ )	(36, 19)	(56, 54)	(80, 46.5)	(56, 54)
V-L( $4 \times 3$ )	(72, 54.5)	(24, 18)	(90, 79.5)	(24, 18)
V-O( $4 \times 3$ )	(98, 76)	(24, 34)	(106, 92.5)	(58, 58)

## V. RESULTS

We shall assume the Rayleigh fading channel is stationary over  $N_{\text{fr}}$  space-time block samples, or equivalently,  $2N_{\text{fr}}$  symbol periods.

### A. Complexity Analysis

All detectors under consideration achieve the same data throughput, i.e., 3 spatial substreams, and they all have equal number of received antennas, i.e.,  $M_r = 4$ . The ACAV detectors under consideration include the linear detector (Lin), OSIC, pair-wise OSIC (PWO), Group-Linear (GL), and Group-OSIC (GO). The detectors for pure V-BLAST signals are the linear and OSIC detectors, denoted as V-L and V-O, respectively. The dimensions of the channel matrices that the detectors deal with are appended to the names of the detectors, as shown in Table IV, where  $M$  and  $N$  correspond to the number of effective transmit and received symbols, respectively. We consider both ZF and MMSE implementations of the detectors, and separated the pre-processing (e.g., computing the pseudoinverse of the channel matrix in (16) and (25)) from the payload processing (e.g., linear convolution of equalizer weights and received signal vector in (18) and the subsequent nulling and canceling operations in (19)). The effective channel matrices that need to be pseudoinversed have dimensions  $8 \times 6$ ,  $4 \times 2$ , and  $4 \times 3$  for the single-stage, dual-stage, and V-BLAST( $4 \times 3$ ) detectors, respectively. After consulting with [12] and [13], we arrived at the following number of pre-processing operations required for the *initialization*, i.e., pseudoinverse of channel matrix, and the *recursion*, i.e., pseudoinverses of subsequent deflated channel matrices, and the number of repetitive payload operations required for every space-time block of  $N$  received symbols to compute the linear convolution (e.g., (17), (18)) and the nulling operations (e.g., (19)). For a channel with  $M$  and  $N$  effective transmit and received symbols, respectively,

#### 1) the Greville-Inverse-Greville algorithm [12]

requires  $(\frac{5}{2}M^2N - \frac{1}{2}MN - 3N)\mathbf{M} + (\frac{5}{2}M^2N - M^2 - \frac{3}{2}MN - \frac{5}{2}N + \frac{3}{2})\mathbf{A}$  operations for the “initialization”, while requiring  $(\frac{1}{4}M^2N + \frac{1}{4}MN - \frac{1}{2}N)\mathbf{M} + (\frac{1}{4}M^2N - \frac{1}{4}M^2 + \frac{1}{4}MN - \frac{1}{4}M -$

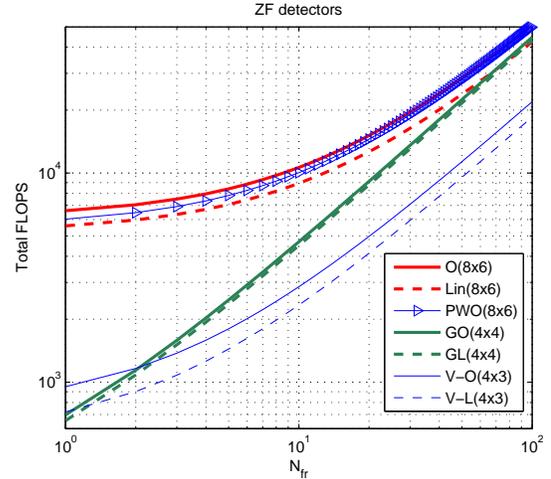


Figure 1. Complexity profile of various detectors with varying  $N_{\text{fr}}$ . Total FLOPS = FLOPS/frame +  $N_{\text{fr}}$ \*FLOPS/6-symbols.

$\frac{1}{2}N + \frac{1}{2})\mathbf{A}$  for the “recursion”. The computations required for payload processing for a block of  $N$  received samples is  $(MN)\mathbf{M} + (MN - M)\mathbf{A}$  for the linear detector and  $(2MN - N)\mathbf{M} + (2MN - M - N)\mathbf{A}$  for the OSIC detector.

#### 2) the Sherman-Morrison algorithm [13]

requires  $(\frac{5}{2}M^2N + \frac{5}{2}MN)\mathbf{M} + (2M^2N - \frac{1}{2}M^2 + MN)\mathbf{A}$  operations for the “initialization”, and  $(\frac{2}{3}M^3 - \frac{2}{3}M)\mathbf{M} + (\frac{1}{2}M^3 - \frac{1}{2}M^2 + M + 1)\mathbf{A}$  for the “recursion”. The computation required for every  $N$  received data sample is  $(MN)\mathbf{M} + (MN - M)\mathbf{A}$  for the linear detector and  $(\frac{1}{2}M^2N + \frac{1}{2}M^2 - \frac{1}{2}MN - \frac{1}{2}M)\mathbf{M} + (\frac{1}{2}M^2N - \frac{1}{2}MN - M + 1)\mathbf{A}$  for the OSIC detector.

The PWO requires less complexity than the original OSIC( $8 \times 6$ ). More significantly, the PWO detects 6 symbols in 3 sequential IC stages instead of 6, thus the detection speed is doubled. The GL and GO detectors require the least pre-processing and are ideal for rapidly varying channels.

We now investigate the effect of increasing  $N_{\text{fr}}$  on the complexity of the detector. Each complex multiplication  $\mathbf{M}$  requires 6 floating point operations (FLOPs); each complex addition  $\mathbf{A}$  requires 2 FLOPs. Although counting FLOPs is not the ideal way to determine the complexity of a detector, it does provide a certain degree of appreciation of the complexity. We compute the total FLOPs by summing the FLOPs per frame and the FLOPs per 6-symbols multiplied by  $N_{\text{fr}}$ . For small frame sizes, the dual-stage detectors are about 10 times simpler than the traditional ACAV detectors, making them very attractive.

### B. Simulation Results

The six detectors are put to test (NB: the performance of OSIC( $8 \times 6$ ) and PWO( $8 \times 6$ ) are identical). For simplicity,

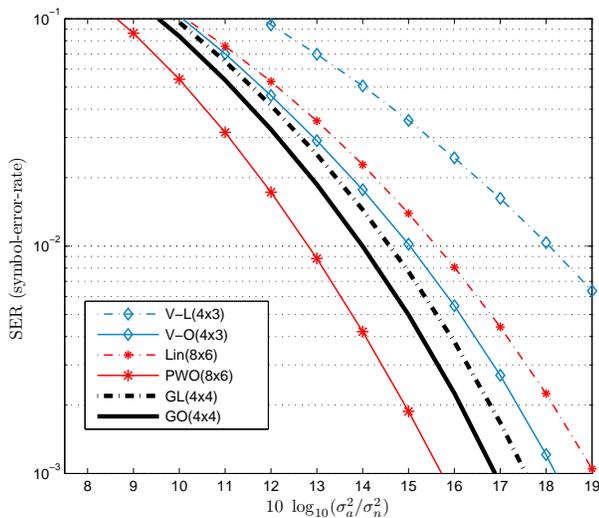


Figure 2. Simulation results comparing the ZF ACAV detectors and V-BLAST  $4 \times 3$  detectors.

all detectors employ the ZF approach. The modulation scheme is 16-QAM and for each simulation point, 10 million symbols over 100 different random Rayleigh faded channel realizations were used. The result is plotted in Fig. 2.

It is clear that without STBC, the V-BLAST detectors perform poorly against the ACAV detectors. Only the linear ACAV detector performs about 0.5dB worse than the V-O.

As for ACAV detectors, PWO performs the best. The second best is the GO detector with a 1dB gap with the PWO. Following closely to the GO is the GL detector. Interestingly, even the GL performs better than V-O, suggesting the advantage of using the ACAV transmission scheme.

## VI. CONCLUSION AND FUTURE WORK

The ACAV has been adopted by IEEE 802.11n-2009 as viable candidates of its hybrid MIMO transmission schemes [5]. Through our simulations we discovered the superiority of ACAV over pure V-BLAST schemes of identical data rate even though similar OSIC or linear detectors were used. That means the ACAV outperforms the V-BLAST as a transmission scheme, and not because of the detector used. In the future, we intend to prove that the ACAV is naturally more resilient to system errors than pure V-BLAST, and perhaps double-space-time-transmit-diversity (D-STTD) [14] too, when the same data rate is enforced. If this is true, the ACAV could replace pure V-BLAST schemes in future broadband modems (WiFi, WiMAX, LTE). Unfortunately the ACAV detectors are relatively complex, so we developed two simple detection schemes in this paper to address this issue.

## REFERENCES

- [1] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [2] V. Tarokh, H. JafarKhani, and A. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Information Theory*, vol. 45, pp. 1456–1467, Jul 1999.
- [3] P. Wolniansky, G. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel," *Signals, Systems, and Electronics, 1998. ISSSE 98. 1998 URSI International Symposium on*, pp. 295–300, 1998.
- [4] W. da C. Freitas Jr., F. R. P. Cavalcanti, and R. R. Lopes, "Hybrid transceiver schemes for spatial multiplexing and diversity in MIMO systems," *Journal of Communication and Information System*, vol. 20, no. 3, pp. 63–76, 2005.
- [5] *IEEE Std 802.11n-2009*, (Last assessed 22 Feb 2011).
- [6] *IEEE Std 802.16-2009*, (Last assessed 22 Feb 2011).
- [7] L. Zhang, S. Li, H. Zheng, and M. Wu, "Alamouti code assisted V-BLAST (ACAV): A new space-time architecture," *Communication Technology, 2006. ICCT '06. International Conference on*, pp. 1–4, 2006.
- [8] *IEEE Std 802.16e-2005*, (Last assessed 22 Feb 2011).
- [9] M. Do, W. Chin, and Y. Wu, "Performance study of a hybrid space time block coded system," *Wireless Pervasive Computing, 2006 1st International Symposium on*, 2006.
- [10] L. Zhao and V. K. Dubey, "Transmit diversity and combining scheme for spatial multiplexing over correlated channels," *IEEE 59th Proc. Veh. Technol. Conf.*, vol. 1, pp. 380–383, Spring 2004.
- [11] —, "Detection schemes for space-time block code and spatial multiplexing combined system," *Communications Letters, IEEE*, vol. 9, no. 1, pp. 49–51, Jan 2005.
- [12] Z. Luo, M. Zhao, S. Liu, and Y. Liu, "Greville-to-Inverse-Greville algorithm for V-BLAST systems," *Communications, 2006. ICC '06. IEEE International Conference on*, pp. 4214–4218, 2006.
- [13] J. Benesty, Y. A. Huang, and J. Chen, "A fast recursive algorithm for optimum sequential signal detection in a BLAST system," *IEEE Trans. Signal Processing*, vol. 51, no. 7, pp. 1722–1730, Jul 2003.
- [14] S. Shim, K. Kim, and C. Lee, "An efficient antenna shuffling scheme for a DSTTD system," *IEEE Commun. Letters*, vol. 9, no. 2, pp. 124–126, Feb 2005.

# Radiation Pattern Behaviour of Reconfigurable Asymmetry Slotted Ultra Wideband Antenna

Yusnita Rahayu

Faculty of Mechanical Engineering  
Universiti Malaysia Pahang  
Kuantan, Pahang 26600, Malaysia  
Email: vannebula2001@yahoo.com

Azlyna Senawi

FIST  
Universiti Malaysia Pahang  
Kuantan, Pahang 26300, Malaysia  
Email: Azlyna@ump.edu.my

Razali Ngah

Faculty of Electrical Engineering  
Universiti Teknologi Malaysia  
Johor Bahru, 81310, Malaysia  
Email: razalin@fke.utm.my

**Abstract**-This paper presents the results of radiation pattern measurement of small reconfigurable slotted ultra wideband (UWB) antennas. The measurements were conducted by using RF measurement and instrumentation facilities, software tools available at WCC of Universiti Teknologi Malaysia. The original antenna's geometry proposed is slotted antenna with L and U slots. Then the slots modifications are applied in order to achieve band-notched characteristics. These proposed antennas are having band notched frequencies at Fixed Wireless Access (FWA), HIPERLAN and WLAN bands. The band-notched operation is achieved by incorporating some small gaps instead of PIN diodes into the slot antenna. It is found that by adjusting the total length of slot antenna to be about a half-wavelength or less at desired notched frequency [1-3], a destructive interference can take place, thus causing the antenna to be non-responsive at that frequency. It was also observed that the measured radiation patterns, H-planes, are omni-directional with slightly gain decreased at boresight direction for measured frequencies. There are also more ripples occurred in the measured pattern compared with the simulated one.

**Keywords**-component; antenna; antenna measurement; radiation pattern; ultra wideband; band-notched antenna.

## I. INTRODUCTION

This section reviews the concept of reconfigurable antennas that reuse their entire geometry for band-notched frequency applications. Various techniques used have been reported in literature [1-9].

In [1], there are two varieties of slotted antennas which have a frequency notched reported, a triangular notch and elliptical notch. Both antennas have frequency notch characteristics where the arc length of slots form a half wavelength resonance structure at particular frequency, thus a destructive interference takes place causing the antennas to be non-responsive at that frequency. Other types of this kind antenna was reported in [4], a band notched UWB antenna using a slot-type split ring resonator (SRR) was found very effective in rejecting unwanted frequency, such as that for WLAN service, in terms of its selectivity and small dimension. The SRR is composed of two concentric split ring slots and proposed for band stop application, since it provides high Q characteristic. The slotted SRR was positioned near the feeding point to provide more coupling

with the field. A multiple band-notched planar monopole antenna using multiple U-shape slots for multi band wireless system was also presented in [5]. The half wavelength U-shape slots were symmetrically inserted in the centre of the planar element. In order to generate the two band-notched characteristics, three U-shape slots were proposed. An alternative antenna design without using slot to obtain band-notched characteristic was proposed in [6]. The antenna consists of two same size monopoles and a small strip bar at the centre showing the band rejection performance in the desired frequency bands. More examples of reconfigurable antenna are available in [7-9].

In this paper, new proposed reconfigurable UWB antennas are designed by adopting the half wavelength slot structure techniques. This paper mainly focuses on reconfigurable notch band through the introduction of new slots, L and U slots, on patch antenna. The current distribution on patch surface is disturbed by the introduction of new slots, which is responsible for the notch in frequency band. Section II will discuss the antenna geometry and the techniques to design new reconfigurable slotted antennas. These proposed antennas are having band notched frequencies. The band-notched operation is achieved by incorporating some small gaps instead of PIN diodes into the slot antenna. The term of small gaps in this paper will refer to switches. The switches are used to short the slot in pre-selected positions along the circumference. The length of the slot antenna can be lengthened or shorted by closing or opening the switches, allowing for a change in the notched frequency. Then the performances of reconfigurable antennas, in terms of VSWR and radiation patterns, will be discussed and evaluated in Section III. Finally, summary will be given in Section IV.

## II. ANTENNA GEOMETRY

Fig. 1 shows the original proposed antenna structure printed on the FR4 substrate of  $\epsilon_r = 4.6$ . The pentagonal antenna is vertically installed above a ground plane ( $l_{\text{ground}}$ ) of 11 mm. The optimum feed gap (h) to the ground plane is found to be 1.5 mm. The dimension of substrate is chosen to be  $30 \times 30 \text{ mm}^2$  ( $W_{\text{sub}} \times L_{\text{sub}}$ ) in this study. Antenna has a pentagonal patch with a width (w) of 15 mm and a length (l) of 12 mm. This shape is as variation of rectangular shape with bevel techniques. The couple slots, L and U, are designed very carefully by studying the current flow distribution which will give input impedance improvement.

The slots proposed on patch effectively change its electrical length over a very wide bandwidth. Slot dimensions of the proposed antenna are listed in Table 1. The slot width is 0.5 mm in order to improve the bandwidth above 10 GHz.

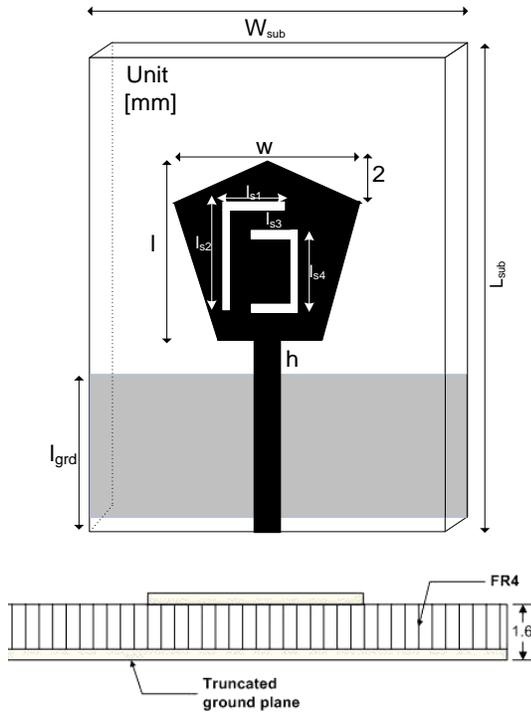


Figure 1. Geometry of L and U slotted antenna

TABLE I DIMENSION OF L AND U SLOT

Description	L and U Slots	
	Symbol	Size [mm]
Slot length	Is1	6
	Is2	9
	Is3	3
	Is4	6.5

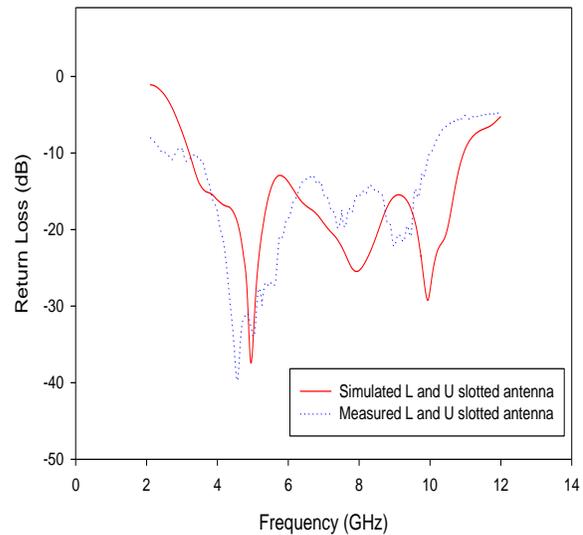


Figure 2. The measured and simulated return loss L and U slotted antenna

The simulated and measured return loss is shown in Fig. 2. The measured return loss is slightly shifted to the simulated one, but they still cover 2.5 GHz to 10.1 GHz as what the UWB required. The length of L slot is 14.5 mm approximately equal to  $0.25\lambda$  at 5.3 GHz, and the length of U slot is 11.5 mm approximately equal to  $0.4\lambda$  at 10.3 GHz.

The original geometry of slotted antenna is taken as a reference to form new modified L and U reconfigurable slotted antennas. The modified antennas are designed for having reconfigurable frequency notched at FWA, HIPERLAN, and WLAN as shown in Fig. 3. There are maximum six switches used to provide the reconfigurable function. No especial matching network is used and the matching properties are solely determined by the placement of the switches. The dimensions of antenna and substrate are kept equal to the original model. The length of L and U slots are similar to the previous length, except two additional slot lengths,  $l_{s20}$  and  $l_{s21}$ . The additional slots are very critically determined by the frequency notched band characteristics.

Licensed band at FWA for point to multipoint radio systems assigned by Malaysian Communications and Multimedia Commissions (MCMC) for 3.4 to 3.7 GHz is considered giving potential interference to UWB application. Therefore, the antenna had notched characteristic at this band is also proposed.

For the simulation purposed, the switches are considered as ideal switches and are modeled as small patches that connect or disconnect the adjacent slot, changing the antennas' slot length.

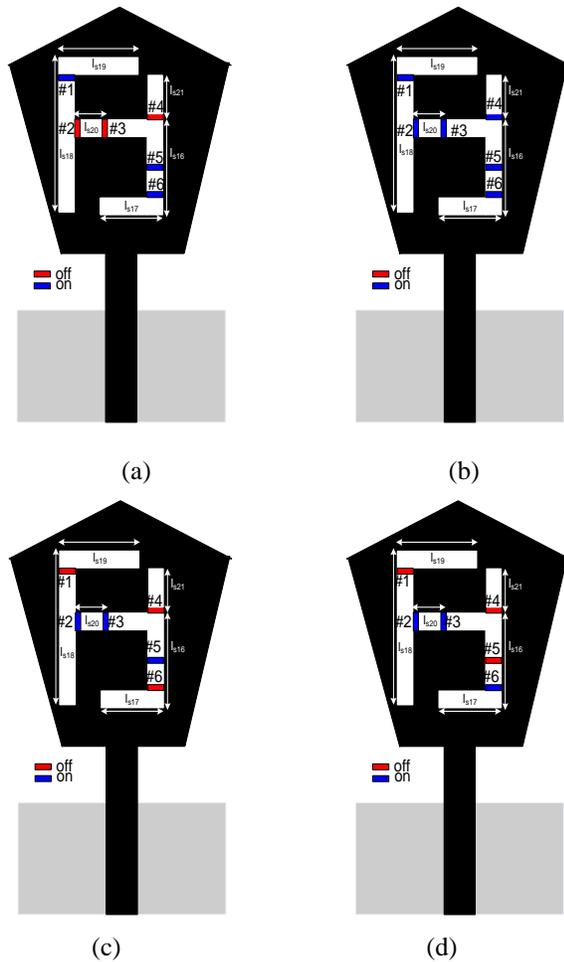


Figure 3. Switching configuration for L and U slotted antennas: (a) without notched, (b) notched at FWA, (c) notched at HIPERLAN, (d) notched at WLAN

For prototype development, the gaps are created in the UWB antenna pattern, which are represented as switches. The selection of PIN diodes as switches is based on their low cost, higher speed and they have better insertion losses at higher frequency than FET switches.

In Fig. 3, the switches have different colors for on and off state condition. Blue color represents the on state condition and red color for the off state condition. In order to provide the UWB characteristic, the switches are placed as shown in Fig. 3(a). Three switches of #2, #3, and #4 are in the off state position. Other switches of #1, #5, and #6 are in the on state condition. When the switches are in the off state condition, the gap between slots occurs and the current flowing to the gap. When the switches are in the on state condition, there is no current flowing to the slots. Thus it forms continuous slots. The switches of #2 and #3 are incorporated to the first additional slot ( $I_{s20}$ ) which is 3.5 mm of slot length. The switch of #4 is attached to the second additional slot ( $I_{s21}$ ) which is 2.5 mm of slot length.

The frequency notched characteristic antenna at FWA is shown in Fig. 3(b). All switches are in the on state position

(continuous slot). Total slot lengths are 32 mm or approximately equal to  $0.4\lambda$  at 3.7 GHz. The total slot lengths mean the sum of slot lengths of L, U and additional slots. Fig. 3(c) and Fig. 3(d) present the frequency notched characteristic antenna at HIPERLAN and WLAN, respectively. To reject interference from HIPERLAN, the switches of #1, #4, and #6 are in the off state position while switches of #2, #3, and #5 are in the on state position. It is investigated that by inserting those switches in the off state condition broke the connection between slots. This break connection has reduced the slot length to be 20.75 mm or approximately equal to  $0.33\lambda$  at 5.2 GHz. Thus, the antenna has frequency notched at HIPERLAN. The total slot length is measured from the length of connecting slots.

The configuration of switches in Fig. 3(d) have resulted an antenna with frequency notched at WLAN. It is shown that the switch of #5 set in the off state position in order to reduce the slot length, while the switch of #6 is set in the on state position. This is the only different while compared to the HIPERLAN configuration. Total slot lengths are 18 mm or approximately equal to  $0.33\lambda$  at 5.75 GHz and measured from the length of connecting slots.

### III. RESULTS AND DISCUSSION

Fig. 4 shows the simulated VSWR for reconfigurable modified L and U slotted antennas. By varying the slot lengths and break the connection between slots using switches, the proposed frequency notched is achieved. The FWA notched band is obtained from 3.57 GHz to 3.86 GHz with the total slot length of 32 mm at 3.7 GHz, which is the centre frequency. While the HIPERLAN and WLAN notched bands are from 4.84 GHz to 5.33 GHz and 5.53 GHz to 6.02 GHz, respectively.

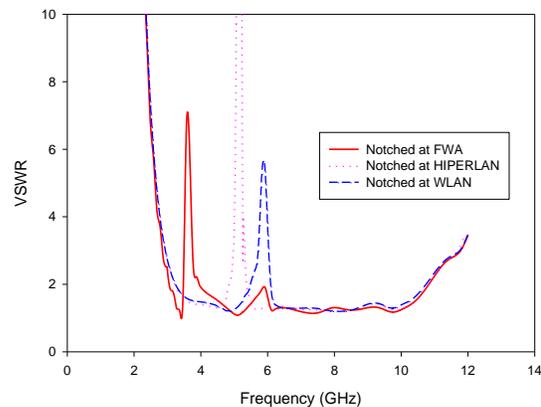


Figure 4 Simulated VSWR for reconfigurable modified L and U slotted antennas

It is noted that beyond the frequency notched bands, the VSWR is kept to be less than 2. With the notched band's characteristic, the antenna has ability to reconfigure its frequency that only responsive to other frequencies beyond the rejection band within UWB bandwidth.

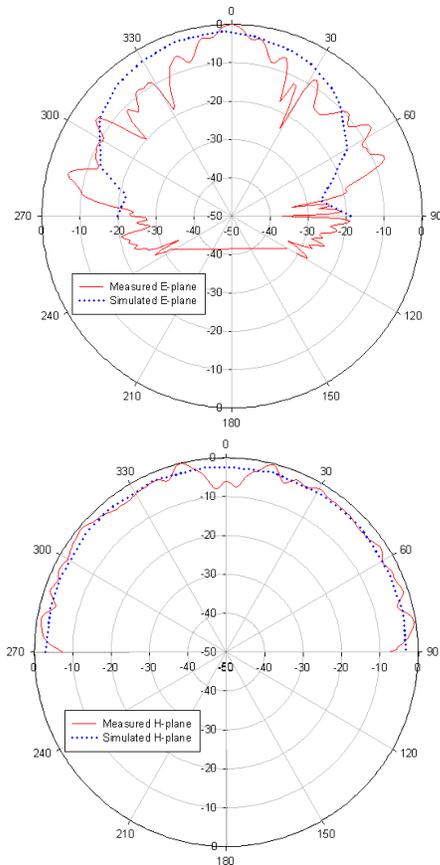


Figure 5. Simulated and measured E and H planes at 4 GHz for antenna notched at FWA

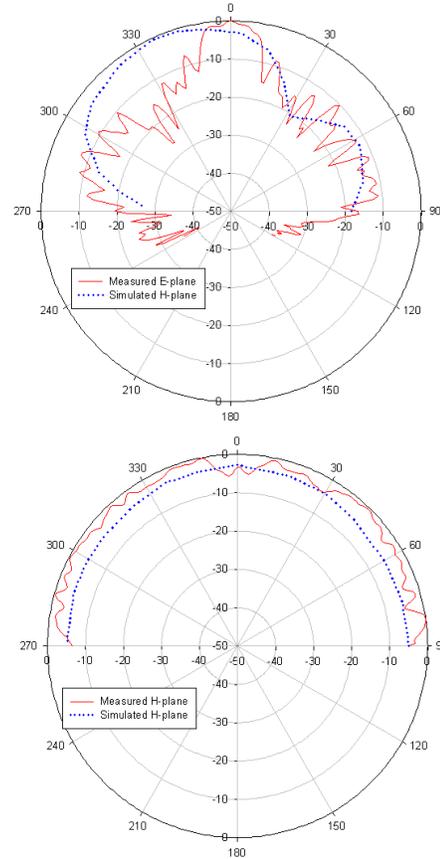


Figure 6. Simulated and measured E and H planes at 5.8 GHz for antenna notched at FWA

Once the resonance frequencies were identified, principal radiation patterns were taken to characterize the operational performance of each antenna. These measurements were conducted in indoor anechoic chamber room. The probes available in the chamber room are in the frequency ranges of 3.95 – 5.85 GHz and 8.95 – 12 GHz, respectively. The existing chamber employed the spherical near field measurement.

Comparison between measured and simulated radiation patterns for these proposed antennas are plotted in Fig. 5 to Fig. 8. All radiation patterns were measured at 4 GHz and 5.8 GHz for both E and H planes. Fig. 5 and Fig. 6 show the radiation patterns of antenna notched at FWA for 4 GHz and 5.8 GHz, respectively.

From Fig. 5 and Fig. 6, there are slightly back-lobes present for the E-planes for both frequencies. Both H-planes are omni-directional with slightly gain decreased at boresight direction. More distortions occur in the measured patterns compared with the simulated ones. This is due to an enhanced perturbing effect on the antenna performance caused by the feeding structure and cable at these frequencies.

During measurement process, several requirements are needed to take into consideration. Obtaining true patterns depends primarily on accurately positioning the probe, accurately measuring the field, and eliminating distortions in the field introduced by the room, tracks, or probe [10]. The room reflections must be lower than the basic side-lobes level and the probe must have low reflections. The probe position must be accurate to give better tolerance corresponding to the side-lobe level. In a spherical near-field range, the spherical measurement surface will be imperfect due to inaccuracies of the positioners and misalignment of these positioners [11].

Fig. 7 shows the radiation patterns for antenna notched at HIPERLAN. The E and H planes measured at 4 GHz and 5.8 GHz.

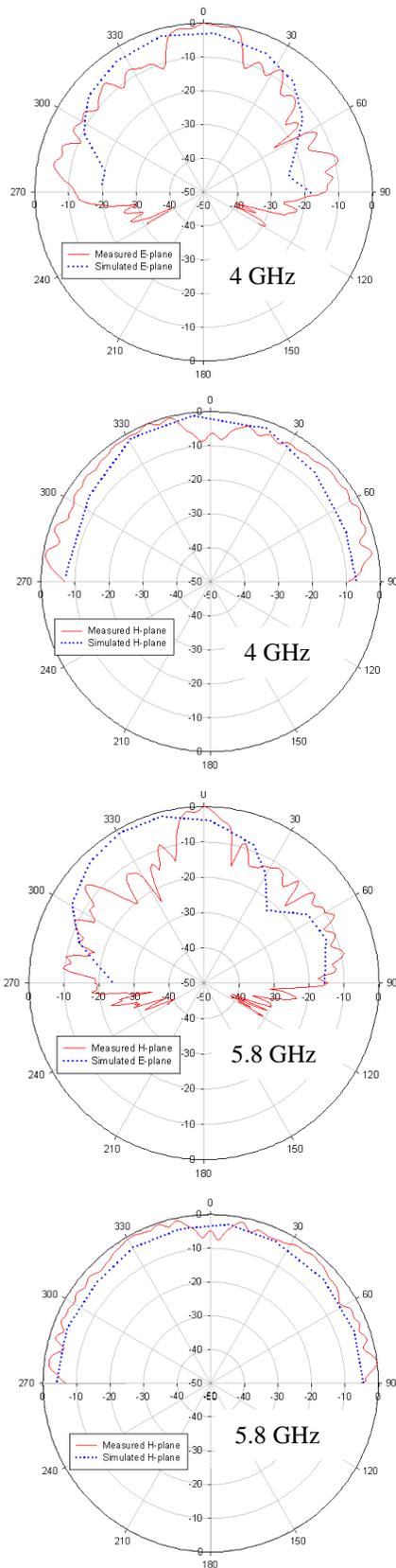


Figure 7. Simulated and measured E and H planes at 4 GHz and 5.8 GHz for antenna notched at HIPERLAN

Fig. 8 and Fig. 9 present the simulated and measured E and H planes for antenna notched at WLAN for both frequencies of 4 GHz and 5.8 GHz, respectively. It is observed that the measured E planes for antenna notched at HIPERLAN broader than the measured E planes for antenna notched at FWA. Both H planes are omni-directional. The overall H-planes patterns retain a satisfactory omni-directionality (less than 10 dB gain variation in most directions) over the entire bandwidth in both simulation and experimental.

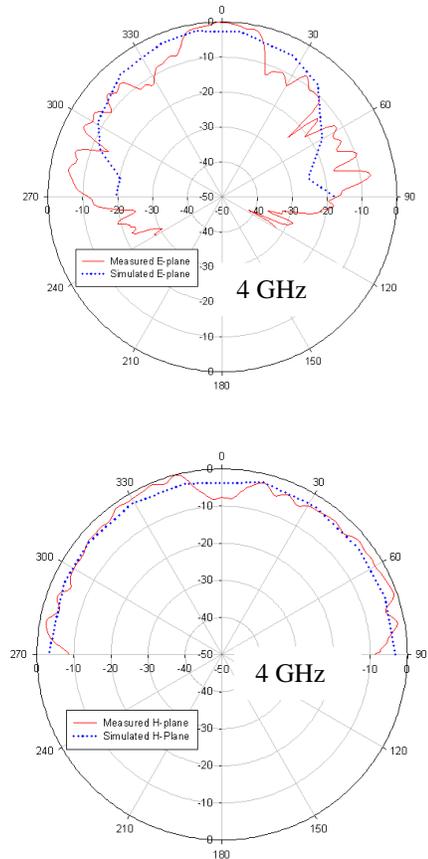


Figure 8. Simulated and measured E and H planes at 4 GHz for antenna notched at WLAN

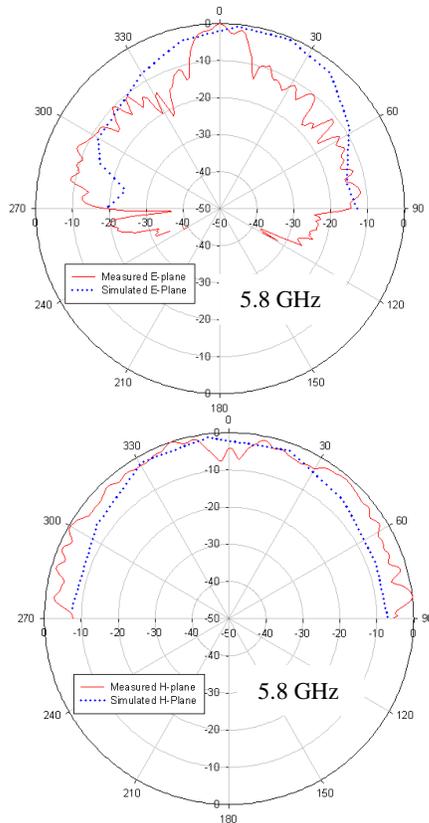


Figure 9. Simulated and measured E and H planes at 5.8 GHz for antenna notched at WLAN

#### IV. SUMMARY

New slotted reconfigurable antennas with band notched characteristics at FWA, HIPERLAN, and WLAN band have successfully designed and developed. The new models are obtained from modification of previous models without degrading their performance. In order to reconfigure their frequency notched band, six switches are attached to the antenna. It is shown that by varying the slot length, the frequency notched antennas are performed at certain frequency.

Simulated and measured radiation patterns for proposed slotted reconfigurable antennas at 4 GHz and 5.8 GHz have been examined. They show acceptable results where the overall H-planes for both frequencies providing omnidirectional patterns. But, more ripples occur in the measured E-planes radiation pattern. This is due to some errors during measurement process.

#### ACKNOWLEDGEMENT

The authors would like to thank Research and Innovation Department of Universiti Malaysia Pahang for supporting this research and Wireless Communication Centre (WCC) of Universiti Teknologi Malaysia for cooperating and facilitating this research measurement.

#### REFERENCES

- [1] H. G. Schant, G. Wolynec, and E. M. Myszka. Frequency Notched UWB Antenna. Proc. IEEE Conference on Ultra Wideband Sys and Tech. 2003. pp. 214-218.
- [2] Y. Kim and D. H. Kwon. CPW-Fed Planar Ultra-Wideband Antenna Having a Frequency Band Notch Function. Electronics Letters. April, 2004. Vol. 40(7): pp. 403-405.
- [3] S. W. Su, K. L. Wong, and F. S. Chang. Compact Printed Ultra-Wideband Slot Antenna with a Band Notched Operation. Microwave & Optical Techni. Letters. April, 2005. Vol. 45(2): pp. 128-130.
- [4] J. Kim, C. S. Cho, and J. W. Lee. 5.2 GHz Notched Ultra-Wideband Antenna Using Slot-type SRR. Electronics Letters. March 16, 2007. Vol. 42(6).
- [5] W. S. Lee, W. G. Lim, and J. W. Yu. Multiple Band-Notched Planar Monopole Antenna for Multiband Wireless Systems. IEEE Microwave and Wireless Components Letters. Sept., 2005. Vol. 15(9).
- [6] S. Hong, K. Chung, and J. Choi. Design of a Band-Notched Wideband Antenna for UWB Applications. Antennas and Propagation Society International Symposium. IEEE. 2006.
- [7] Kenneth C. L. Chan and Yi Huang. A Compact Semi-Circular Disk Dipole with Notched Band for UWB Applications. The Institution of Engineering and Technology Seminar on Ultra Wideband Systems, Technologies, and Applications. 2006. pp. 226-230.
- [8] K. Chung, J. Kim, and J. Choi. Wideband Microstrip-Fed Monopole Antenna Having Frequency Band Notch Function. IEEE Microwave and Wireless Components Letters. November, 2005. Vol. 15(11).
- [9] Y. Kim and D. H. Kwon. Planar Ultra Wideband Slot Antenna with Frequency Band Notch Function. Antennas and Propagation Society International Symposium. IEEE. 2004.
- [10] Gary E. Evans. Antenna Measurement Techniques. Boston, London: Artech House. 1990.
- [11] G. Hindman and Allen C. Newell. Simplified Spherical Near-Field Accuracy Assessment. Technical Paper. Nearfield System, Inc. 2004.

# Radio Resource Unit Allocation and Rate Adaptation in Filter Banks Multicarrier System

Faouzi Bader

Centre Tecnològic de Telecomunicacions de Catalunya-  
CTTC.

Av. Carl Friedrich Gauss 7, 08860 Barcelona-Spain.  
Phone: +34 936452901, fax: +34 936452901, e-mail:

[faouzi.bader@cttc.es](mailto:faouzi.bader@cttc.es)

Mohammad Banat

Department of Electrical Engineering, Jordan  
University of Science and Technology (JUST)  
P.O.Box 3030. Irbid 22110-Jordan. Email:

[m.banat@ieee.org](mailto:m.banat@ieee.org)

**Abstract**— In this paper, the effect of the effective channel gain and the SNR using the improved joint resource unit (RU) allocation and Bit loading (JRAB) on a filter banks multicarrier (FBMC) system are analyzed. Computer simulations have been performed assuming a hypothetical WiMAX scenario in which an FBMC system substitutes OFDM by maintaining as much as possible the physical layer compatibility. From obtained simulation results it has been demonstrated that it is possible to upper-bound the maximum delay for delay-sensitive applications (rtPS and nrtPS) using the above mentioned effective metrics in a FBMC system.

**Keywords**-RRM; Filter bank; WiMAX; JRAB.

## I. INTRODUCTION

Adaptive modulation is considered as one of the main techniques to increase the data rate that can be reliably transmitted over a fading channel. Many forms of adaptive modulation and transmission techniques have been proposed and implemented in recent wireless systems [1] [4] [5], however, adaptive bit loading is one of the key features of very recent wireless communication systems (i.e., WiMAX, etc), and its importance will increase in the near future (i.e., LTE). Adaptive modulation technique is possible thanks to adjustability of many system parameters according to the channel fading state variability, the transmit power, the data rate, and channel coding rate.

Two adaptive approaches are widely considered in the scientific literature. The first one is the rate adaptation (RA) which has been treated in [1] and [2]. The RA approach is based on the bite error rate (BER) (or the packet error rate - PER) which is bounded while the maximum throughput is attained by allocating different transmit powers into different users. The second approach is the margin adaptation (MA). The MA method is based on the use of minimum transmit power while the minimum required quality of service (QoS) is guaranteed. It is also possible to generalize the (*multiuser*) link adaptation process using either the RA or the MA objective functions. The outcomes of these adaptations are the resources assigned to each user in the time, frequency or space domains, the transmitted power per user, and the optimum modulation and coding scheme (MCS). Using an orthogonal frequency division

multiplexing (OFDM) scheme [2] resources can be efficiently assigned to different users without the need to use guard bands or time gaps (when perfect synchronization between the mobile station (MS) and the base station (BS) is assumed). The whole frequency and time domains are segmented into different resource units (RU) which can be arbitrarily (or using specific policies) assigned to different users. The minimum RU is a single symbol, the resource allocation algorithm inputs are the channel state information (CSI) of all users, and the maximum allowed transmission power. The main considered output is the power and MCS assigned to each RU.

To the best knowledge of the authors, this is the first time the radio resource unit allocation is treated using both the effective channel gain and the effective signal to noise ratio ( $SNR_{eff}$ ) concept in a filter banks multicarrier (FBMC) system. The FBMC communication scheme has been subject of intense researches during last year's mainly within the Information and Communications Technology (ICT) European project PHYDYAS (Physical layer for dynamic spectrum access and cognitive radio) [9].

The remainder of this paper is organized as the following: in Section II, the model and main features of the used FBMC system is summarized. In Section III, the expression of the effective channel transfer functions of both OFDM and FBMC systems is derived, and the resource unit (RU) capacity is calculated in Section IV. The main results obtained by simulation are presented in Section V, and finally, main conclusions are outlined in Section VI.

## II. INTRODUCTION TO FBMC SYSTEM

Filter banks multicarrier system can be realized via a digital transmultiplexer configuration, where a synthesis filter bank (SFB) is used at the transmitter side while an analysis filter banks (AFB) is used at the receiver side [3] [9]. In FBMC applications, the use of a critical sampled filter banks would be problematic, since the aliasing effect would make it difficult to compensate the channel imperfections' effects by processing the sub-channel signals only after the AFB. Therefore, a factor of two oversampling is commonly applied into the sub-channel signals at the AFB [3].

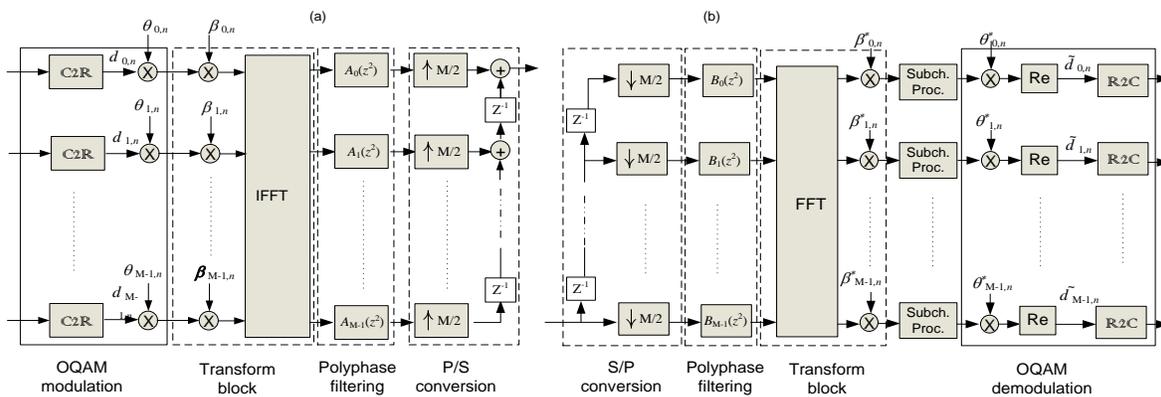


Figure 1: Multicarrier polyphase filter banks for SISO case, a) Synthesis filter banks (SFB), b) Analysis filter banks (AFB)

In this paper, a uniform modulated filter banks is assumed where the prototype filter  $p[m]$  of length  $L$  is shifted to cover the whole system bandwidth. The output signal from the synthesis filter bank is defined as in [3] by,

$$s[m] = \sum_{k=0}^{M-1} \sum_{n=-\infty}^{\infty} d_{k,n} \theta_{k,n} \beta_{k,n} p \left[ m - \frac{nM}{2} \right] e^{j \frac{2\pi}{M} km} \quad (1)$$

where

$$\beta_{k,n} = (-1)^{kn} e^{-j \frac{2\pi k L - 1}{M} \frac{nM}{2}} \quad (2)$$

where  $M$  is the total number of subcarriers (IFFT/FFT size),  $d_{k,n}$  is the real-valued symbol (of rate  $2/T$ ) modulated over the  $k$ -th subcarrier and the  $n$ -th time symbol interval. The time signaling interval  $T$  is defined as the inverse of the subcarrier spacing  $\Delta f$ . The symbols  $d_{k,n}$  and  $d_{k,n+1}$  can be seen as the in-phase and quadrature (I/Q) components of a complex-valued symbol  $c_{k,l}$  (of rate  $1/T$ ) chosen from a multilevel quadrature amplitude modulation (M-QAM) alphabet. Note that the sign of the sequence  $\{\theta_{k,n} = j^{k+n}\}$  sequence can be chosen arbitrarily, but the pattern of the real and the imaginary samples have to follow definition (1) and (2) to maintain (near) orthogonality [3] [9].  $L$  is the length of the prototype filter  $p[m]$  and is equal to the product of the filter bank size  $M$  and the overlapping factor  $K$  ( $L=KM$ ) [3]. The ‘‘C2R’’ and ‘‘R2C’’ blocks depicted in Figure 1, indicate the conversion from complex to real form, and the inverse operation respectively.

As it can be observed in (1), the synthesized signal is a composite of  $M$  sub-channel signals each one is a linear combination of time-shifted (by multiples of  $T/2$ ) and the overlapped impulse response of the prototype filter weighted by the respective data symbol  $d_{k,n}$ . When a real (imaginary) part of a subcarrier symbol is used (to carry an information

symbol) the unused imaginary (real) part is at the receiver a fairly complicated function of surrounding data symbols effect.

### III. EFFECTIVE OFDM AND FBMC CHANNEL TRANSFER FUNCTIONS

In order to obtain the optimal power and the bit rate adaptation we need the bit error rate ( $BER$ ) expression in additive white Gaussian noise (AWGN) which is easily invertible in terms of bit rate and power. Unfortunately, for most of non-binary modulation techniques, e.g., multi-level QAM (MQAM), and multi-level phase shift keying (M-PSK), an exact expression for the  $BER$  is hard to obtain. Often, the  $BER$  with Gray bit mapping at high SNRs can be approximated as the symbol error rate ( $SER$ ) divided by number of bits per symbol [4]. The equivalent subcarrier approach developed by C. Tang *et al.*, in [5] allows a group of subcarriers containing spread data symbols to be represented by a single equivalent subcarrier to handle the bit and power loading mechanism in a more compact and simpler way.

To better understand this concept let first consider an OFDM system where the M-QAM bit error estimation approximation developed in [4] is used to obtain the  $BER$  of the  $k$ -th equivalent subcarrier,

$$BER_k \approx 0.2 \exp \left( \frac{-1.6 |H_{eff,k}|^2 P_k}{(2^{b_k} - 1) \sigma^2} \right) \quad (3)$$

where  $BER_k$  is the approximate  $BER$ ,  $|H_{eff,k}|^2$  means the square magnitude of the effective channel transfer function,  $P_k$  is the transmit power at the  $k$ -th subcarrier,  $b_k$  is the number of transmitted bits, and  $\sigma^2$  is the additive white Gaussian noise (AWGN) power. Using this expression as equality instead of an approximation, the expressions for the assigned power and number of bits may be solved as in [4] by,

$$P_k = \frac{-(2^{b_k} - 1)\sigma^2}{1.6|H_{eff,k}|^2} \ln\left(\frac{BER_k}{0.2}\right) \text{ for } BER_k < 0.2 \quad (4)$$

$$\max b_k = \max \left\{ \log_2 \left( 1 - \frac{1.6|H_{eff,k}|^2 P_k}{\sigma^2 \ln\left(\frac{BER_k}{0.2}\right)} \right) \right\} \quad (5)$$

Subject to:  $BER_k < \overline{BER}_k$

Subject to:  $\sum_{k=1}^M P_k - P_T \leq 0$

where  $\overline{BER}_k$  is the upper bound BER, and  $P_T$  is the maximum allowed total transmit power. Knowing the value  $|H_{eff,k}|^2$  of each active user, the number of bits that can be loaded without exceeding a certain BER threshold is estimated using (4), and it's equal to,

$$b_k = \log_2 \left( 1 - \frac{1.6|H_{eff,k}|^2 P_k}{\sigma^2 \ln\left(\frac{BER_k}{0.2}\right)} \right) \quad (6)$$

The first step of the adaptation process is the calculation of the effective channel gain over the  $M$  subcarriers and transmitted frames (such parameters will be fixed in the simulation section, e.g., by using WiMAX standard specifications [8]). For purpose of an easier comprehension we started calculating the effective channel gain of an OFDM system, and after the FBMC system.

#### A. Calculation of Effective Channel Transfer Function: OFDM Case

We assume the single user case, where the modulated data symbol of the active user at the  $k$ -th carrier is  $S_k$ . By removing the cyclic prefix at the demodulator the  $k$ -th received data symbol in frequency domain is,

$$Y_k = S_k H_k + \eta_k \quad (7)$$

where  $H_k$  means the channel gain, and  $\eta_k$  is the AWGN component at the  $k$ -th subcarrier index. After using a zero-forcing (ZF) equalization, the enhanced noise term at the  $k$ -th subcarrier is given by the term  $\eta_k/H_k$ . We assume that the noise power is equally distributed over all the subcarriers with a value  $\sigma^2$ . Therefore, the power of the enhanced noise term at  $k$ -th subcarrier is  $\sigma^2/|H_k|^2$ . If we define  $P_s = E[S_k S_k^*]$  as the total power transmitted by a single subcarrier, the instantaneous signal to noise ratio will be equal to:

$$SNR_k = \frac{P_s}{\sigma^2 / |H_k|^2} \quad (8)$$

The effective channel power attenuation for the modulated symbol is,

$$|H_{eff,k}|^2 = |H_k|^2 \quad (9)$$

Substituting (9) into (8) yields to the following

$$SNR_k = \frac{P_s}{\sigma^2} |H_{eff,k}|^2 \quad (10)$$

Knowing the CSI values, the base station (BS) can use the calculated power in (9) to define the bit and power loading values for transmission. Then the packet scheduler is in charge of formatting the symbols to fit into one or several RUs. The MCS of the burst frame is fixed based on the effective SNR ( $SNR_{eff}$ ) of the sub-channel and the symbol where the burst is allocated. The effective SNR function ( $SNR_{eff}$ ) is a function of different instantaneous SNRs, and is defined as,

$$SNR_{eff} = \phi^{-1} \left\{ \frac{1}{M} \sum_{k=1}^M \phi(SNR_k) \right\} \quad (11)$$

#### B. Calculation of Effective Channel Transfer Function: FBMC Case

In contrast to the OFDM scheme, where complex valued symbols are transmitted at a given symbol rate the FBMC transmits real symbols at twice the OFDM's rate. Therefore, FBMC is a scheme that preserves the spectral efficiency and even allows the optimization of the carrier pulse shape according to the channel characteristics [3] [9].

Hereafter we assume the use of the Zero-Forcing equalization. The frequency-time pair  $(k, n)$  denotes the subcarrier  $k$  and symbol time  $n$  (with  $T/2$  spacing time) position respectively within a transmitted frame (see Figure 2). In FBMC, every frequency-time position suffers interference from neighboring sub-channels (Figure 2). For an ideal channel this interference affects only the sub-channels of the imaginary symbols, while the real part of the symbols yields the originally transmitted symbol  $S_{k,n}$ . Note that this interference could be considered as a (sometimes close to zero) random variable that depends on the transmitted symbols around the symbol position  $(k, n)$ . From (1) and (2), and with some mathematical arrangements we obtain in (12) the received FBMC signal expression

$$y_{k,n} = H_{k,n} S_{k,n} + j \sum_{(p,q) \neq (k,n)} H_{p,q} w_{p,q} S_{p,q} + \eta_{k,n} \quad (12)$$

where the  $\{w_{p,q}\}$  values are the constant filter banks coefficients depicted in Table I, their effect constitute the main interference component in (12) (right summation value) on every transmitted data symbol.

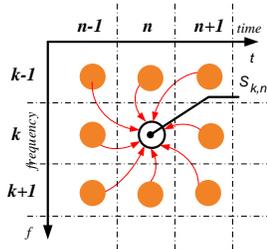


Figure 2: First time order neighbors in time-frequency representation for PHYDYAS FBMC system [9].

TABLE I. TRANSMULTIPLEXER RESPONSE OF THE FBMC SYSTEM USED IN PHYDYAS PROJECT [9]. ROWS REPRESENT TIME DIRECTION AND THE COLUMNS THE FREQUENCY DIRECTION

0.0006	0.0001	0	0	0	0.0001	0006
-0.0429j	0.1250	<b>0.2058j</b>	<b>0.2393</b>	<b>0.2058j</b>	0.1250	0.0429j
-0.0668	0.0002	<b>0.5644</b>	<b>1.000</b>	<b>0.5644</b>	0.0002	0.0668
0.0429j	0.1250	<b>0.2058j</b>	<b>0.2393</b>	<b>0.2058j</b>	0.1250	0.0429j
0.0006	0.0001	0	0	0	0.0001	0.0006

The neighborhood set of positions that affects a given frequency-time  $(k, n)$  position of a transmitted symbol is defined as,

$$\Omega_{\Delta k, \Delta n} = \{(p, q), |p| \leq \Delta k, |q| \leq \Delta n, H_{k+p, n+q} \approx H_{k, n}\} \quad (13)$$

We define the set  $\Omega_{\Delta k, \Delta n}^*$  such that  $\Omega_{\Delta k, \Delta n}^* = \Omega_{\Delta k, \Delta n} - (k, n)$ . Note that both  $\Delta k$  and  $\Delta n$  should be chosen taking into account the channel time coherence  $T_c$ , and the bandwidth  $B_c$ . It is worth mentioning that when  $B_c$  decreases the value of  $\Delta k$  also decreases. The same can be concluded for  $T_c$  and  $\Delta n$ . Having a well-dimensioned real system,  $B_c$  encompass few subcarriers ( $\Delta k \geq 1$ ) while  $T_c$  is generally larger than  $T$  ( $\Delta n \geq 1$ ). This allows us to rewrite the received signal in (12) using zero-forcing (ZF) equalization as,

$$\begin{aligned} \frac{y_{k,n}}{H_{k,n}} &= S_{k,n} \\ &+ j \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} \frac{H_{k+p, n+q}}{H_{k,n}} w_{k+p, n+q} S_{k+p, n+q} \\ &+ j \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} \frac{H_{k+p, n+q}}{H_{k,n}} w_{k+p, n+q} S_{k+p, n+q} + \frac{\eta_{k,n}}{H_{k,n}} \end{aligned} \quad (14)$$

If we consider that the generated filter banks prototype is well-localized in time and frequency domain, we consequently have

$$\begin{aligned} &\left| \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} \frac{H_{k+p, n+q}}{H_{k,n}} w_{k+p, n+q} S_{k+p, n+q} \right| \\ &\square \left| \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} \frac{H_{k+p, n+q}}{H_{k,n}} w_{k+p, n+q} S_{k+p, n+q} \right| \end{aligned} \quad (15)$$

Therefore, (14) can be rewritten as (see weights' value at columns 2 and 6 in Table 2)

$$\frac{y_{k,n}}{H_{k,n}} \approx S_{k,n} + j \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} \frac{H_{k+p, n+q}}{H_{k,n}} w_{k+p, n+q} S_{k+p, n+q} + \frac{\eta_{k,n}}{H_{k,n}} \quad (16)$$

using  $(p, q) \in \Omega_{\Delta k, \Delta n}^*$  as the summation range in (16) means that we can approximate the channel gain at  $(k+p, n+q)$  position by that experienced at  $(k, n)$ . Therefore, using the ZF equalizer the received symbol at the  $k$ -th sub-carrier and  $n$ -th time is equal to,

$$\hat{S}_{k,n} = S_{k,n} + j \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} w_{k+p, n+q} S_{k+p, n+q} + \frac{\eta_{k,n}}{H_{k,n}} \quad (17)$$

$$\hat{S}_{k,n} = S_{k,n} + j I_{k,n} \quad (18)$$

From (15), the enhanced noise term after equalization is

$$I_{k,n} = \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} w_{k+p, n+q} S_{k+p, n+q} - j \frac{\eta_{k,n}}{H_{k,n}} \quad (19)$$

where the power term here is,

$$E[I_{k,n} I_{k,n}^*] \approx \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} w_{k+p, n+q}^2 P_{k+p, n+q} + \frac{\sigma^2}{|H_{k,n}|^2} \quad (20)$$

If  $P_{k,n} = |S_{k,n}|^2$  is the total transmitted power by a single carrier, then, the instantaneous signal to noise ratio is

$$SNR_{k,n} = \frac{P_{k,n} |H_{k,n}|^2}{\sigma^2 + \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} |H_{k,n}|^2 w_{k+p, n+q}^2 P_{k+p, n+q}} \quad (21)$$

where  $P_{k+p, n+q}$  is the total transmitted power by each symbol belongs to the set  $\Omega_{\Delta k, \Delta n}^*$ . From (20), it can be seen that the effective power for the modulated symbol is:

$$|H_{eff, k, n}|_{FBMC}^2 = \frac{|H_{k,n}|^2}{1 + \frac{\sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} w_{k+p, n+q}^2 P_{k+p, n+q}}{\sigma^2}} \quad (22)$$

Then the  $SNR_{eff}$  is calculated and is equal to,

$$SNR_{eff, k, n} = \frac{P_{k,n}}{\sigma^2} |H_{eff, k, n}|_{FBMC}^2 \quad (23)$$

Note that  $|H_{eff, k, n}|_{FBMC}^2$  has a random behavior as it's strongly dependent of the interference part

$$\left\{ \sum_{(p,q) \in \Omega_{\Delta k, \Delta n}^*} w_{k+p, n+q}^2 P_{k+p, n+q} \right\}.$$

#### IV. FBMC RESOURCE UNIT ALLOCATION USING THE $SNR_{eff}$

In time division duplex (TDD) approach the communication frame consists on  $N_s$  symbols of duration  $T_{frame}$  seconds. The numbers of downlink and uplink OFDM/FBMC symbols usually follow the ratio 2:1 or 3:1. However, this parameter can be adjusted at the BS according to user demands and the available resources. The total system bandwidth  $BW$  consists of  $N_c$  subcarriers where only a limited number equal to  $N_{used}$  are active, while the remaining carriers are used as guard tones. Active subcarriers include both pilot subcarriers and data subcarriers, which will be mapped over different sub-channels according to specific subcarrier permutation schemes.

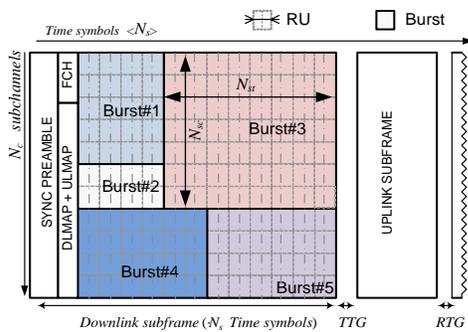


Figure 3: FBMC frame in TDD mode like burst structures based on IEEE 802.16e standard [8].

For the full usage of subcarriers (FUSC) scheme, pilot subcarriers are allocated first, and the remainders are grouped into sub-channels, where the data subcarriers are mapped. On the other hand, in partial usage subcarriers scheme (PUSC), and in adjacent subcarrier permutation schemes (usually referred as band AMC) map first all the pilots and the data subcarriers into the sub-channels, and therefore, each sub-channel contains its own set of pilot subcarriers.

For the FUSC and PUSC modes, the assigned subcarriers to each sub-channel are distant in frequency, whereas for AMC scheme the subcarriers belonging to one sub-channel are adjacent. Note that both FUSC and PUSC increase the frequency diversity and average the interference effect, whereas the AMC is more convenient for bit loading and beamforming as an increase in multiuser diversity is demanded. As depicted in Figure 3, and similar to WiMAX standard [8], the minimum RU assigned to any data stream within a frame has a two dimensional shape constructed by at least one sub-channel and two symbols<sup>1</sup>.

<sup>1</sup> one OFDM symbol in case of OFDM scheme

We define a RU as a resource unit formed by a set of  $N_{sc} \times N_{st}$  subcarriers and FBMC symbols, respectively. Once the size of the RU is defined it's possible to obtain the total number of RUs per frame  $Q \times T$ , where  $Q = N_c / N_{sc}$  is the number of sub-channels and  $T = N_s / N_{st}$  is the number of time slots. Note that both the RU and the data region always follow a rectangular shape structure. In the IEEE 802.16 standard, the specific size of the RU varies according to the permutation scheme, concretely for the AMC scheme the RU may take the sizes;  $9 \times 6$ ,  $18 \times 3$  or  $27 \times 2$ , where one ninth of the subcarriers are dedicated to pilots [8]. By analogy, and taking into account that the OFDM symbol duration is twice that of the FBMC symbol (Figure 2) for the AMC scheme, the FBMC RU may takes the sizes  $9 \times 12$ ,  $18 \times 6$  or  $27 \times 4$ .

The effective  $SNR_{eff}$  for FBMC is given by (23). The effective SNR merges the SNR from the different subcarriers, i.e., in a sub-channel or in a burst. Therefore, the BER, the packet error rate or the channel capacity can be obtained directly by assuming an AWGN channel from an equivalent SNR equal to the  $SNR_{eff}$ . The power is assumed uniformly distributed over all the subcarriers. Therefore, the effective channel gain can be obtained via the geometric mean of the subcarriers gain (considering the FBMC structure in the capacity calculation described in [4]) the total capacity  $C_{RU}$  within each RU is calculated by

$$C_{RU} = \sum_{i=1}^{N_{st}/2} \Delta f N_{sc} \left[ \log_2 \left( 1 - \frac{1.6 H_{eff}^2 SNR}{\ln \left( \frac{BER}{0.2} \right)} \right) \right] \quad (24)$$

where  $H_{eff}^2$  is the geometric mean of the effective channel experienced over each  $i = \{1, 2, \dots, N_{st}/2\}$ , and the  $SNR_{eff}$  is the geometric mean of the SNR values over all the  $N_{sc}$  sub-channels of the RU unit.

For the resource allocation scheduling we used the improved Joint RU Allocation and Bit Loading (JRAB) by Scheduling described in [11], where two competing aspects exist during the RU allocation and the scheduling process. These are

- the guarantee of the different service QoS constraints, and
- the maximization of the spectral efficiency.

The packet scheduling functions described by Shakkottai *et al.* in [10] maximize both the spectral efficiency (therefore the bit loading) and the delay effect based on the CSI values. Furthermore, according to described functions in [10] the prevalence of the channel over the distribution of the RUs (to maintain the QoS) or the opposite is difficult to assure. From a system administrator perspective, this approach

might be difficult to implement, and furthermore, it has been shown that when the physical layer is in charge of the bit allocation process of each active user the spectral efficiency increase [6]. For these reasons, the scheme proposed in [10] has tackled the problem from a different perspective. It is usually unavoidable that the packets might be fragmented to fit into the physical layer burst.

We assume in this paper that any packet can be arbitrarily fragmented as many times as necessary (obviously this will affect the spectral efficiency due to the fragmentation of headers). Based on this assumption, and assuming that each packet is delivered within a certain time interval (no matter which class of service it belongs to, with either constant bit rate (CBR) or variable bit rate (VBR) [10]), by following the development in [12], we can obtain the minimum number of bits  $b_k^{(u)}$  that the system should assign to each active FBMC user during each frame by

$$b_k^{(u)} = \begin{cases} T_{\text{frame}} \sum_{p=1}^P \frac{L_p^{(u)}}{\tau_{\text{max}}^{(u)} - \Delta\tau - \tau_p^{(u)}}, & \text{if } \forall p' \rightarrow \tau_{p'}^{(u)} < (\tau_{\text{max}}^{(u)} - \Delta\tau) \\ T_{\text{frame}} \left( \sum_{p=1}^P \frac{L_p^{(u)}}{\tau_{\text{max}}^{(u)} - \Delta\tau - \tau_p^{(u)}} + \sum_{p'} L_{i,p'} \right), & \text{otherwise} \end{cases} \quad (25)$$

where  $T_{\text{frame}}$  is the frame time period (in seconds), and  $L_p^{(u)}$  is the number of bits still queued from the  $p$ -th packet. Then, if  $b_k^{(u)}$  bits are allocated during each frame to each active user  $u$  and the  $k$ -th RU, the delay is certainly under its upper bound. Note that if any  $p'$  packet has waited period more than  $(\tau_{\text{max}}^{(u)} - \Delta\tau)$  all the remained bits of the packet will be considered for transmission in the following frame. The RU allocation and the bit loading problem can be solved for the a minimum rate  $R^{(u)}$  based on the allocated bits  $b_k^{(u)}$  as defined in (25).  $b_k^{(u)}$  is also used to determine the priority assigned to each user within each RU. Hence, for the  $k$ -th sub-channel of the  $u$ -th user the hereafter priority assignment is defined as

$$\varphi_k^{(u)} = \begin{cases} \min \left( \frac{b_k^{(u)}}{b_{\text{max}}}, 1 \right) \frac{\bar{\omega}_k^{(u)}}{\bar{\omega}_{\text{max}}}, & \text{if } \forall p' \rightarrow \tau_{p'}^{(u)} < (\tau_{\text{max}}^{(u)} - \Delta\tau) \\ P_{\text{urgency}} \frac{\bar{\omega}_k^{(u)}}{\bar{\omega}_{\text{max}}}, & \text{otherwise} \end{cases} \quad (26)$$

where  $b_{\text{max}}$  is a normalization factor and is equal to the maximum number of bits that can be transmitted within a frame using the highest MCS scheme. Furthermore, when a packet from the  $u$ -th user/service flow is close to exceed its maximum delay, the term  $b_k^{(u)}/b_{\text{max}}$  in (26) is substituted by an urgency factor  $P_{\text{urgency}}$ , which is a fixed constant

satisfying  $P_{\text{urgency}} > (\bar{\omega}_{\text{min}}/\bar{\omega}_{\text{max}})^{-1}$  inequality. As a result, the packets close to their maximum delay are put ahead in the allocation process in order to avoid the packet drops due to the excessive packet delay. The  $\bar{\omega}_k^{(u)}$  is the achievable throughput or the rate from the  $u$ -th user on the  $k$ -th RU, which is obtained based on the  $SNR_{\text{eff}}$  and the available MCSs.

## V. SIMULATION RESULTS

Table II summarizes the simulation parameters used to verify the results of the RU allocation using the JRAB for FBMC system.

TABLE II. FBMC AIR INTERFACE AND SYSTEM LEVEL CONFIGURATION PARAMETERS

Parameter	Values/Quantities
Carrier Frequency, Bandwidth	3.5GHz, 20MHz
Sampling Frequency	22.857Msps
Subcarrier Permutation	Band AMC
FFT Length	2048
# of Used Subcarriers	1728
# of Subcarriers per Sub-Channel	18
# of FBMC Time Symbols per RU	6
# of Data Symbols per RU	48
Modulation	{4,16,64}-QAM
Channel Coding	Punctured Convolutional with coding rates 1/2, 2/3
Channel Model	Pedestrian B
MS Velocity	10Km/h
Channel Estimation and CQI	Ideal
Shadowing Standard deviation	5dB
BS Transmit Power	49dBm

Figure 4 shows the cumulative density function of the packet delay for 50 and 100 active users. Let first focus on the case when the number of users is  $K=50$ . Figure 4 demonstrates that all the schemes achieve a delay lower than the maximum (50ms), in fact the 99<sup>th</sup> percentile is achieved at 25ms using the JRAB procedure and for the PFS (Proportional Fair Scheduler) [12]. Furthermore, the packet loss rate performance for each scheme is almost zero for the JRAB, and about  $1.6 \times 10^{-5}$  for the PFS.

For the case  $K=100$ , we can observe that the PFS is the algorithm that achieves the lower packet delays whereas the JRAB sends the packets mainly when the urgency factor is applied. During the simulations, the guard time  $\Delta\tau$  is fixed equal to  $0.2 \tau_{\text{max}}^{(u)}$ , and thus the urgency factor is activated when  $\tau_p^{(u)} > (\tau_{\text{max}}^{(u)} - \Delta\tau) = 0.4 \text{ ms}$ . Now, for  $K=100$  the packet loss rate for each scheduling function and bit loading procedure is 0.0824 and 0.1375 for the JRAB and the PFS respectively. Therefore, although most of the packets are sent when they are near to expire with the JRAB, a lower packet loss rate is achieved.

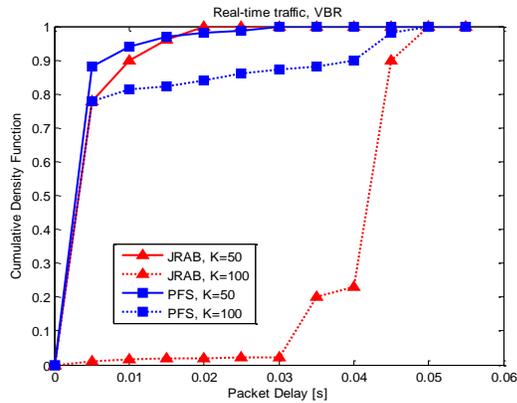


Figure 4: Cumulative density functions of the packet delay for the PFS, and JRAB algorithms with  $K=50$  and  $100$  active users

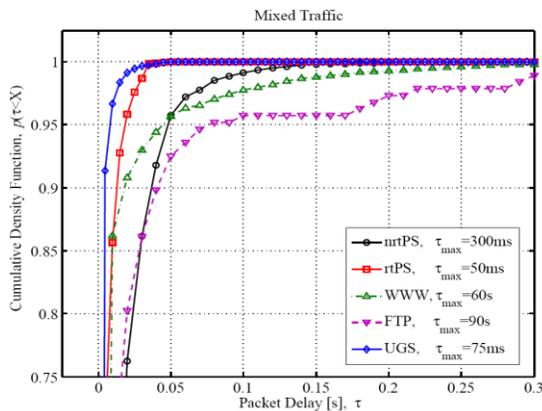


Figure 5: Cumulative density functions of the packet delay for the JRAB algorithm with mixed traffic and  $K=50$  users.

The performance of the JRAB in case of mixed traffic is shown in Figure 5. In this scenario  $K=50$  users are simulated, where ten users require non-real time test service (*nrtPS*), 13 users require real-time test service (*rtPS*), ten users are browsing internet files (World Wide Web service), ten users are downloading files using the File Transfer Protocol (FTP), and 12 users require UGS connections for applications such as Voice over IP. The delay for the www and the FTP services has been assumed as  $\tau_{max}=60s$  and  $\tau_{max}=90s$  respectively.

It is clearly shown in Figure 5 that each traffic type achieves a maximum packet delay lower than its defined maximum value. The 99th percentile for the delay sensitive applications is found to be 100ms, 35ms and 20ms for the *nrtPS*, the *rtPS* and the UGS, respectively, much lower than the fixed maximum values.

## VI. CONCLUSION AND FURTHER WORK

The analysis of using the effective channel information and the  $SNR_{eff}$  metrics in each RU using the JRAB algorithm has shown that it is possible to upper-bound the maximum delay for delay sensitive applications (*rtPS* and *nrtPS*) in a FBMC system. This was achieved despite the interference effect experienced at each subcarrier (or sub-channel)  $(k,n)$  position due to the proper characteristics of filter banks prototype. Besides the higher achieved efficiency by using FBMC compared to OFDM ([3]

[7]), it is even possible to obtain an extra spectral efficiency margin by exploiting the multiuser diversity on those unallocated resources. Future works will be focused on evaluating the effect of having partial CSI information on the assignment of the RU in the FBMC system.

## REFERENCES

- [1] A. Czylik, "Adaptive OFDM for Wideband Radio Channels" IEEE Global Communications Conference (GLOBECOM'96), Vol 1, pp.713-718, Nov 18-22, 1996. London, UK.
- [2] S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Adaptive Resource allocation in Composite Fading Environments," IEEE Global Communications Conference (GLOBECOM 2001), Vol2, pp-1312-1316. San Antonio, Texas. USA. Nov 25-29, 2001.
- [3] M. G. Bellanger, "Transmit diversity in multicarrier transmission using QAM modulation," International Symposium on Wireless Pervasive Computing (ISWPC-2008), pp. 1430 - 1469 Santorini (Greece), May 2008.
- [4] S. T. Chung and A. J. Goldsmith, "Degrees of Freedom in Adaptive Modulation: A Unified View", IEEE Transactions on Communications, Vol. 49, N° 9, pp. 1561-1571, September 2001.
- [5] C. Tang and V. J. Stolzman, "Multiple User Adaptive Modulation Scheme for MC-CDMA," IEEE Global Communications Conference (Globecom'04), pp. 3823-3827. Dallas, Texas-USA. 29<sup>th</sup> Nov to 3<sup>rd</sup> Dec 2004.
- [6] I. Gutierrez, F. Bader, and J. L. Pijoan, "Radio Resource Allocation in MC-CDMA Under QoS Requirements". Book title: Multi-Carrier Spread Spectrum 2007. Chapter 5: Adaptive Transmission, pp. 207-216. Editors: S. Plass, A. Dammann, S. Kaiser and K. Fazel, Ed. Springer © 2007. ISBN: 978-1-4020-6128-8. Netherlands.
- [7] M. Shaat and F. Bader, "Power Allocation and Throughput Comparison in OFDM and FBMC Based Cognitive Radio", 22<sup>nd</sup> Meeting of the Wireless World Research Forum (WWRF'2009), Paris, France. May 2009.
- [8] IEEE 802.16e-2005, "IEEE standard for local and Metropolitan Networks. Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, amendment 2: Physical and Medium Control Layers for Combined Fixed and Mobile Operation in Licensed Bands and Corrigendum 1". Feb 2006.
- [9] PHYDYAS European project ICT-211887, [www.ict-phydvas.org](http://www.ict-phydvas.org). (visited 31<sup>st</sup> January 2011)
- [10] S. Shakkottai and A. Stolyar, "Scheduling for Multiple flows Sharing a Time-Varying Channel: The Exponential Rule", Analytic Methods in Applied Probability. In Memory of Fridrih Karpelevich. Yu. M. Suhov, Editor. American Mathematical Society Translations, Series 2, Volume 207, pp. 185-202. American Mathematical Society, Providence, RI, 2002.
- [11] I. Gutiérrez, F. Bader, and J.L. Pijoan, "New Prioritization Function for Packet Data Scheduling in OFDMA Systems", IEEE Radio and Wireless Symposium 2009 (IEEE RWS'09), pp. 405-408. San Diego, USA, Jan. 2009.
- [12] I. Gutiérrez, F. Bader, R. Aquilué, and J. L. Pijoan, "Contiguous Frequency-Time Resource Allocation and Scheduling for Wireless OFDMA Systems with QoS Support", EURASIP Journal on Wireless Communications and Networking, Vol. 2009, Article ID 134579, doi:10.1155/2009/134579. June 2009.

# An Experimental Study of ICI Cancellation in OFDM Utilizing GNU Radio System

Miftahur Rahman, Sajjadul Latif, and Tasnia Ahad

Dept. of Electrical Engineering and Computer Science

North South University

Dhaka, Bangladesh

e-mails: [mrahman@northsouth.edu](mailto:mrahman@northsouth.edu), [netpavel@hotmail.com](mailto:netpavel@hotmail.com) and [tasnia\\_ahad@hotmail.com](mailto:tasnia_ahad@hotmail.com)

**Abstract**— Inter-Carrier-Interference (ICI) is an effect that noticeably degrades the quality of the Orthogonal Frequency Division Multiplexing (OFDM) signal. In this work, several proposed ICI cancellation schemes have been tested using open source Software Defined Radio (SDR) named GNU Radio. The GNU Radio system used in the experiment had one Universal Software Radio Peripheral (USRP) module connected to a computer. The USRP had two-daughterboard (RFX-400) for both transmission and reception of radio signals in the 400 MHz band. The input data to the USRP was prepared in compliance with IEEE-802.11b specification. The experimental results were compared with the theoretical results of the proposed Inter-Carrier Interference (ICI) cancellation schemes. The comparison of the results revealed that the new schemes are suitable for high performance transmission. The results of this paper open up new opportunities of using OFDM in 400MHz band, where channels are heavily congested. The new ICI cancellation schemes can be used for Digital TV applications or for secured government communication services in the 400MHz to 500MHz band.

**Keywords**-OFDM; USRP; GNU radio; ICI cancellation; 400MHz, software defined radio;

## I. INTRODUCTION

The OFDM is a technique, where a large number of orthogonal, overlapping subcarriers are transmitted in parallel, dividing the available transmission bandwidth into narrowband sub-channels [1, 2, 3]. The separations of the subcarriers are kept as minimal as possible to get a very compact spectral utilization. Due to the use of narrow bandwidth, each sub channel requires a longer symbol period. The orthogonality of the OFDM is lost when there are channel impairments and frequency mismatch in the transmitter and receiver. This leads to inter-carrier interference, phase rotation and performance degradation of the channel [4]. To reduce the effect of ICI, different cancellation schemes [5, 6] have been proposed and were tested in the GNU radio system. GNU radio is an open source software toolkit for building software radios by making use of software that defines the transmitted waveforms and demodulators instead of hardware components [7].

The paper is organized as follows: Section II illustrates the OFDM modulation and demodulation techniques; Section III describes different ICI cancellation schemes; Section IV explains the need of the experiment, Section V discusses the experimental setup; Section VI explains the binary data preparation; Section VII is about the unique modulation

technique used; Section VIII explains software and hardware setup; Section IX describes the data retrieval process; Section X analyses the results; Section XI discusses the future scope of the technology and Section XII makes the concluding remarks.

## II. OFDM MODULATION & DEMODULATION

OFDM is a technique as shown in Fig. 1, where the input data is converted to parallel bits and mapped according to predefined standard [9]. Inverse Fast Fourier Transform (IFFT) is a vital part to convert signal from frequency domain to time domain. After IFFT the parallel data is again converted to serial data. Cyclic prefix is also added before it gets converted from digital to analog data. The input data should be prepared maintaining specific standard. According to IEEE-802.11b specification [9], the data is transferred using 64 subcarriers. In the IFFT mapping, the total 64 subcarriers in frequency domain are converted to time domain. In order to preserve the orthogonality of OFDM signal, preamble bits [9] are added. Also the cyclic prefix enables synchronization as the bits are used to detect the beginning and end of each frame and it appends the OFDM symbols one after another [9].

The received signal is demodulated according to the steps shown in Fig. 2, which is the opposite of OFDM modulation technique shown in Fig. 1.

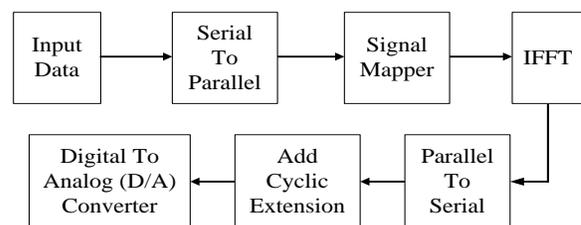


Figure 1. OFDM Modulation

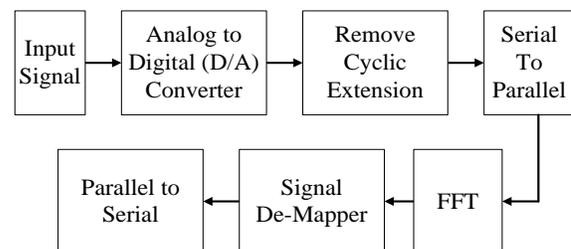


Figure 2. OFDM Demodulation

### III. ICI CANCELLATION SCHEMES

Different cancellation schemes are used in order to reduce the effect of Inter-Carrier-Interference. Among the established models, self-cancellation and modified self-cancellation techniques [11] are considered for the experiment. Also, two other new schemes have been developed and tested during the research [11].

#### A. ICI self-cancellation

In this scheme each data bit is sent through two adjacent sub-carriers, one with weight '+1' and another with '-1' [11].

$$\begin{aligned}
 Y'(K) &= Y(K) - Y(K+1) \\
 &= X(K)[-S(-1) + 2 \times S(0) - S(1)] \\
 &+ \sum_{\substack{l=0 \\ l \neq \text{even} \\ l \neq K}}^{N-2} X(l)[-S(1-K-l) \\
 &+ 2 \times S(1-K) - S(1-K+1)] \\
 &+ \{w(K) - w(K+1)\}
 \end{aligned} \quad (1)$$

In order to cancel the ICI at the receiver, adjacent sub-carriers are subtracted,  $X(K) - (-X(K)) = 2X(K)$  [11]. This enhances the data value and reduces the noise level as given in equation (1).

#### B. ICI modified self-cancellation

In this scheme each data is sent through two sub-carriers one with weight '+1' and another with '-1' [7]. The  $K^{\text{th}}$  and  $(N-1-K)^{\text{th}}$  subcarriers are used.

$$\begin{aligned}
 Y'(K) &= Y(K) - Y(N-1-K) \\
 &= X(K)[2 \times S(0) - S(N-1-2K) - S(2K-N+1)] \\
 &+ \sum_{\substack{l=0 \\ l \neq K}}^{\frac{N-1}{2}} X(l)[S(1-K) - S(N-1-l-K) \\
 &- S(1-N+1+K) + S(-1+K)] \\
 &+ \{w(K) - w(N-1-K)\}
 \end{aligned} \quad (2)$$

Similar to ICI self-cancellation, the interfering components are removed by subtracting the  $K^{\text{th}}$  and  $(N-1-K)^{\text{th}}$  sub-carriers in the receiving end as shown in equation (2).

#### C. New ICI cancellation scheme-1

In this scheme each data is sent through four adjacent sub-carriers. The  $K^{\text{th}}$  and  $(K+3)^{\text{th}}$  sub-carrier with weight '+1' and  $(K+1)^{\text{th}}$  and  $(K+2)^{\text{th}}$  with weight '-1' [1].

$$\begin{aligned}
 Y'(K) &= Y(K) - Y(K+1) - Y(K+2) + Y(K+3) \\
 &= X(K)[4 \times S(0) - S(1) - 2 \times S(2) + S(3) - S(-1) \\
 &- 2 \times S(-2) + S(-3)] \\
 &+ \sum_{\substack{l=0 \\ l=1+4 \\ l \neq K}}^{N-4} X(l)[4 \times S(1-K) - S(1-K-1) \\
 &- 2 \times S(1-K+2) + S(1-K+3) \\
 &- S(1-K-1) - 2 \times S(1-K-2) + S(1-K-3)] \\
 &+ \{w(K) - w(K+1) - w(K+2) + w(K+3)\}
 \end{aligned} \quad (3)$$

To retrieve the data at receiving side, 1st and 4th sub-carriers are added and 2nd and 3rd sub-carriers are subtracted as stated in equation (3) [1],  $X(K) + X(K) - (-X(K)) - (-X(K)) = 4X(K)$ . So the original data values become prominent.

#### D. New ICI cancellation scheme-2

In this scheme each data bit is sent through four adjacent sub-carriers, each with  $90^\circ$  phase shift [4]. This gives symmetry to the signal and greatly reduces the effect of the inter carrier interference.

$$\begin{aligned}
 Y'(K) &= Y(K) - e^{-j\frac{\pi}{2}} Y(K+1) - Y(K+2) + e^{-j\frac{\pi}{2}} Y(K+3) \\
 &= X(K) \left[ \begin{aligned} &[3 \times S(0) - 2 \times S(2) - S(-2)] \\ &+ e^{-j\frac{\pi}{2}} [2 \times S(1) - 3 \times S(-1) + S(-3)] \\ &+ e^{j\frac{\pi}{2}} [S(3) - S(1)] + e^{-j\pi} [-S(0) + S(-2)] \end{aligned} \right] \\
 &+ \sum_{\substack{l=0 \\ l=1+4 \\ l \neq K}}^{N-4} X(l) \left[ \begin{aligned} &[3 \times S(1-K) - 2 \times S(1-K+2) - S(1-K-2)] \\ &+ e^{-j\frac{\pi}{2}} [2 \times S(1-K+1) - 3 \times S(1-K-1) \\ &+ S(1-K-3)] \\ &+ e^{j\frac{\pi}{2}} [S(1-K+3) - S(1-K+1)] \\ &+ e^{-j\pi} [-S(1-K) + S(1-K-2)] \end{aligned} \right] \\
 &+ \{w(K) - w(K+1) - w(K+2) + w(K+3)\}
 \end{aligned} \quad (4)$$

In the receiving end, the 3<sup>rd</sup> and 4<sup>th</sup> are phase shifted by  $90^\circ$  and added to 1<sup>st</sup> and 2<sup>nd</sup> sub-carriers respectively. Then the results of 2<sup>nd</sup> and 4<sup>th</sup> phases are again shifted by  $180^\circ$  to get the original data as shown in equation (4).

### IV. THE NEED FOR THE EXPERIMENT

In OFDM technique a large number of orthogonal subcarriers are transmitted in parallel dividing the available transmission bandwidth into narrowband sub-channels [1,2,3]. The separations of the subcarriers are kept as minimal as possible to get a very compact spectral utilization. The channel impairments and frequency mismatch in the transmitter and receiver, sometimes lead to inter-carrier interference (ICI) and phase rotation, resulting in performance degradation of the channel [4]. In order to reduce the effect of ICI, different cancellation schemes [5,6] have been proposed. These schemes help to reduce the effects of ICI by cancelling out the interference from carrier. As the data itself is used to cancel the interference, no extra encoding is needed for error correction. The results from hardware simulations give better insight about the real life problems than does software simulations. The GNU radio is a cost effective and flexible implementation platform that can be used to verify the functionalities as well as the performance of advanced models of wireless technologies in a real time setup. The advent of high speed processors with increased computational capability is making the software and hardware of GNU radio much closer to the antenna. One of the main advantages of SDR is that it is software reconfigurable, which leads to significant design simplification.

V. EXPERIMENT

In the experiment, an input audio data was modulated using OFDM technique. The setup of the experiment is illustrated in Fig. 3 where initially a recorded audio voice signal was taken and it was imported into MATLAB where all the OFDM processing along with ICI cancellation encoding was performed and the resulting data was sent to a GNU Radio Companion (GRC) running computer. That computer was connected to a USRP transmitter. GNU Radio processed and transmitted the signal over the air. At the other end, an USRP receiver captured the signal, processed in GNU Radio and sent it back to MATLAB. The MATLAB demodulated and decoded the data and regenerated the sound and sent it to the speaker.

VI. BINARY DATA PREPARATION

The steps of preparing binary data in MATLAB are shown in the Fig. 4. In the experiment the input signal was a pre-recorded audio voice signal as shown in Fig. 5. The duration of the audio signal was two seconds and it was sampled at 8 kHz. The signal was compressed using the A-law companding technique [2] with A=80. The signal then quantized using pre-determined step size. In the experiment the step size was 512 in the range of -256 to 255. Then 256 were added to all the quantized values to make it positive integer and the sample quantized. The quantized data then converted into binary string using MATLAB function. The system used 9 bits for each level. With total 512 levels ranges from 0 to 511. The produced data was binary strings. The strings were converted to individual number bits for the use of modulation. The numbers were then modulated either using Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK). In case of BPSK [9], the simplest form of digital modulation technique, the phase of a constant amplitude carrier signal is switched between two values of 1 and 0. The two phases are separated by 180°. Here all the 0's became -1 and 1's remained as 1. After the BPSK/QPSK modulation, each of the different ICI cancellation schemes (as discussed in section III) were implemented.

VII. MODULATION

Then OFDM modulation was performed according to IEEE-802.11b specification [9] as discussed in section II.

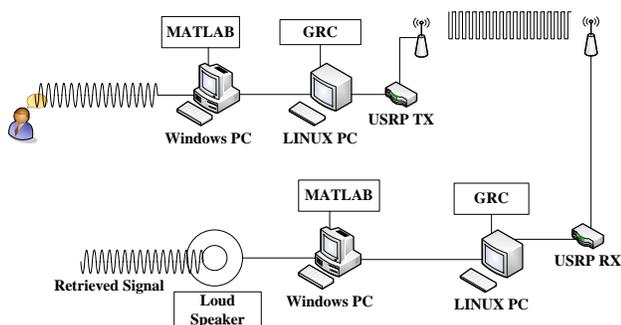


Figure 3. Experimental setup

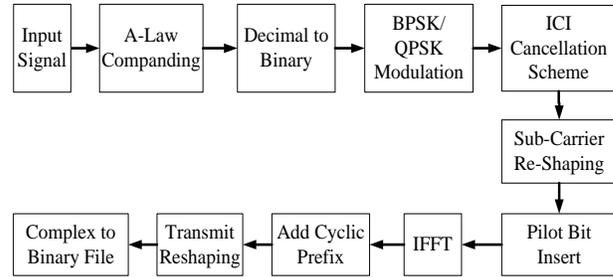


Figure 4. Steps of preparing binary data.

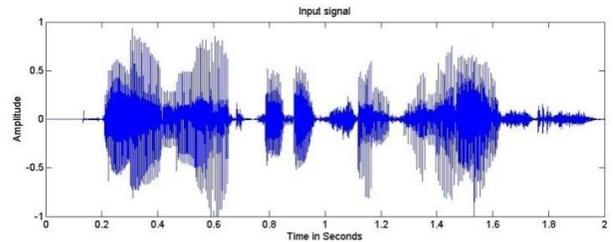


Figure 5. Message signal

Total 52 subcarriers were divided into two groups of 26 bits. A null value was added between the two groups. The subcarrier number 27 to 52 including the null value was placed in front of the signal. The eleven null bits or pilot bits were added, and the rest of the subcarriers were placed after the pilot bits [9]. This made 64 subcarrier long signals as shown in Fig. 6. Here each number denotes subcarrier number.

In the IFFT mapping, the total 64 subcarriers in frequency domain were converted to time domain by IFFT. A 64-point IFFT was used in the experiment. The coefficients 1 to 26 were mapped to the same numbered IFFT inputs, while the coefficients -26 to -1 were copied into IFFT inputs 38 to 63. The rest of the inputs, 27 to 37 and the 0 (dc) input, were set to 0 or null as pilot bits [9]. The IFFT performs the process of transforming a spectrum (amplitude and phase of each component) into a time domain signal. An IFFT converts a number of complex data points of length that is a power of 2, into the time domain signal of the same number of points [9]. After IFFT the signal domain value of 64 subcarriers were placed, where the 49<sup>th</sup> and 64<sup>th</sup> subcarrier numbers are copied at the beginning of the signal as preamble as shown in Fig. 7 [9]. The addition of preamble bits allows the orthogonality of OFDM signals to be preserved. The total numbers of bits after a preamble addition was 80 bits. The addition of these bits enables synchronization as the bits were used to detect the beginning and end of each frame appending the OFDM symbols one after another. This was the 80 subcarrier long signal that was transferred. Fig. 8 was the signal that was transmitted. This signal was converted from complex to binary and was saved in a file to be exported to GNU Radio system. A special method was used to make binary conversion where the real and imaginary parts of the complex signals were separated

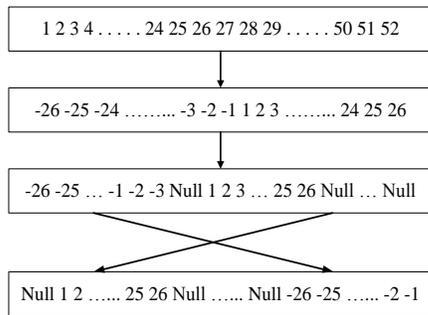


Figure 6. Sub-Carrier Re-Shaping

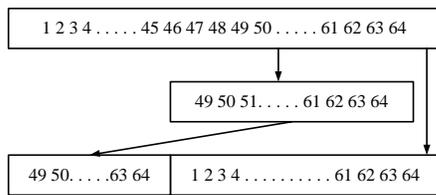


Figure 7. Adding Cyclic Prefix

and written in a file interleaved the data. This way string of real and complex valued signal was mapped in a binary formatted file shown in Fig. 9. The GNU Radio can utilize the file as a source to transmit over the air.

VIII. GNU RADIO AND HARDWARE SETUP

GNU Radio is an open source software toolkit used for the experiment [7, 8]. In the GNU Radio Companion (GRC) software the file source block was used for feeding the binary file which was passed through a constant multiplier and was transferred to the USRP sink to transmit Over-the-Air (OTA). Fig. 10 illustrates the implementation of the hardware parts of the GNU radio system. The binary file was transferred to the “low cost” hardware called USRP and was passed onto the transmitter daughterboard which are RF frontends. The signal was received by a receiver daughterboard. Similarly, like the transmitter, the signal passed through the receiver USRP and this captured signal was saved in a binary file using the File Sink block [7]. The spectrum can be viewed using the FFT Sink block [7]. All the ICI OFDM BPSK/QPSK modulations were performed in MATLAB and the binary file was extracted from there.

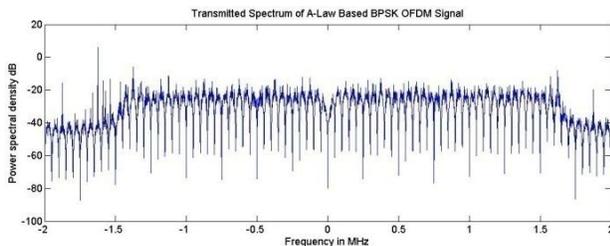


Figure 8. Transmitted Signal

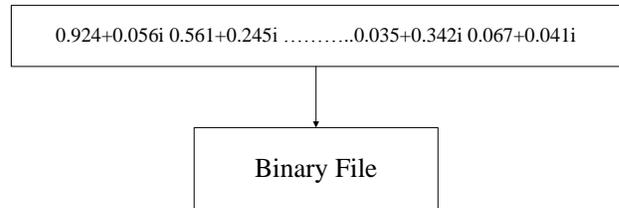


Figure 9. Complex value to Binary File Creation

IX. DATA RETRIEVAL

The received binary file was imported to MATLAB and the resulting received spectrum as shown in Fig. 11 was displayed. From the spectrum, densely congested ICI cancellation signal was noticed and also noise components were visible. Fig. 12 is the opposite of the data preparation as shown in Fig. 4. Here the received signal was saved to a binary file. From the binary file, the complex values were extracted, some representative samples are shown in Fig. 13. The 16 bits of cyclic prefix that was added was removed followed by FFT which converts the time domain signal into frequency domain. The subcarrier bits have been extracted from this signal and the 12 pilot bits are removed. Step by step, the ICI cancellation schemes were performed followed by the BPSK/QPSK demodulation as shown in Fig. 12. The data was then converted to binary digits and A-law decompression was implemented and the audio voice signal was regenerated.

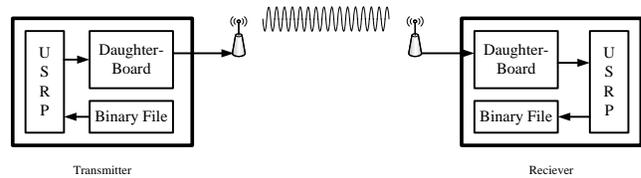


Figure 10. Complex value to Binary File Creation

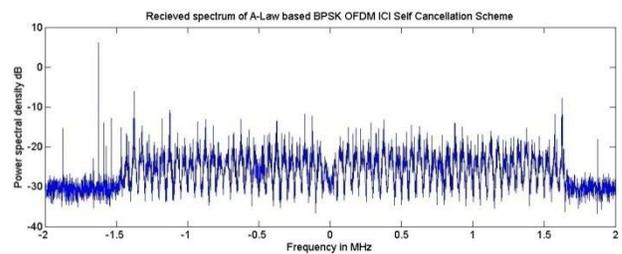


Figure 11. Received Spectrum

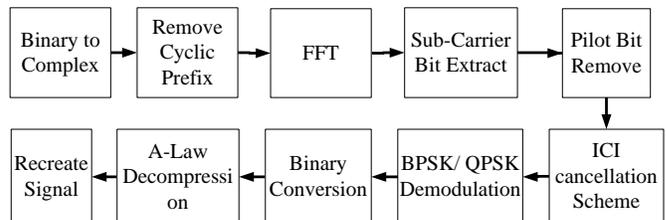


Figure 12. Data Retrieval.

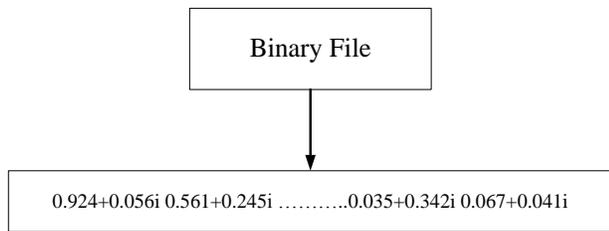


Figure 13. Binary to Complex Conversion

### X. RESULT AND ANALYSIS

Fig. 8 shows the transmitted spectrum of the OFDM BPSK signal which was generated in compliance to IEEE specification [9]. The spectrum has 80 sub-carriers according to the standard. There are 16 cyclic prefixes or preamble (8 on each side). The middle portion contains the 64 IFFT data sub-carriers with 12 pilot bits. In total 52 data sub-carriers were embedded in frequency domain [9]. Fig. 11 shows the received spectrum of the same signal at the receiving end. The signal suffered attenuation and inter-carrier-interference (ICI) during transmission. From the figure the effect of Additive White Gaussian Noise (AWGN) noise can be clearly seen by looking at the cyclic prefixes. The visible figure can be said to have dominant noise components. The total power level also decreases significantly. There are many peaks and sudden drops in the spectrum which were not present in the original transmitted spectrum. The effects of ICI are also clearly visible from the figure as the adjacent carriers are almost overlapping. Fig. 14 is the original message signal, which is very nicely shaped and does not have any distortion pattern or sudden peaks or valleys. The received signal, however, is not fully free of noises. So, the reconstructed signal could not be retrieved as such from the received spectrum. Fig. 15 shows the reconstructed signal in time domain. The figure clearly displays the noises and ICI effects on the message. There are many drops across the message. However these problems can be resolved by using the new schemes. Fig. 16 shows received spectrum with ICI new scheme – 1 BPSK OFDM. From the figure it can be deduced that the spectrum has improved from the previous scheme. The regenerated signal also has low distortion. Fig. 17 is the regenerated signal.

In scheme-2, the performance improved a lot. Basically in the test, the scheme-2 was the most efficient signal. In that scheme the retrieved signal was almost same the original signal. From Fig. 18 the improvement in the dense sub-carriers can be easily seen. The improvement is clearly visible in the cyclic prefix and also in the data sub carriers.

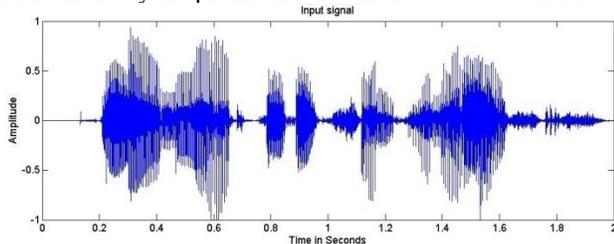


Figure 14. Transmitted Message

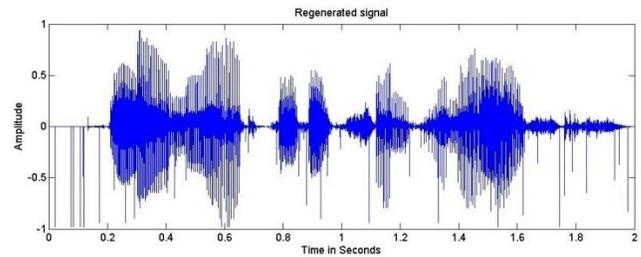


Figure 15. Regenerated ICI-self cancelled signal

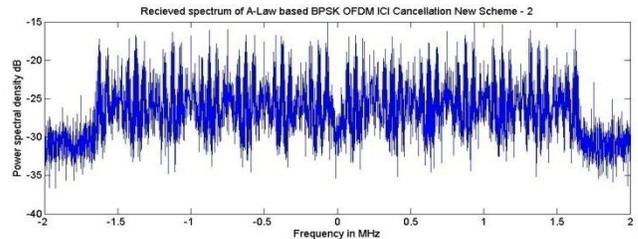


Figure 16. Received Spectrum of New Scheme – 1

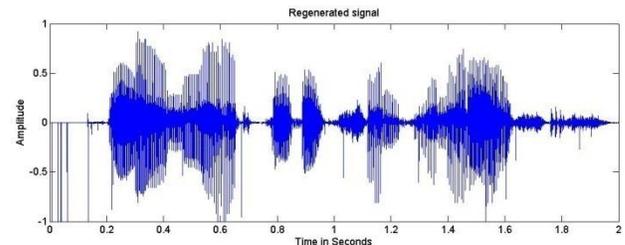


Figure 17. Regenerated New scheme –1 signal

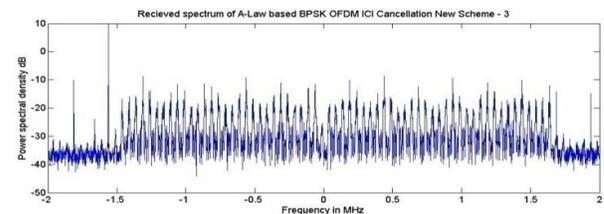


Figure 18. Received Spectrum of New Scheme – 2

The sub carrier signals have retained their original shapes and power level. The signal has not been distorted severely due to AWGN noise and ICI as observed from the spectrum. The regenerated signal also shows the improvement as seen from Fig. 19. There are not many drops in the signal. The regenerated signal is almost 99% close to original signal.

This improvement in the performance justifies the feasibility of the new schemes in OFDM to cancel the ICI and improve performance. The error percentages in new schemes are shown in Table 1 where the new scheme–2 gave the best CIR values. The performance improves because of the symmetrical transmission of the subcarriers.

Though the new schemes improve the performance to a great extent, the bandwidth utilization has not been efficient. Normal OFDM uses full bandwidth of the channel whereas the self-cancellation schemes use only half of it. The new schemes are less efficient with the use of only quarter of the

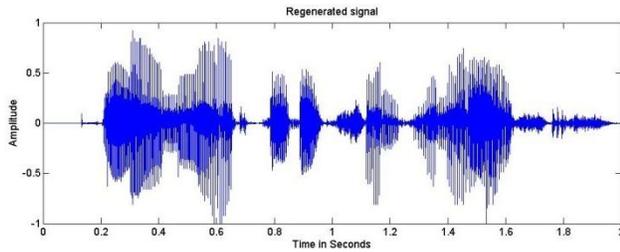


Figure 19. Regenerated signal for New Scheme – 3

TABLE I. ERROR PERFORMANCE

Schemes	Results		
	No. of transmitted bits	Error bits	% bit error
ICI self	184320	293	0.159%
New scheme -1	184320	171	0.093%
New scheme -2	184320	1	0.00001%

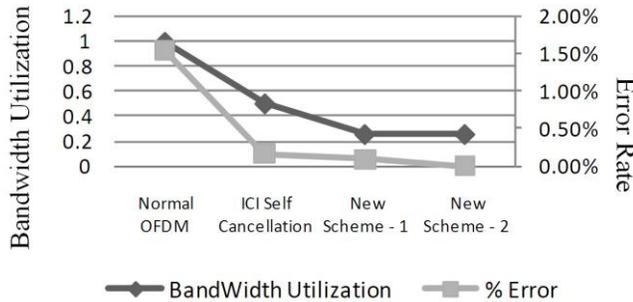


Figure 20. Performance comparison of schemes

bandwidth as shown in Fig. 20. Theoretically, the data rate also decreases and new schemes has almost quarter the data rate than the normal OFDM. If the transmission gives more emphasis on performance and data rate is not a concern, the new schemes are better choice than the already established self-cancellation schemes. The New Schemes also require greater resources to process the cancellation on both transmitter and receiver.

### XI. FUTURE SCOPE

The outcome of the research shows the feasibility of the implementation of OFDM in GNU Radio system, which provides a new field to explore. The project was used to verify the viability of the schemes that was considered for testing. Implementing the ICI cancellation scheme gives an opportunity to retrieve signals with better accuracy, so further test can be done using the GNU Radio involving ICI schemes. The availability of the hardware will enable the researchers to perform more practical implementation and hence to obtain more accurate results.

This experiment was conducted in 400 MHz frequency band. Further investigation of the OFDM systems also can be done in 2.4 GHZ or 5GHZ frequency bands in the future. The use of OFDM in newer technologies like LTE, WiMAX, Wi-Fi, Digital TV can be tested for practical feasibility in

GNU Radio using the techniques used in this experiment. The research also opens up the opportunity to think and explore the options for 400 MHz frequency band similar to the initiatives taken in Australia [11]. The use of OFDM can help utilize the full capacity of this band to a greater extent. The new schemes can be tested in Digital TV services for getting the best possible clarity of the channel.

### XII. CONCLUSION

The OFDM system, because of its robustness to multipath fading, has become popular in recent years. The project concentrated on testing the established schemes which were already developed and seemed to have improved the performance of the OFDM system. The schemes included ICI self-cancellation, and modified ICI self-cancellation. The two new schemes as mentioned in reference [10] were also tested in hardware radio transmission. The focus of this project was to check the viability of these schemes. The result shows significant improvements in new schemes. But the inefficient bandwidth utilization and low data rate restrict its use only in high performance systems. The digital HD TV, secured government communication network on 400 MHz band, possibilities of data transfer through amateur radio band, emergency surveillance and rural telephone services [11] can utilize the effectiveness of the new schemes because in terms of performance the new schemes give far better performance than already established ICI cancellation techniques.

### ACKNOWLEDGMENT

We would like to express our gratitude especially to Mr. Shabbir Ahmed of Victoria University, Australia, for his endless hours of discussion, suggestions, ideas, advice, and proper guidance throughout the experiment. We would like to thank Mr. Preyom Kanti Dey and Mr. M. F. Rabbiur Rashid of North South University, Dhaka, Bangladesh, for their valuable advices and many useful feedbacks in conducting this research. We also like to thank the entire GNU Radio development team [7] and GNU Radio forum members [7] for their support and suggestions.

### REFERENCES

- [1] Vijay K. Carg, Wireless Communications and Networking, First Ed., Morgan Kaufmann Publishers, 2008.
- [2] T. S. Rappoport, Wireless Communication principles and practice, 2nd Ed, Prentice- Hall, 2002.
- [3] Richard Van Nee and Ramjee Prasad, "OFDM for wireless multimedia communications", Artech House Publishers, 2000.
- [4] Miftahur Rahman, Preyom Kanti Dey, and M. F. Rabbiur Rashid, "Improved ICI Self Cancellation Scheme for Phase Rotation Error Reduction in OFDM System", 2011 International Conference on Network Communication and Computer (ICNCC 2011), March 19-20, 2011, New Delhi, India.
- [5] Hen-Geul Yeh, Yuan-Kwei Chang, and Babak Hassibi, "A Scheme for Cancelling Intercarrier Interference using Conjugate Transmission in Multicarrier Communication Systems", IEEE Transactions on Wireless Communications, Vol. 6, No. 1, January 2007.
- [6] Hen-Geul Yeh and Yuan-Kwei Chang, "A Conjugate Operation for Mitigating Intercarrier Interference of OFDM Systems", IEEE Xplore, 0-7803-8521-7/04/\$20.00 © 2004 IEEE.

- [7] GNU Radio. Retrieved June 12, 2010, from GNU Radio: <http://www.gnuradio.org/>
- [8] Ettus product. Retrieved June 27, 2010, from Ettus Research LLC. : <http://www.ettus.com>
- [9] IEEE Computer Society, "IEEE Standard for Information Technology—Telecommunications and information exchange between systems— Local and metropolitan area networks— Specific requirements", IEEE Standard 802.11, 12 June 2007, Chapter-17, pp. 591-635. Retrieved May 17, 2010, from IEEE: <http://www.ieee.org/>.
- [10] Australian Communications and Media Authority (ACMA), "Spectrum Options: 403-520 MHz". Retrieved June 20, 2010, from ACMA: <http://www.acma.gov.au/>.
- [11] Brahmaji T.A.R.K., "An Efficient ICI Cancellation Technique for OFDM Communications Systems", M.S. Thesis, Dept. of Electrical Engineering National Institute of Technology Rourkela, 2009.
- [12] Mutsawashe Gahadza, Minseok Kim, and Jun-Ichi Takada, "Implementation of a Channel Sounder using GNU Radio Opensource SDR Platform", Grad. School of Engineering, Tokyo Institute of Technology, March 2009.
- [13] Timothy J. O'Shea, T. Charles Clancy, and Hani J. Ebeid, "Practical Signal Detection and Classification in GNU Radio", Laboratory for Telecommunications Sciences, US Department of Defense.
- [14] Álvaro Palomo Navarro, Rudi Villing, and Ronan Farrell, "Software Defined Radio Architectures Evaluation", Proc. of the SDR '08 Technical Conference and Product Exposition Chapter-09, pp. 251-256.
- [15] Naveen Manicka, "GNU Radio Testbed", MS in Computer Science, University of Delaware, Spring 2007.

# Dynamic Group PCI Assignment Scheme

Xu Zhang<sup>1</sup>, DiBin Zhou<sup>2</sup>, Zhu Xiao<sup>3</sup>, Enjie Liu<sup>1</sup>, Jie Zhang<sup>1</sup> and Andres Alayon Glasunov<sup>4</sup>

<sup>1,2,3</sup>Institute for Research in Applicable Computing, <sup>4</sup>Electromagnetic Engineering

<sup>1</sup>University of Bedfordshire, <sup>2</sup>Hang Zhou Normal University, <sup>3</sup>Hunan University, <sup>4</sup>School of Electrical Engineering

<sup>1</sup>Luton, UK, <sup>2</sup>Hang Zhou, China, <sup>3</sup>Chang Sha, china, <sup>4</sup>Stockholm, Sweden

Email: <sup>1</sup>[xu.zhang1,Enjie.Liu, Jie.Zhang]@study.beds.ac.uk, <sup>2</sup>dibinz@zju.edu.cn, <sup>3</sup>zhxiao@hnu.edu.cn, <sup>4</sup>aag@ee.kth.se

**Abstract**—The Physical Cell Identity (PCI) is used to identify femtocell in LTE femtocell as the physical layer identity. Due to the fact that a) the PCIs are normally allocated without planning; and b) the limitation of the number of PCI, the cross-tier PCI confusion problem has arisen. The current solution in 3GPP Release 9 is to use Cell Global Identity (CGI) associated with PCI to solve this problem of confusion. However, using CGI has many serious drawbacks. Especially CGI might cause the inbound-handover is failure. In order to mitigate the CGI problem, this research proposes a dynamic PCI assignment scheme of DG-PCI in the macrocell and femtocell environment. The proposed scheme is tested by using system-level simulation. The simulations showed that the approach offers an optimal PCI distribution between macrocell and femtocell. The test showed that the CGI problem can be solved, and hence increase the successful rate of inbound-handover.<sup>1</sup>

**Index Terms**—PCI, CGI, LTE femtocell, Graph Colouring, PCI release

## I. INTRODUCTION

A femtocell is a small cellular base station used by operators to extend indoor service coverage and enhance overall network performance. Since 2008, it has gradually become a mainstream technology supported by industrial standardization bodies such as the 3rd Generation Partnership Project (3GPP). As a consequence, an explosive deployment has been anticipated by industry leaders, which pose new challenges [1].

In the LTE femtocell systems, the Physical Cell Identity (PCI) is used by User Equipment (UE) to identify a femtocell. Unfortunately, the number of PCIs is limited to 504 and they are normally allocated without planning and network operator intervention. Reflecting this, the PCI assignment problems which are called single-tier PCI collision and confusion (conflict) have recently been addressed for LTE network [2][3]. In LTE femtocell system, if a large number of femtocells are deployed under a macrocell coverage area and there are more than one femtocell-neighbour cells assigned with the same PCI due to PCI reuse, the network is unable to identify the femtocell in the measurement report. This may cause inbound-handover (handover from macrocell to femtocell) preparation to be directed to a wrong target and results in handover failure [4]. This is called cross-tier PCI confusion. Moreover, the PCIs

reuse in femtocell identification is very common and the cross-tier PCI confusion problem could potentially become quite destructive.

The remainder of this paper is as follows. Section II introduces some related work on PCI allocation; Section III introduces the proposed approach and some useful definitions; Section IV gives the scheme discussion and also proposed a method as Unutilised femtocells PCI release; Section V is the performance evaluation for the scheme. In Section VI, we have the conclusion.

## II. RELATED WORK

Some related works are described as below:

### A. Review on the Latest Standardisation

In order to solve the cross-tier PCI confusion, a Cell Global Identity (CGI) has been proposed to identify the cells in 3GPP release 9[5]. This approach temporarily solves the cross-tier PCI confusion problem. However, compared to PCI, using CGI has many serious drawbacks.

Firstly, PCI is a reference signal sequence which means that the UE reads the identity of the target cell in a very short time - up to 20ms in LTE. Yet, CGI is not a reference signal sequence, and it needs to be obtained by reading the system information which requires a large measurement time gap (e.g., up to 160 ms (milliseconds) for LTE) [5][6].

Secondly, during this relatively long measurement gap (e.g. 160ms), the UE cannot receive or transmit any data to or from the serving cell, thus, it probably leads to unnecessary service interruption, such as a call drop in the situation where the signal strength fades rapidly [5]. This becomes more critical in a busy and dense network scenario and causes many handover failures [7].

Lastly, a long measurement time is required to obtain the system information which causes concerns regarding the UE battery life[4]. The UE battery life is important for battery-hungry multimedia applications.

It is therefore preferable to use PCI rather than CGI to identify cells. However, as mention previously, the PCI has a conflict problem. This research aims to solve the drawbacks of CGI and meanwhile, mitigate PCI conflict in the cross-tier network environment, with co-existence of macrocell and femtocell.

<sup>1</sup>Our work is supported by the Fundamental Research Funds for the Central Universities (531107040276) and State Key Laboratory of Integrated Services Networks Open Project (ISN12-05).

## B. Review of the Literature

In [8], the authors proposed a solution for PCI conflict problem in single-tier LTE networks. A Graph Colouring-based mathematical method has developed for the PCI auto-configuration of LTE network. Each nodes ID (PCI) is assigned a color, and the neighboring nodes get different ID based on graphic coloring theory.

In [2], the authors proposed to use mobile measurements to update the Neighbour-Cell-List (NCL) in single-tier LTE networks, in order to detect PCI conflict. The solution proposes that if the PCI conflict appears, the mobile send this information to Core Network (CN) and Operation Support System (OSS) will require the involved conflict cells to change their PCIs.

In the above papers, the authors analysed the PCI conflict issue in single-tier LTE system. However, none of them have included the impact of the layered structure of a heterogeneous network with a combination of macrocell and femtocell (cross-tier network) on the PCI auto-configuration.

In [9], the authors proposed an automation PCI allocation system (APCIAS) and APCIA method to allocate the PCI in a cross-tier LTE networks in order to reduce the planning time of PCI. In the paper, the researchers used the cell information which includes cell state information, type information and neighbour list information to create the PCI resource and also allocate the PCIs.

In [10], the authors proposed an automatic assignment of femtocell PCIs depending on different access modes for network optimization in order to reduce the operational expenditure for PCI allocation. The proposed scheme autonomously detects the neighbour cells of target femtocell and sends the neighbourhood information to the centre controller. By using a centre controller, the PCIs can be assigned in an optimal way.

In [6], the authors proposed an approach to reduce the time spent on femtocell cell selection/reselection. This approach uses two groups of PCIs, a femtocell group which is a reuse-PCI-group and a macrocell group which is a unique-PCI-group. When the UE moves into a new macrocell service, it automatically obtains the network information of this macrocell, which sets certain PCI numbers for macrocell and femtocell. During the handover process, the UE easily detects whether the target device is a macrocell or a femtocell by using this informations and leading to a reduction in unnecessary signalling with the CN and identification time.

In the above papers, the authors proposed the cross-tier PCI allocation scheme. However, none of them solve the cross-tier confusion problem.

## III. DYNAMIC GROUP PCI ASSIGNMENT SCHEME

This scheme is described as below:

### A. Busy Femtocell

In Dynamic Group PCI Assignment Scheme (DG-PCI), the traffic density information is the most important features to describe the network situation. Busy Femtocell (BFemtocell)

concept is introduced to describe this. The principle of this concept is that there might be some femtocells which have a higher numbers of inbound-handovers than others during the same time period. For the operators, more attention should be paid to these BFemtocells in order to maintain a high quality of service since they have a potentially higher probability to be involved in cross-tier PCI confusion and result in handover failure due to CGI drawback.

### B. Dynamic PCI Group Assignment

In the proposed DG-PCI, PCIs are categorised into three groups - CSG group (Closed Subscriber Group), non-CSG group, and macrocell groups. Furthermore, in CSG and non-CSG groups, PCIs are further categorised into two sub-groups - unique and reuse group, shown in Fig. 1. The PCIs are set according to their group and sub-group, it would be updated depending on the various environments.

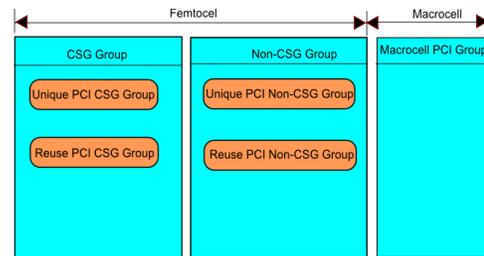


Fig. 1. Proposed Dynamic PCI Group Structure

Since this dynamic PCI group method is similar to the profile of proposed the PCI distribution in [6] and [11], it inherits their advantages such as the fast CSG cell, non-CSG cell and macrocell selection/reselection.

The unique PCI group is used to offer unique PCI to BFemtocells. As we know cross-tier PCI confusion may happen when reusing PCIs under the large range of macrocell service, thus, if an approach guarantees unique PCIs being allocated to femtocells, there will be no confusion in inbound-handover. For obvious reason, it is sensible to offer the unique PCI to BFemtocells whenever the system has unique PCI available.

### C. The Scheme used for Different Access Policies

3GPP defined three types of cell access modes: closed, hybrid and open mode. The following section discusses applying the scheme in different access scenarios.

#### 1) The cross-tier confusion case in different access modes:

a) For the closed and hybrid access modes: Cross-tier PCI confusion occurs as follows: during inbound-handover, after UE has determined that this cell is a closed or a hybrid cell, the UE needs to obtain the PCI of the target femtocell to achieve handover. Since there might be multiple femtocells within the coverage of the Serving Radio Network Control (SRNC) that have the same PCI ID, cross-tier confusion may occur.

b) *For the open access mode:* Cross-tier PCI confusion occurs as follows: during the inbound-handover, after the UE has determined that this cell is an open cell, the UE needs to obtain the PCI of the target femtocell to achieve handover. As the femtocell with open access mode operates just like a normal LTE cells, these cells' PCI could be included in NCL of the macrocell. There might be multiple same femtocell PCI within the NCL of the macrocell which involve the cross-tier confusion issue.

c) *Comparing these confusion cases in different modes:* Three modes contain almost the same process involving the PCI confusion by using the same PCI. Therefore, giving a unique PCI to identify BFemtocells in order to mitigate the drawbacks of CGI is fit for both of the confusion cases.

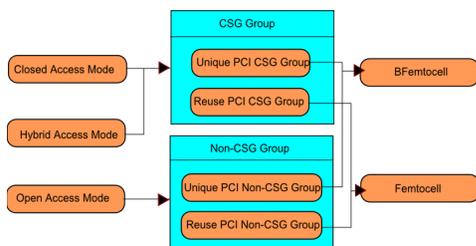


Fig. 2. PCI Distribution in femtocell PCI Group

2) *The PCI IDs in different access modes distribute in CSG/non-CSG sub-groups:*

a) *According to release 9, the System Information Block Type1 (SIB1) uses two parameters: CSG-Indication and CSG-Identity:* The closed mode cells have a CSG-Identity and CSG-Indication bit set to TRUE, hybrid model cells have a CSG-Identity and CSG-Indication bit set to FALSE, open mode cells do not have the CSG-Identity and CSG-Indication bit set to FALSE.

b) *The PCI IDs in different access modes distribute in CSG/non-CSG sub-groups:* Since the CSG-Identity ID involves using the PCI ID [14] and both of the hybrid and close cell mode support CSG-identity ID, closed and hybrid access mode can be treated as the same CSG group, using the set range of PCI IDs (CSG-identity ID). The open access mode is called non-CSG-group also using the set range of PCI IDs (CSG-identity ID). Moreover, there are two sub-groups in each CSG-group/non-CSG group, the PCI in unique PCI sub-groups is used in identification of the BFemtocell which may be CSG or non-CSG mode in order to mitigate the drawbacks of CGI, and also the reuse PCI sub-group is used in identification of the normal femtocell which may be in CSG or non-CSG mode as shown in Fig. 2.

#### D. The Proposed Scheme

The DG-PCI Flow Chat is shown in Fig. 3.

In the graph,  $P_{CSG-unique}$  is the number of unique PCIs for CSG cells;  $P_{CSG-reuse}$  is the number of reuse PCIs for

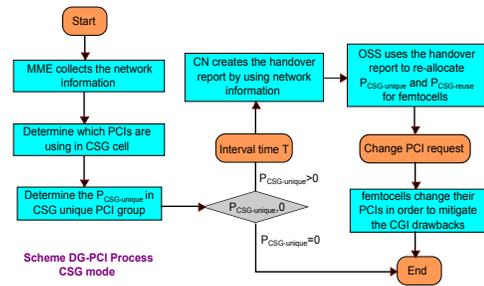


Fig. 3. DG-PCI Flow Chat

CSG cell. Since non-CSG and CSG have most of the same PCI confusion mentioned early, the process in the proposed scheme for non-CSG and CSG cell scheme can be considered the same process. The author provides the CSG mode flow chat of the proposed scheme process instead of the chat of both the non-CSG and CSG mode.

## IV. DISCUSSIONS OF THE PROPOSED SCHEME

In this section, we will analyse some important issues for DG-PCI.

### A. Determination of the BFemtocell in LTE Femtocell System

In a real network scenario, the handover procedure is associated with the Mobility Management Entity (MME). During the inbound-handover, severed femtocell PCI or CGI information is stored in MME. Due to the fact that PCI is dynamically allocated and CGI is statically allocated, in order to ensure the data on the list is available at any time, it is desirable to use CGI ID for data recording.

MME is connected with a Home Subscriber Server (HSS), which is responsible for femtocell management, authentication and authorization, the mapping of PCI to the unique CGI and the NCL of each femtocell could be obtained from them [12]. This CGI list will be build by using HSS in order to find the corresponding CGI ID depending on the temporary PCI stored in the MME.

After a interval time of T, the MME will send the recorded CGI list to CN. CN will build the BFemtocell list (Inbound-Handover Report) according to the CGI list and permit OSS [2] to send the updating PCI request to the BFemtocell.

### B. Graph Colouring-Based Method in Single-Tier PCI Conflict Free

The number of unique PCIs is quite an important issue in our scheme. The  $PCI_{unique}$  is equal to the total number of PCIs take off the  $PCI_{reuse}$  and  $PCI_{reuse}$  depends on the single-tier conflict. Reflecting this, the  $PCI_{unique}$  is relative to single-tier conflict. We support using Graph Colouring-based mathematical for PCI assignment in order to find the minimum of  $PCI_{reuse}$  and also avoid single-tier PCI conflict in cross-tier LTE femtocell. The method is described as below:

```

Begin
Do Sort the new femtocells in order by which femtocell has the highest nNeighbours
For (set i=0; i<=Cell_list_size()-1; i++) // Choose a new cell from the list Cell_list in order.
Do D_set = function Build_PCiset(Cell_list[i].dNeighbours);
Do N_set = function Build_PCiset(Cell_list[i].nNeighbours);
Do NN_set = function Build_PCiset(Cell_list[i].nnNeighbours);
If (NN_set[D_set]&&N_set!=null)
Do P_set=NN_set[D_set]&&N_set;
Do call function PCI_assignment(P_set);
Else
If (U_set[D_set]&&N_set!=null)
Do P_set=U_set[D_set]&&N_set;
Do call function PCI_assignment(P_set);
Else
If (A_set[D_set]&&N_set!=null)
Do P_set=A_set[D_set]&&N_set;
Do call function PCI_assignment(P_set);
Else
Do process failure; // PCI run out
//start loop again;
End
Function Build_PCiset(Neighbours) //function to find the PCI_set for different Neighbours list;
For (set j=0; j<=Neighbours.size()-1; j++)
If (in Neighbours[j], there are some new cells without PCIs)
Do PCI_set will not collect this cell's PCI;
Else
Do PCI_set collect there PCIs in Neighbours;
End
Return PCI_set;
Function PCI_assignment(P_set) //function to assign the PCI to new cell;
Do Sort P_set in C_PCI descending order;
Do Select the first value;
Do PCI assignment success;
End
    
```

The main idea of this method is that to find the proper PCI IDs in the  $NN_{set}$  of the target femtocell. If we find the PCI IDs in the  $D_{set}$ , it will result in the single-tier conflict and if we find the PCI IDs  $N_{set}$ , it will result in cross-tier confusion. Since the PCI ID are reused to allocate the femtocells, this method could find the minimum of  $PCI_{reuse}$ .

TABLE I  
PARAMETERS IN PROPOSED METHOD

Name	Meaning
dNeighbours	New cell's direct neighbours
nNeighbours	New cell's dNeighbours' neighbours
nnNeighbours	Neighbours or nNeighbours
$A_{set}/U_{set}$	All available/used PCI in the network
$D_{set}$	PCI set of dNeighbours
$N_{set}$	PCI set of nNeighbours
$NN_{set}$	PCI set of neighbours of nNeighbours
$P_{set}$	Result set which PCI is choosing from
$PCI_{set}$	Number of PCI in target Neighbours list
$C_{PCI}$	Number of times the PCI is used

C. Dynamic PCI Group using in PCI Assignment

Femtocells are plug-and-play devices, if a new femtocell joins the network, it might interfere with the system balance which would result in the single-tier conflict. Therefore, ideally the PCIs should be able to transfer between different groups to cope with the dynamic number of femtoells. However, the number of PCIs in a marocell PCI group is set at the system level [6], thus this transfer will not be considered in this research. In this sub-section, we will only consider the PCI transfer between unique PCI and reuse PCI group.

Unique PCI transfer to reuse PCI group. If there are not enough reuse PCI to sign new femtocells, the unique PCI group dynamically transfers its PCIs to reused PCI group in order to ensure the network has enough PCIs to remain single-tier conflict free. However, if the unique PCI numbers run out,

that means all PCIs need to be used to ensure the system is conflict free. Unfortunately, our approach does not work in this extreme case, yet this kind of scenario is rare.

Reuse PCI transfers to unique PCI group. When some femtocells are turned off, the number of reused PCIs could be more than the system requested, therefore, some of the reuse PCIs should be released, this will be analysed in detail in the next section.

D. Unutilised femtocell PCI Release Method

The PCI release is the decisive feature to obtain unique PCI resource for scheme DG-PCI. A PCI release method is proposed in this research to tackle this problem.

This method includes unique and reuse PCI released. The unique PCI release method releases the femtocells that have unique PCIs identity. The reuse PCI release is used in dynamic PCI group to release reused PCIs. The proposed PCI release is described as below:

```

Begin
Do Determining the Femtocell_off list;
Do Determining the Femtocell_on list;
For (set i=0; i<=Femtocell_off_size()-1; i++) // Choose a femtocell from the list
For (set j=0; j<=Femtocell_on_size()-1; j++)
Do Set a=true;
//Check the PCIi in femtocelloff is used in femtocellon list;
If (PCIi in femtocellon[j]==PCI in femtocelloff[i]) // PCIi is the PCI of chosen femtocell;
Do break; //jump out current loop;
Else
//Check the PCIi in Femtocelloff is also used by other members in Femtocelloff;
For (set k=0; k<=Femtocelloff.size()-1; k++)
If ((i!=k)&&(PCIi in femtocelloff[j]==PCI in femtocelloff[k]))
Do Set a=false;
Do break; //jump out current loop;
End
If (a==false)
Do break; //jump out current loop;
Else
Do PCIi = PCIi in femtocelloff[j]; // PCIi is the unique PCI
Do Add PCIi to Listunique;
End
End
    
```

However, there is a problem. If the femtocell is normally in turn off mode, the CN can easily detect this femtocell and also execute the PCI release procedure. If the femtocell is not in normal turn off mode, meaning that the users just unplugged the power socket to shut down the femtocell, it is difficult for CN to identify the femtocell statics and results in a failed reuse PCI release procedure.

In order to solve this problem, ping can be used to check the femtocell status. Since ping transmits in very small packages, it will not cause any large signal overload. The MME could ping the registration femtocells at small constant intervals of time to check their statics and then report to CN.

V. PERFORMANCE EVALUATION

A dynamic simulator developed by CWIND [13] group is used to evaluate the proposed scheme. The goal of the performance evaluation is to see if the proposed scheme mitigates the drawbacks of CGI and improves the inbound-handover quality in a realistic LTE Femtocell network.

A. Simulation Scenario

Since the femtocell is a free plug-and-play device, we have set the scenario based on the number of active (means switched on) femtoells during 24 hours (shown in Fig. 4) which is similar to the human life timetable. Moreover, there is not a

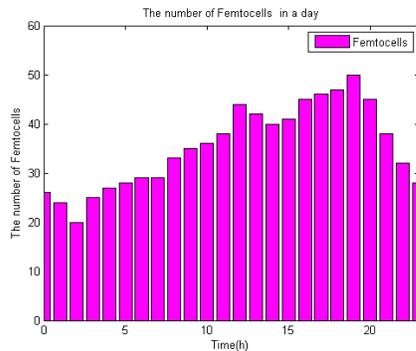


Fig. 4. The Various Number of Femtocells in 24 hours

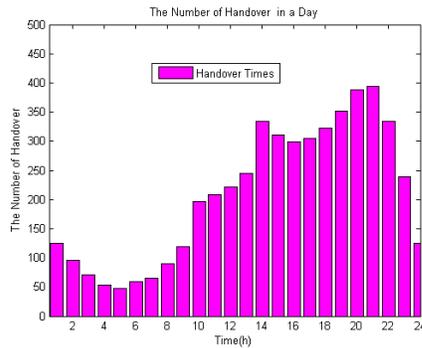


Fig. 5. The Various Number of Inbound-handover Events in 24 hours.

femtocell traffic model found in the literature, so, the inbound-handover events are also following the the same pattern, shown in Fig. 5.

Since this scheme is aiming to reduce the usage of CGI in the handover procedure, the number of CGI readings in an hour is chosen to evaluate the performance of this scheme. In the simulation, the UEs and femtocells will both randomly occur in the coverage of a macrocell (radius: 400 meters). If an UE occurs in the coverage of a femtocell, this UE would be determined as it is doing inbound-handover. Reflecting this, the MME would store the PCI and CGI which are from the serving femtocell and build the inbound-handover List for this network. The PCI updating interval time is initially set to 900 seconds, this time gap will be further analysed in future research. Some parameters are listed in Table II.

Moreover, two types of Femtocell access modes have been set in this simulation, CSG and non-CSG mode. The difference between the two modes is that when an UE has been detected as doing inbound-handover, the CSG femtocells will check if this UE is accessible or not; where as the non-CSG femtocells won't check it. Two distances of femtocell

coverage are simulated: 15 and 50 meters. The different values of coverage have impacted on amount of single-tier conflict between neighbouring femtocells, the higher the value of coverage distance, the greater the conflict. Public femtocells such as ones intalled in shopping malls belong to the non-CSG mode and have 50 meters coverage. Residential femtocells install in homes are mostly CSG mode which can serve up to 4 users and have 15 meters coverage, but some non-CSG residential femtocells are also simulated.

The evaluations are undertaken in two cases: Case 1 is a CSG mode only simulation, Case 2 simulates both CSG and non-CSG modes. In Case 1, the number of femtocells is up to 50. In Case 2 , the number of shopping Mall femtocell is 3, the number of non-CSG residential femtocells will increase from 10 to 27 and the number of CSG residential femtocells will increase from 10 to 20.

TABLE II  
PARAMETERS FOR SCHEME DG-PCI SCHEME

Parameter	Value
Coverage of macrocell/femtocell	400/15 meters
Coverage of shopping Mall/public place femtocell	50 meters
T	900 Seconds
Number of CSG fetocells (Case 1)	20-50
Number of shopping Mall/public place (case 2)	3
Number of non-CSG fetocells (Case 2)	10-27
Number of CSG fetocells (Case 2)	10-20

B. Simulation Analysis

As can be seen in Figs. 6a and 6b, across the board, in both cases, the proposed DG-PCI scheme consistently performs better compared to the conventional approach. We also observed that the performance differs according realistically to the time of the day. This is due to the fact that, when there are less inbound handovers, the boundary of determining the busy or not busy femtocell is unclear, which results in inaccurate judgement.

From Figs. 7a and 7b, our scheme obviously reduces the number of CGI reading compared to the conventional method. One can observe that the performance does not proportionally increase with the number of femtocells, but with the number of inbound handovers. This is perfectly understandable, because inbound-handover is what we are concerned with regarding to PCI allocation.

Comparing Case1 and Case 2, the performance in case 2 is better than case 1. First, this is due to the number of the UEs that a femtocell supports. In case 1, for home usage, again, the boundaries of busy or not busy femtocell are not clear. Reflecting this, the scheme is more suitable for open access which is currently high favoured femtocell deployment. Second, in case 2, we assigned fixed unique PCI to open access femtocells. Therefore, we suggest that in the reality, the network system should offer some independent unique PCIs for the openly accessed femtocells. We also suggest that unique PCIs could be considered as a commercial resource

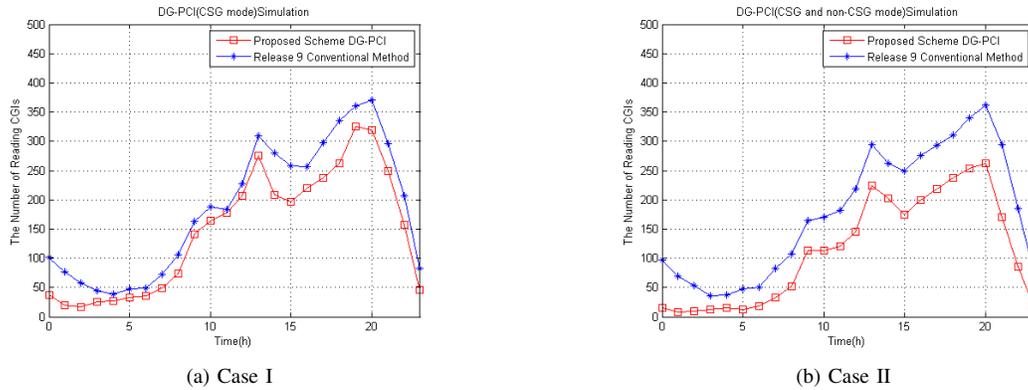


Fig. 6. DG-PCI and Rel. 9 simulation based on Time

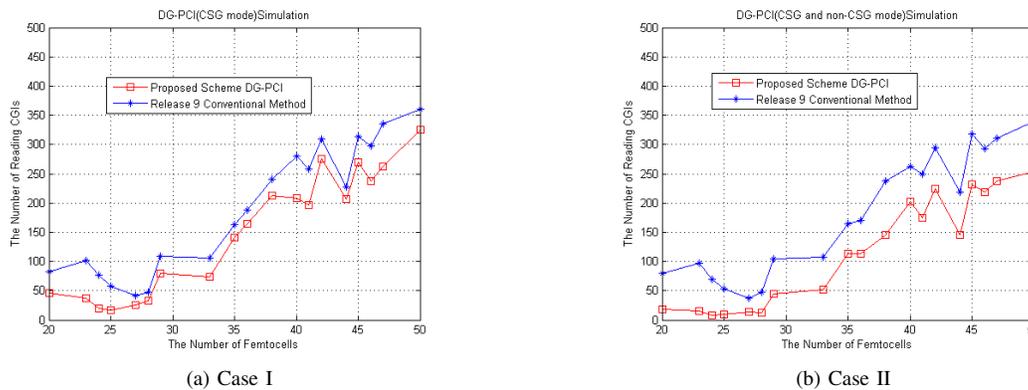


Fig. 7. DG-PCI and Rel. 9 simulation based on Number of Femtocells

which Shopping Malls would need to purchase in order to have better quality of wireless service for customers.

### VI. CONCLUSION AND FUTURE WORK

This paper proposed the DG-PCI scheme which based on the concept of the BFemtocell and dynamic PCI allocation and Graph Colouring-based mathematical method to mitigate the CGI problems and avoid cross-tier PCI confusion. The proposed scheme is tested by simulation. The results showed that the approach out-performs the existing approach in: reducing the number of CGI reading, hence reducing the possible of handover disconnection. The approach achieved dynamic PCI allocations in the cross-tier environment.

In terms of the future work, work will be carried out to optimise the PCI allocations, especially for released PCIs. Work will also be carried out on optimising the parameter T, which is the time interval representation of PCI updating, used in the scheme.

### REFERENCES

[1] Jie Zhang and Guillaume de la Roche, "Femtocells: Technologies and deployment", 1st ed, 2009.  
 [2] Mehdi Amirijoo, Pal Frenger, Fredrik Gunnarsson, Harald Kallin, Johan Moe and Kristina Zetterberg, "Neighbor Cell Relation List and Physical Cell Identity Self-Organization in LTE", IEEE International Conference on Communications (ICC), pp. 37-41, 2008.

[3] 3GPP TS 36.211 V8.6.0, "Physical Channels and Modulation", 2009.  
 [4] Assen Golaup, Mona Mustapha and Leo Boonchin Patanongpibul, "Femtocell Access Control Strategy in UMTS and LTE", IEEE Communications Magazine, Vol. 47, No. 9, pp. 117-123, 2009.  
 [5] 3GPP TS 36.300, "E-UTRAN Overall Description", 2009  
 [6] Poongup Lee, Jangkeun Jeong, Navrati Saxena and Jitae Shin, "Dynamic Reservation Scheme of Physical Cell Identity for 3GPP LTE Femtocell Systems", Journal of Information Processing Systems, pp. 207-220, 2009.  
 [7] David Lopez-Perez, Alvaro Valcarce, Akos Ladanyi, Guillaume de la Roche and Jie Zhang, "Intracell Handover for Interference and Handover Mitigation in OFDMA Two-Tier Macrocell-Femtocell Networks", EURASIP Journal of Wireless Communications and Networking, Vol. 2010, pp. 1-6, 18-21, 2010.  
 [8] Tobias Bandh, Georg Carl and Henning Sanneck, "Graph Coloring Based Physical-Cell-ID Assignment for LTE networks", ACM International Conference on Wireless Communications and Mobile Computing (IWCMC), pp. 116-120, 2009.  
 [9] Ting Wu, LanLan Rui, Ao Xiong and ShaoYong Guo, "An Automation PCI Allocation Method for eNodeB and Home eNodeB Cell", Wireless Communications Networking and Mobile Computing (WiCOM), pp. 1-4, 23-25, Sept. 2010.  
 [10] Yi Wu, Hai Jiang, Ye Wu and Dongmei Zhang, "Physical Cell Identity Self-Organization for Home eNodeB Deployment in LTE", Nokia Siemens Networks wicom-meeting, pp. 1-6, 2010.  
 [11] 3GPP TS 25.367, "Mobility procedures for Home Node B (HNB)", 2009.  
 [12] 3GPP TR 23.830, "Architecture aspects of Home NodeB and Home eNodeB (Release 9)", 2009.  
 [13] D. Lopez-Perez, A. Valcarce, G. De La Roche, E. Liu, and J. Zhang, "Access methods to wimax femtocells: A downlink system-level case study.", In IEEE International Conference on Communications Systems, pp. 1657-1662, 2008.

# Performance Analysis of Pilot Patterns in Coded OFDMA Systems over Multi-user Channels

Jihyung Kim, Jung-Hyun Kim, Kwang Jae Lim, and Dong Seung Kwon

*Internet Research Division*

*Electronics and Telecommunications Research Institute (ETRI)*

*Daejeon, KOREA*

*Email: savant21@etri.re.kr; jh.kim06@etri.re.kr*

**Abstract**—This paper presents a systematic approach for analyzing the bit error probability of a coded orthogonal frequency-division multiple access (OFDMA) system in multi-user channels. A comparative analysis with conventional pilot patterns, the overlapped pilot pattern and the interlaced pilot pattern, is developed. Simulation results show that the interlaced pilot pattern is more sensitive to the effect of interference, but also it outperforms the overlapped pilot pattern with serious interference scenarios.

**Keywords**—OFDMA, interference, pilot pattern.

## I. INTRODUCTION

Multi-user orthogonal frequency division multiple access (OFDMA) is a strong candidate transceiver scheme for high-speed packet-based multi-user wireless communication systems, due to its flexibility on subcarrier allocation and multi-user diversity utilization over multi-path fading channel [1]. A dynamic channel estimation is necessary before the demodulation of OFDMA signals since the radio channel is frequency selective and time-varying for wideband mobile communication systems [2]. The channel estimation can be performed by either inserting pilot tones into same positions or different positions for each user. In the rest of the paper, we call them as the overlapped pilot pattern and the interlaced pilot pattern respectively. After channel estimation is performed in the pilot locations, the channel frequency response is typically determined by interpolating the response between samples obtained using pilot tones. Thus, the design of pilot pattern is a critical factor in determining the performance of the OFDMA system. It is well-known that to achieve theoretically perfect channel estimation, pilots must be inserted into transmitted OFDMA symbols densely enough to fulfil the Shannon-Nyquist sampling theorem both in the frequency domain and in the time domain [3] [4]. Conversely, for a specific OFDMA configuration and pilot placement in the frequency domain, the channel propagation delay dispersion is limited to a specific range if theoretically perfect channel estimation is expected. For channels with larger delay dispersion, there will be irrecoverable channel estimation error, whose the magnitude is dependent on the distribution of power conveyed by the channel within and outside the acceptable propagation delay range [5] [6]. As

justified above, in designing modern OFDMA systems, it is crucial to be able to estimate the magnitude of such unavoidable errors in realistic channel conditions.

In this work our objective is to analyze the performance of the coded OFDMA system in multi-user channels. The analysis and results can be extended to the design of pilot patterns for mesh network [7].

The rest of the paper has been organized as follows. Section II contains the system model and pilot assisted channel estimation for the coded OFDMA in multi-user channels. Section III provides the bit error probability (BEP) analysis for the coded OFDMA system. Simulation results are in Section IV. Finally, Section V summarizes our main results.

## II. SYSTEM MODEL

In the discrete time domain, the  $n$ th sample of the  $l$ th output symbol of the OFDMA transmitter is given by the  $N$  point complex modulation sequence

$$x_l[n] = \sum_{k=0}^{N-1} X_l[k] e^{-j2\pi kn/N}, \quad (1)$$

where  $X_l[k]$  is a data signal and  $k$  is a subcarrier index.

The received OFDMA signal can be written as

$$\begin{aligned} Y_l[k] &= \sum_{n=0}^{N-1} \left\{ \sum_{m=0}^{\infty} h_l[m] x_l[n-m] + z_l[n] \right\} e^{-j2\pi kn/N} \\ &= \sum_{n=0}^{N-1} \left\{ \sum_{m=0}^{\infty} h_l[m] \left\{ \frac{1}{N} \sum_{i=0}^{N-1} X_l[i] e^{j2\pi kn/N} \right\} \right\} \\ &\quad \cdot e^{-j2\pi kn/N} + Z_l[k] \\ &= H_l[k] \cdot X_l[k] + Z_l[k], \end{aligned} \quad (2)$$

where

$$H_l[k] = \sum_{m=0}^{N_p-1} h_l[m] e^{-i2\pi k\tau_m/N}, \quad (3)$$

$$Z_l[k] = I_l[k] + N_l[k], \quad (4)$$

$H_l[k]$  is the channel response of the  $k$ th signal in the  $l$ th OFDMA symbol and  $Z_l[k]$  is the summation of interference

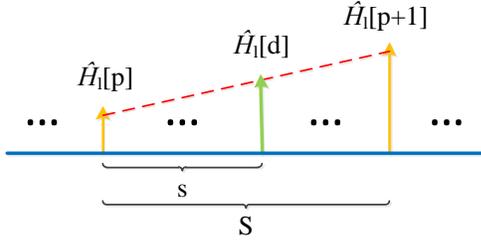


Figure 1. Linear interpolation.

$I_l[k] = H_l'[k] \cdot X_l'[k]$  according to other users using same resource and Gaussian noise  $N_l[k]$ .

### A. Channel Model

We consider the wide sense stationary uncorrelated scattering (WSSUS) Rayleigh fading channel. The channel impulse response is represented as

$$h(t) = \sum_{m=0}^{N_p-1} c_m(t) \delta(t - \tau_m), \quad (5)$$

where  $N_p$  is the number of multiple paths,  $\tau_m$  and  $c_m(t)$  are the time delay and the channel response of the  $m$ th multipath respectively.

### B. Pilot Assisted Channel Estimation

For the channel estimation, the known symbol, so-called pilot signal, is usually employed. Before the transmission, pilot signals are uniformly inserted into the data stream. Upon receiving the corrupted pilot signals at the receiver, the channel impulse response at pilot locations is estimated. The channel impulse response at data locations can then be obtained through interpolation with the pilot channel estimations. Here, we consider the least-squares (LS) estimate for the channel estimation at a pilot location. The  $l$ th estimated channel response can be obtained as follow:

$$\hat{H}_l[p] = \frac{Y_l[p]}{X_l[p]} = H_l[p] + \frac{Z_l[p]}{X_l[p]} = H_l[p] + V_l[p], \quad (6)$$

where  $p$  denotes the index of pilot subcarrier,  $X_l[p]$  is the  $p$ th pilot signal at the  $l$ th OFDMA symbol and  $Y_l[p]$  is the received symbol corresponding to pilot signal  $X_l[p]$ .

For the data location, the channel response can be estimated by taking interpolation between the pilot channel estimate. There are several forms to interpolate: uniform, spline interpolation, and 2D Wiener interpolation etc. Here, linear interpolation is used [4]. In the linear interpolation, the data channel estimate is given by

$$\hat{H}_l[d] = \left(1 - \frac{s}{S}\right) \hat{H}_l[p] + \frac{s}{S} \hat{H}_l[p+1], \quad (7)$$

where  $d$  denotes the index of data subcarrier,  $S$  is the interval between pilot subcarriers and  $s$  is the distance between the

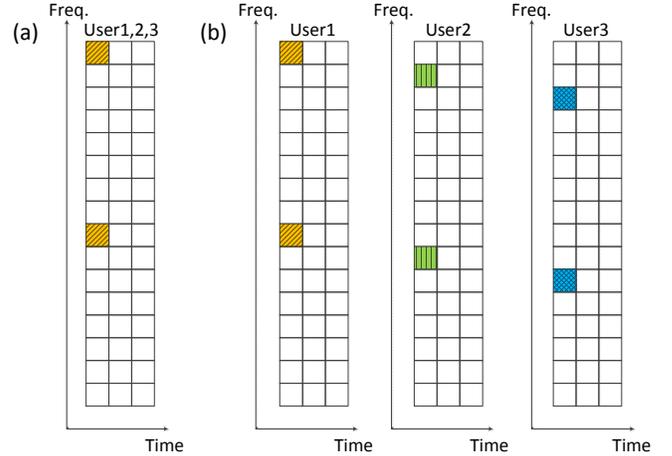


Figure 2. Pilot Patterns for OFDMA systems: (a) overlapped pilot pattern, (b) interlaced pilot pattern

$p$ th pilot subcarrier and the  $d$ th data subcarrier as shown in Figure 1.

Thus, the transmitted data signal at the  $d$ th data subcarrier in the  $l$ th OFDMA symbol can be estimated by

$$\hat{X}_l[d] = \frac{Y_l[d]}{\hat{H}_l[d]}. \quad (8)$$

For two patterns, which are the overlapped pilot pattern and the interlaced pilot pattern as shown Figure 2, the above formula can be re-written as follows:

1) Overlapped pilot pattern:

$$\hat{X}_l[d] = \frac{Y_l[d]}{\hat{H}_l[d]} = X_l[d] + \frac{Z_l[d]}{\hat{H}_l[d]} - \frac{V_l[d]X_l[d]}{\hat{H}_l[d]}, \quad (9)$$

$$\hat{Z}_l[d] = I_l[d] + N_l[d], \quad (10)$$

$$I_l[d] = \sum_{i \in \kappa} H_l^{(i)}[d] X_l^{(i)}[d], \quad (11)$$

where  $I_l[d]$  is the interference of the  $d$ th data signal in the  $l$ th OFDMA symbol,  $H_l^{(i)}[d]$  and  $X_l^{(i)}[d]$  are the channel response and the data signal of the  $i$ th user respectively and  $\kappa$  is the set of neighbor users.

2) Interlaced pilot pattern:

$$\hat{X}_l[d_1] = \frac{Y_l[d_1]}{\hat{H}_l[d_1]} = X_l[d_1] + \frac{Z_l[d_1]}{\hat{H}_l[d_1]} - \frac{V_l[d_1]X_l[d_1]}{\hat{H}_l[d_1]}, \quad (12)$$

$$\hat{Z}_l[d_1] = I_l[d_1] + N_l[d_1], \quad (13)$$

$$I_l[d_1] = \sum_{i \in \kappa, i \neq j} H_l^{(i)}[d_1] X_l^{(i)}[d_1] + H_l^{(j)}[p] X_l^{(j)}[p], \quad (14)$$

$$\hat{X}_l[d_2] = \frac{Y_l[d_2]}{\hat{H}_l[d_2]} = X_l[d_2] + \frac{Z_l[d_2]}{\hat{H}_l[d_2]} - \frac{V_l[d_2]X_l[d_2]}{\hat{H}_l[d_2]}, \quad (15)$$

$$\hat{Z}_l[d_2] = I_l[d_2] + N_l[d_2], \quad (16)$$

$$I_l[d_2] = \sum_{i \in \kappa} H_l^{(i)}[d_2] X_l^{(i)}[d_2], \quad (17)$$

where  $j$  is the user index that the  $j$ th user's pilot signal is the interference.

For the interlaced pilot pattern, we can divide the data signals into two parts. The first part is the region that one of neighbor user's pilot signal and another neighbor users' data signals are interference. The second part is the region that all of neighbor users' data signals are interference. The first part gets more interference than second part because the pilot power is boosted for accurate channel estimation. At the above equations, we denote  $d_1$  as data signal index in the first part and  $d_2$  as data signal index in the second part.

### C. Pilot Assisted Channel Estimation for Multi-user

For a single user case, it's obvious that more pilot signals leads to better performance but with sacrificing in symbol rate. Therefore the number of pilots is a trade-off between channel estimation accuracy and bandwidth efficiency. However, for multi-user case, it's not obvious that more pilot signals outperform less pilot signals due to interference between users. If pilot signals are corrupted, data fail to be demodulated irrespective of correcting processes such as despreading and decoding. Also, signal to interference ratio remains the same in the overlapped pilot pattern regardless of pilot power boost-up. The solution is to make the pilots from different users not to collide each other because pilot signals are relatively stronger than spread data signals. The channel performance estimation is improved by reducing the number of pilot signal collisions. However better channel performance estimation dose not always guarantee better data detection performance. The interference of pilot region and the interference of data region is trade-off.

## III. PERFORMANCE ANALYSIS OF CODED OFDMA SYSTEMS

### A. Coded OFDMA System

Error correcting coding is an essential part of an OFDMA system for mobile communication. OFDMA in a fading environment is almost always used with coding to improve its performance and as such is often referred to as Coded OFDMA or COFDMA. For an uncoded OFDMA system in a frequency selective Rayleigh-fading environment, each OFDMA subcarrier has a flat-fading channel. Accordingly, the average probability of error for an uncoded OFDMA system is the same as that for a flat-fading single-carrier system with the same average geometric mean of SNR.

Just as we can introduce time diversity through coding and interleaving in a flat-fading single-carrier system, we can introduce frequency diversity through coding and interleaving across subcarriers in an OFDMA system. However, since OFDMA in itself does not increase the system bandwidth it can never introduce frequency diversity on flat fading channels.

With coding and interleaving across subcarriers, the strong subcarriers help the weak ones as the block fading effect is decreased. Thus overall data detection performance is dependent on the ratio of strong part and weak part such as coding rate.

### B. Bit Error Probability Analysis with Simple Bound

To compare the performance of the overlapped pilot pattern and the interlaced pilot pattern, we consider the simple bound derived in [8]. Here, we use  $(n, k)$  block code that maps each  $k$  information bits into  $n$  codeword bits.

Then the simple bound on the bit-error rate is given by

$$P_b(E) = \sum_{h=h_{min}}^{n-k+1} \min \left\{ e^{-nE(c,h)}, e^{nr(\delta)} Q \left( \sqrt{2ch} \right) \right\}, \quad (18)$$

and the exponent of the simple bound is expressed by

$$E(c, h) = \begin{cases} \frac{1}{2} \ln[1 - 2c_0(\delta)] + \frac{cf(c, \delta)}{1 + f(c, \delta)}, \\ \quad \text{if } c_0(\delta) < c < \frac{e^{2r(\delta)} - 1}{2\delta(1 - \delta)}, \\ -r(\delta) + \delta c, \quad \text{otherwise,} \end{cases}$$

with  $\delta = \frac{h}{n}$ ,  $r(\delta) \triangleq \frac{\ln \sum_w [w/k] A_{w,h}}{n}$ ,  $c_0(\delta) = (1 - e^{-2r(\delta)}) \frac{1-\delta}{2\delta}$ ,  $c = \bar{\gamma}_b(k/n)$ , and  $f(c, \delta) = \sqrt{\frac{c}{c_0(\delta)} + 2c + c^2 - c - 1}$ .  $w$  is the input Hamming weight,  $h$  is the output Hamming weight, and  $A_{w,h}$  is the input-output weight distribution. For  $(n, k)$  random block code, the code spectrum  $A_{w,h}$  is known to follow a binomial distribution:

$$A_{w,h} = \frac{\binom{k}{w} \binom{n}{h}}{2^n}.$$

The average signal to interference noise ratio(SINR) per bit defined by

$$\bar{\gamma}_b = \frac{E[|HX|^2]}{KE[|Z|^2]}, \quad (19)$$

where  $K$  denotes the number of bits represented by one symbol.

For the overlapped and interlaced pilot patterns, the average SINR per bit defined by (19) becomes

1) Overlapped pilot pattern:

$$\bar{\gamma}_b = \frac{2\sigma_1^2}{2 \left( \sigma_N^2 + \sum_{i \in \kappa} \sigma_{D,i}^2 \right)}. \quad (20)$$

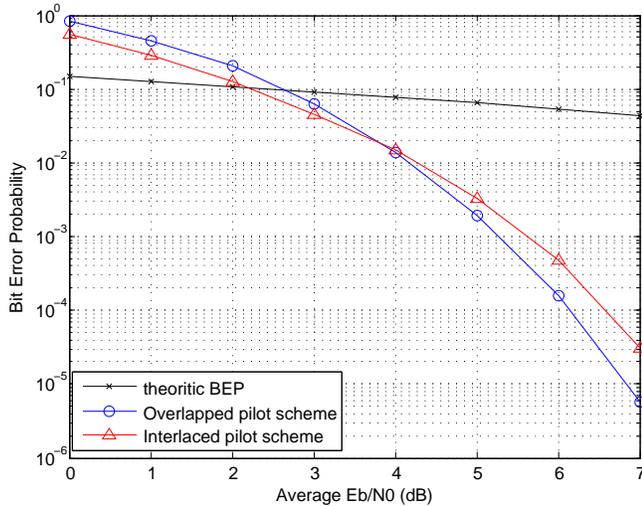


Figure 3. Pilot Patterns for 2 Users, pilot power boosting 4dB, pilot gap 20: (a) overlapped pilot pattern, (b) interlaced pilot pattern.

2) Interlaced pilot pattern:

$$\bar{\gamma}_{b,1} = \frac{2\sigma_1^2}{2\left(\sigma_N^2 + \sigma_{P,j}^2 + \sum_{i \in \kappa, i \neq j} \sigma_{D,i}^2\right)}, \quad (21)$$

$$\bar{\gamma}_{b,2} = \frac{2\sigma_1^2}{2\left(\sigma_N^2 + \sum_{i \in \kappa} \sigma_{D,i}^2\right)}, \quad (22)$$

$$P_b(E) = \frac{P_{b,1}(E) + \epsilon P_{b,2}(E)}{1 + \epsilon}, \quad (23)$$

where  $\epsilon = \epsilon_D/\epsilon_P$ ,  $\epsilon_D(\epsilon_P)$  is the number of symbols that affected by data(pilot) symbols of another users as an interference.

#### IV. SIMULATION RESULTS

For the simulation, we use (7,4) hamming code as an error correcting code. The channel estimation is performed by using LS estimator and the flat fading channel model is used.

Figures 3-7 give the BEP performance of two coded patterns using overlapped pilots and interlaced pilots for channel estimation, and uncoded BPSK in [3].  $E_b/N_0$  is the ratio of Energy per Bit ( $E_b$ ) to the Spectral Noise Density ( $N_0$ ). These results show that the BEP performance of the overlapped pilot pattern is relatively less susceptible to the number of users, but the BEP performance of the interlaced pilot pattern can be sensitive to the number of users, mainly due to the effect of interference.

The effect of pilot power boosting can be observed from Figure 3 and Figure 6. For the overlapped pilot pattern, the performance dose not improved as pilot power increase. The reason for this is that the pilot power of interference users

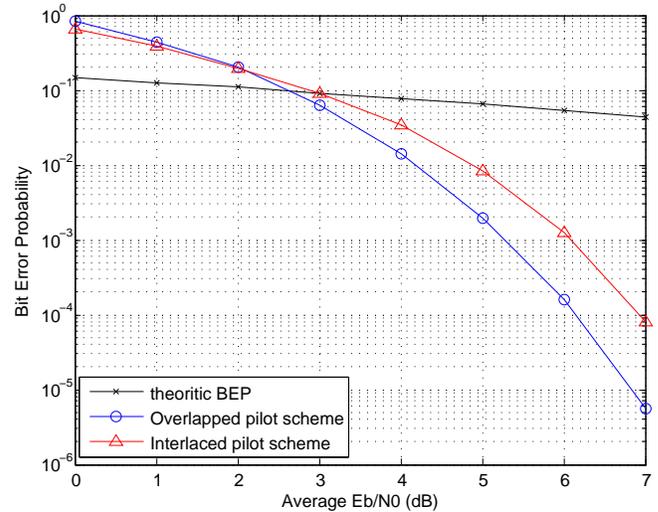


Figure 4. Pilot Patterns for 2 Users, pilot power boosting 4dB, pilot gap 8: (a) overlapped pilot pattern, (b) interlaced pilot pattern.

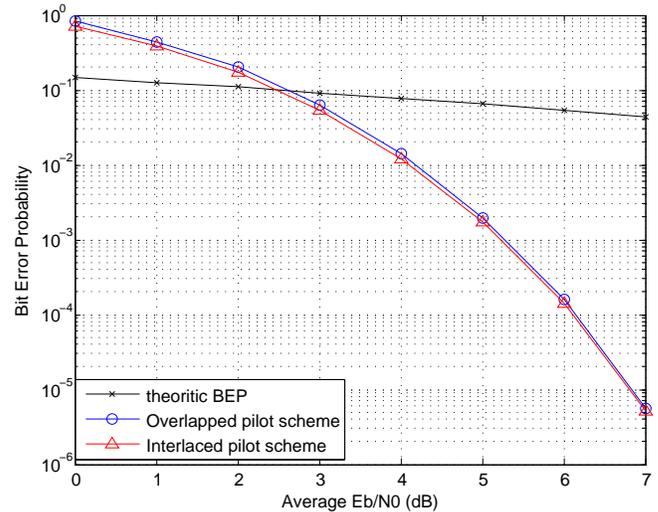


Figure 5. Pilot Patterns for 5 Users, pilot power boosting 4dB, pilot gap 8: (a) overlapped pilot pattern, (b) interlaced pilot pattern.

are also increased. Also, Figures show that the region whose the performance of the interlaced pilot pattern is better than that of the overlapped pilot pattern is increased as the pilot gap and the number of user are increased.

#### V. CONCLUSION

In this paper, we analyze the performance of the overlapped pilot pattern and the interlaced pilot pattern for channel estimation. This comparison is of special interest since the pilot pattern affects the performance of OFDMA systems. Simulation results in terms of BEP corroborate our theoretical analysis. We notice that the interlaced pilot pattern is more suitable for the multi-user networks like mesh network in which serious interference exists. We

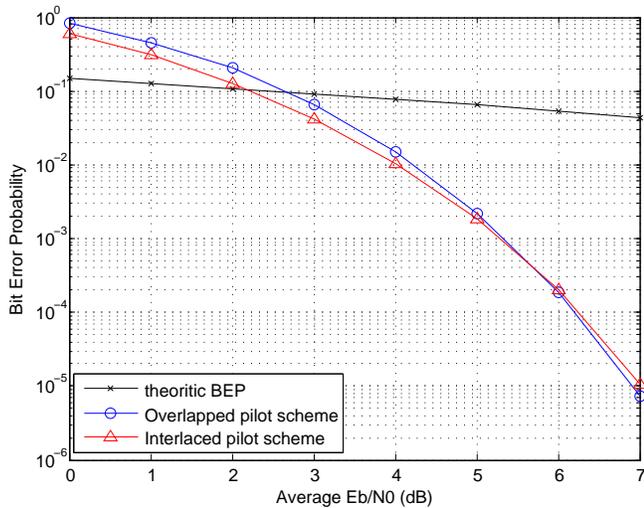


Figure 6. Pilot Patterns for 2 Users, pilot power boosting 3dB, pilot gap 20: (a) overlapped pilot pattern, (b) interlaced pilot pattern.

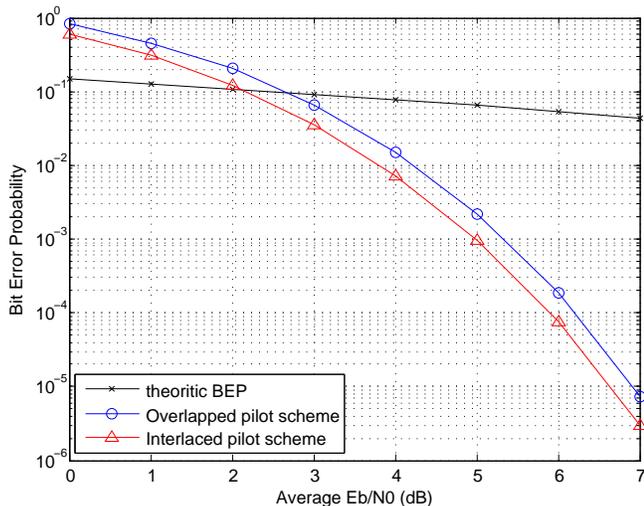


Figure 7. Pilot Patterns for 5 Users, pilot power boosting 3dB, pilot gap 20: (a) overlapped pilot pattern, (b) interlaced pilot pattern.

expect that the results obtained here can be directly applied to evaluate the performance of coded OFDMA systems in mesh network as well.

ACKNOWLEDGMENT

This research was supported by the KCC(Korea Communications Commission), Korea, under the R&D program supervised by the KCA(Korea Communications Agency)(KCA-2011-10913-04003).

REFERENCES

[1] R. van Nee and R. Prasad, *OFDM Wireless Multimedia Communications*, Norwood, MA: Artech House, 2000, ch. 9.

[2] O. Edfors, M. Sandell, and J. J. van de Beek *et al.*, "Robust channel estimation for by singular value decomposition," *IEEE Trans. on Comm.*, vol. 46, pp. 931-939, July 1998.

[3] J. G. Proakis, *Digital Communications* 3rd ed., Englewood Ciffs, NJ: Prentice-Hall, 1995.

[4] Jihyung Kim, Jeongho Park, and Daesik Hong, "Performance analysis of channel estimation in OFDM systems," *IEEE Signal Processing Letters*, Vol.12, No.1, pp. 60 - 62, Jan. 2005.

[5] Kun-Chien Hung and D.W. Lin, "Pilot-Aided Multicarrier Wireless Channel Estimation via MMSE Polynomial Interpolation," *IEEE GLOBECOM 2008*, pp. 1-5, Nov. 2008.

[6] Byoung Jun Lee, et al., "System-Level Performance of MIMO-based Mobile WiMAX System ," *2009 IEEE Mobile WiMAX Symposium*, pp. 189-194, July 2009.

[7] Lok Tiing Tie, et al., "Performance evaluation of Mesh-based mobile WiMax PHY baseband for above the sea propagation," *2010 European Wireless Conference (EW)*, pp. 83-87, Apr. 2010.

[8] D. Divsalar, "A simple tight bound on error probability of block codes with application to turbo codes," *TMO Progress Reports* 42-139, Jet Propulsion Lab, Nov. 1999.

# A System Design for The Integration of RFID Systems with Wireless Network Technologies

Mehmet Erkan Yüksel  
Department of Computer Engineering  
Istanbul University  
Istanbul, Turkey  
e-mail: eyuksel@istanbul.edu.tr

Asım Sinan Yüksel  
Department of Computer Science & Engineering  
Texas A&M University  
Texas, USA  
e-mail: asyuksel@cse.tamu.edu

**Abstract**— RFID (Radio Frequency Identification) is a special type of wireless communication that allows automatic identification of objects without requiring a line-of-sight. It cannot provide the fullest benefits as a standalone technology. Efficient and successful RFID applications must be developed by smoothly incorporating RFID systems into existing infrastructures to improve business processes and data management capabilities. In this paper, a system architecture that achieves the integration of RFID systems with wireless network technologies (e.g., ZigBee, Wi-Fi, WiMAX, cellular networks) is presented. A control unit is designed to identify objects, gather and manage RFID data, and keep the track of RFID tagged objects without any human involvement from long distances. A middleware is developed to improve mobile applications, provide monitoring our system, route RFID information to the related units automatically in a faster, more efficient, secure, wide capacitated way without any limitation. Our system design resolves the problems caused by lack of information, ensures to achieve competitive advantages, has longer operating range. In our study, a system design is illustrated to show how RFID systems can be integrated with wireless communication technologies to collect and manage secure real-time information.

**Keywords**- Radio frequency identification; active/passive tag; Wi-Fi; WiMAX; TinyOS; data gathering; object monitoring

## I. INTRODUCTION

RFID is an automatic identification (Auto-ID) and data capturing technology that gives an opportunity to identify and monitor objects by using a tag that carries information. It provides easy, secure and quick data entry, storage and transmission. It is used everywhere, such as shops, stores, hospitals, pharmaceutical companies, logistic services, etc. where real-time data should be used. It improves data gathering and management capabilities, raises security level of information about objects and resolves problems caused by lack of information. The main feature of RFID technology is its ability to identify, locate, track, and monitor objects automatically without a clear line of sight between tags and readers [1].

Wireless networks provide the infrastructure to support mobile connectivity to all information sources with high performance and availability. They allow users to communicate and access applications and information from

long distances without wires. Most wireless network technologies are license free and provide freedom of movement and the ability to extend applications to different locations. They provide significant cost savings to deploy. They offer device mobility and higher reliability because of increases in efficiencies, less error-prone cabling and less downtime as compared to a wired network. Wireless networks have several categories depending on user requirements and the size of the physical area that they are covering.

In this study, we present a system design that combines the RFID system and wireless networks. Main contribution of our study is to design and develop intelligent hardware and software components for the integration of RFID systems with wireless networks to gather and manage data, keep track of objects carrying RFID tags from short or long distances. Our design provides a cost-effective solution and economic power consumption. It also supports real-time detection of RFID tags and remote data collection through the underlying wireless network. Our system structure consists of spatially distributed devices called control units. These devices are equipped with wireless communication protocols that allow them to communicate with other devices either directly or via multiple hops. They can also use sensors to monitor conditions (e.g., temperature, pressure, motion, etc.) about objects. Having a control unit enables the network devices to identify, track and monitor a wider range of objects. Additionally, the wireless network can expand the RFID system's range and provides an RFID system in areas where a network infrastructure (e.g., Internet) does not exist.

The paper is organized as follows: Section II presents an overview of basic RFID system components. Section III presents related work discussing the integration of RFID systems with wireless communication technologies. In section IV we illustrate various examples of system designs that use our control unit and show how RFID systems are integrated with wireless networks. Finally, a discussion of new challenges and suggestions for future work are presented.

## II. OVERVIEW OF RFID SYSTEM COMPONENTS

Main hardware components of a basic RFID system are RFID tags (passive, active and semi-passive), RFID readers, and various antenna types.

An RFID tag consists of a microchip where the information about an object is stored, an antenna connected to the chip, on-board electronics, and a protecting film layer that covers these components. It is used as an electronic data carrier, and different information can be written and read in its environment. The microchip in RFID tag can store information from 64 bit to 8 MByte. This means that the tag can carry some important information, such as shipping history, order number, customer information, company/staff information and serial number. For the purpose of tracking the movement of objects, tags can be placed anywhere, such as containers, pallets, materials handling equipment, cases or even on individual products. Tags can be classified as passive (no battery), active (with battery) or semi-passive according to their power supply. While active tags use an energy source that is integrated to a tag physically, passive tags obtain this energy from the readers in the communication field. Today, semi-passive tags that have some properties of both active and passive tags can be also used [1-3].

An RFID reader is a specialized radio transmitter and receiver that act as a central location for the RFID system. It reads tag data by the help of the RFID antennas at a certain frequency. The RFID reader is basically an electronic device that emits and receives radio signals. It is designed for fast and easy system integration without losing performance, functionality and security. The reader has a real-time processor, an operating system, a memory, and a transmitter/receiver unit. The reader is usually classified into two types: (1) Fixed reader also called RFID gate is set to a definite place. (2) Mobile RFID reader includes a wireless interface, precisely Bluetooth, ZigBee or Wi-Fi. This device uses short or long-range radio links. It can identify, read/write, remotely control and monitor RFID tags over wireless communication. It contains some software tools to communicate with other mobile RFID readers, PDAs, laptops, etc. The Mobile RFID reader facilitates the identification of the tags that are in dangerous fields where the reading process is difficult [2], [3].

An RFID antenna is used as a medium for the purpose of tag reading and data collection. In many situations, the use of an antenna is important because tag-reading ranges are generally small. Although an antenna has very simple structure because of its concept, it must be able to receive the best signal in low power and adapt to special conditions. Antenna must be designed in different sizes, shapes and frequency intervals according to the properties and distances of the environment where the application will be implemented. The antenna can be designed considering several factors, such as reading distance, particular product types, specific operating conditions, known orientation, speed of the tagged objects, reader/controller, arbitrary orientation, antenna polarization, environmental changes, etc. [2], [3].

## III. RELATED WORK

In this section, we present several related works and projects that are taken place in industrial and academic research areas.

In a study performed by Intel Corporation [4], a Wireless Identification and Sensing Platform (WISP) has been developed. This platform consists of passive RFID tags and includes sensors that provide a very small-scale computing platform. It is a viable alternative system for smart dust applications, such as monitoring and recognizing human indoor activities, tracking items, informing of disasters, detecting poisonous gas or radioactivity, etc.

In [5], researchers from Intel Corporation have presented a method and its requirements, design and early experiences to obtain and use knowledge of human-object interactions. They have developed, built and deployed hands-on RFID readers that are embedded in wearable gloves and bracelets for detecting use of tagged objects. These autonomous readers are connected to sensors, gather data and report sensed events wirelessly to a base station. This method serves to support real-time analysis of data streams and can be applied to activity-based applications, such as health monitoring.

In [6], researchers in NESL (Networked and Embedded Systems Laboratory) in UCLA (University of California, Los Angeles) have presented a method that provides a cost-efficient solution for object recognition using the integration of RFID systems with mobile sensor networks. They have created an application called Ragobots (Real action gaming robots) that use small RFID readers. The Ragobots are wirelessly collaborated and coordinated to achieve a final goal while navigating in a terrain. They move in a random walk, search for tags, detect the tags and determine the object type based on the information stored on the tags.

In [7], Bluesoft, a company building Wi-Fi-based wireless security and location RFID technology, has deployed its real-time location system called AeroScout within Legoland Denmark, one of Europe's largest amusement parks. In this system, company's active tags operate using 802.11b wireless LAN technology and can be tracked over much larger areas with much fewer readers compared with traditional active RFID systems. In the Legoland deployment, readers consist of 38 Bluesoft AeroScout location receiver units. These units can read an AeroScout tag's 2.4 GHz signal and any other 802.11b-enabled device. Bluesoft's system's capability to use their preexisting Wi-Fi infrastructure allows AeroScout location receivers to be connected to the same cabling used for Wi-Fi access points. KidSpotter, a theme park application developer, has provided two different software programs. One of them links the AeroScout system with mobile phone networks while the other one enables the park to analyze tag location data to optimize the distribution of visitors within the park in real time.

In study [8], the authors have proposed a prototype that combines the RFID systems with wireless communication technologies for an in-home health care system to gather data and monitor the medication of patients. The prototype

includes two parts: an RFID reader node and a base station node used for wireless communication.

There are numerous other real-world applications related to the integration of RFID with wireless networks. Examples consist of US Navy's wireless RFID system to monitor the condition of valuable aircraft parts in storage [9], Siemens IT Solutions and Services' cargo-tracking system that combines RFID, wireless sensors, GSM and satellite services [10], ZigBeef's long-range RFID system that helps ranchers and rodeos track animals from various distances [11].

Most of the related works presented above are the applications in which RFID systems are used together with wireless sensor networks. However, our study handles this issue from a different point of view. The aim of our study is to demonstrate how active and passive RFID systems can be integrated to wireless networks that contain different communication technologies, such as ZigBEE, Wi-Fi, WiMAX, and cellular networks. In our implementation, several alternative systems that comprise different network topologies were designed to accomplish RFID integration. Moreover, a microcontroller based adaptable control unit that uses intelligent techniques to effectively operate in various integrations of RFID with wireless networks were developed.

#### IV. OUR SYSTEM ARCHITECTURE

Our system architecture offers an effective, a reliable, comprehensive, and low-cost solution to facilitate automatic detection and identification of objects from long distances. It comprises of one or more control units, a wireless network infrastructure, and a middleware. Fig. 1 illustrates our basic system architecture.

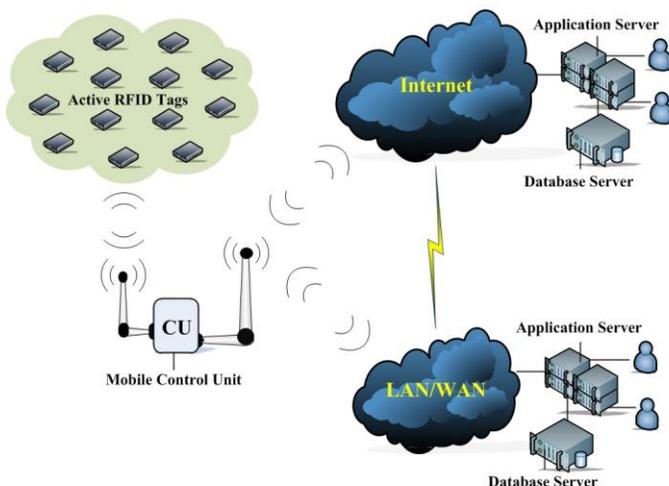


Figure 1. Our System Architecture

##### A. Our Control Unit

Our control unit is the brain of our system. It has long reading/writing ranges. It serves as a mobile mini application server between the RFID system and a wireless network. It gathers data from tagged objects, processes tag information and sends this information to related remote system that

consists of computers, servers, or base stations. It has more advanced specifications than current RFID readers. It runs an open-source operating system (TinyOS) [12] which is designed for low-power wireless devices. Our control unit has several properties:

- operating anti-collision protocols
- monitoring RFID system
- rerouting data about objects to related units on which database management system or application software works (if necessary)
- collaborating with wireless network devices and other control units
- remote controlling and management by users
- ensuring a safe data transfer between RFID tags and the back end system
- real-time data capturing and analyzing
- sensitive object tracking and localization

##### B. Wireless Network Infrastructure

The wireless network infrastructure can use different communication technologies, such as Bluetooth, ZigBee, Wi-Fi (IEEE 802.11 a/b/g/n), WiMAX (IEEE 802.16/x), and cellular networks (e.g., GSM, GPRS, EDGE, CDMA, etc.) to provide fast data communication, larger coverage area, and Internet access. This infrastructure can be a cloud that consists of various network topologies (or clouds) and many spatially distributed devices that produce convenient communication among themselves. Using a wireless network infrastructure will offer several important advantages for our system architecture as below:

- It will enable exchange of communication between clouds owned by multiple service providers and system resources.
- It will expand the RFID system's range and enable the RFID system in large areas where a network infrastructure (e.g., Internet) does not exist.
- It will enable both RFID and wireless network devices operate collaboratively and coordinately to provide the best services and applications.
- It will create a heterogeneous wireless environment to serve the purpose of different requirements, such as mobile automatic identification, human-object interaction, remote data collection, object monitoring and tracking, etc.
- It will ensure a safe and compressed data transfer, information sharing, real-time data management between our control units and the back end systems.

##### C. Our Middleware

For our system architecture, we developed a middleware that combines active RFID tags, our control unit, the wireless network and the back-end system. As shown in Fig. 2, our middleware has six modules as follows: data management, device management, process management, application development, communication, and administration. Each module does specific tasks on its own.

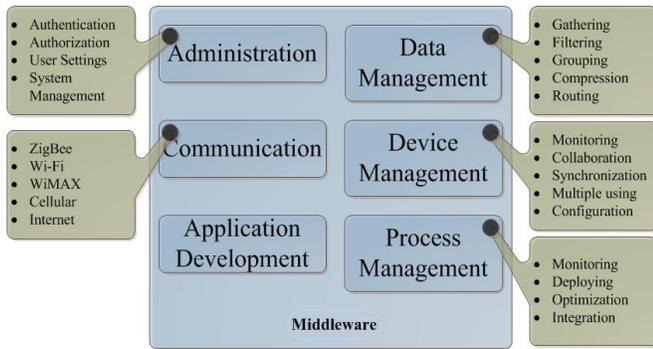


Figure 2. Our Middleware Architecture

Data management module has built-in rules and provides gathering, filtering, grouping, and routing of RFID data. Gathering helps us to gather data from RFID tags using our control unit. Filtering is used to extract definite data that we need. It checks data source when data contains unknown or undesired data. Grouping uses a clustering algorithm to classify data and allows splitting data into groups based on identical values in a field. Compression ensures a safe and compressed data transfer between the control unit and the back-end system. Routing allows to route information to related back-end system in a network.

Device management module is used to configure our control unit without having direct access to the device. It allows monitoring RFID system, collaboration with network devices and other control units, synchronization to manage data flow between control unit and mobile devices. It enables multiple users to share mobile devices and access RFID data.

Process management module offers monitoring, deploying, optimizing, and integration our system. It determines the resources required to operate at the back-end system and collect information for carrying out subsequent activities of deployment process. It consists of a set of services that allow interaction of multiple processes running on one or more RFID system.

Application development module involves creating a set of program modules to perform our future tasks. It supports an open-source programming model which mobile RFID applications can be developed on TinyOS operating system. It provides adding new functionalities in the form of loadable modules.

Communication module provides various wireless communication protocols to gather and manage data, monitor objects due to wireless network infrastructure. It also supports RFID anti-collision protocols for best Tag readings. It makes the platform independent of both the control unit and the wireless network. Several mobile devices such as Smart phones, Personal Digital Assistants (PDAs), laptops are supported by communication module.

Administration module defines a set of rules (read, write or change) for users to access specific data in every mobile application. Depending on the user rules and logon data, the end user can make changes in the following settings: number formats, language, time zone, control and monitor processes, manage user/group policies and applications, etc.

## V. IMPLEMENTATION

The objective of our study is accomplished by integrating hardware and software components. The hardware components are divided into two parts. The first part consists of active RFID tags, our control unit, and other RFID equipment that are shown on the left sides of Fig. 5 and 6. For the RFID equipment, we used several RFID evaluation kits [13-15] which includes the RFID reader, active and passive RFID tags, antennas, sensors, adaptors, etc. Second part is the wireless network that is shown on the right sides of Fig. 5 and 6. For creating a wireless network, we used a wireless router and network evaluation kit [16] which includes various wireless network devices. The Motorola development kit [17-21] that is shown at the center of Fig. 5 and 6 acts as a bridge to combine these two parts. The software part consists of a program module developed in NesC [22] using Eclipse with NESCDT plug-in [23] to control the Motorola development kit and our control unit, and a database management system to manage the RFID tag IDs of objects attached to them. We used TinyOS operating system in our control unit to achieve several operations such as sending commands, running anti-collision protocols, controlling data signals and communication between RFID system and wireless network. Our research has the potential of being adapted for use with secure real-time data gathering and management applications involving wireless network and RFID technologies.



Figure 3. Motorola G24 Developer Kit (exterior)

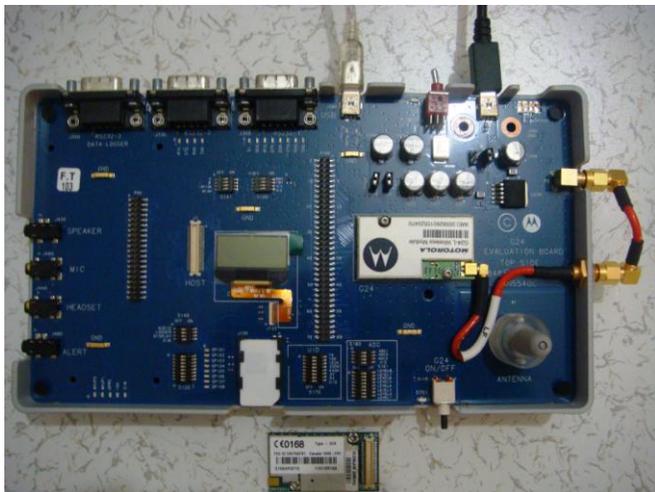


Figure 4. Motorola G24 Developer Kit (interior)

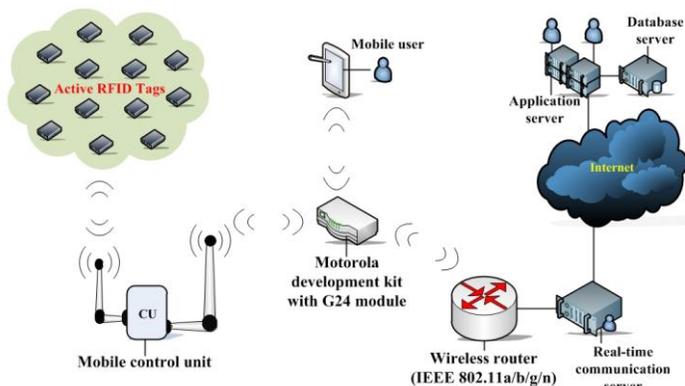


Figure 5. Integrating RFID Systems with Wi-Fi Networks

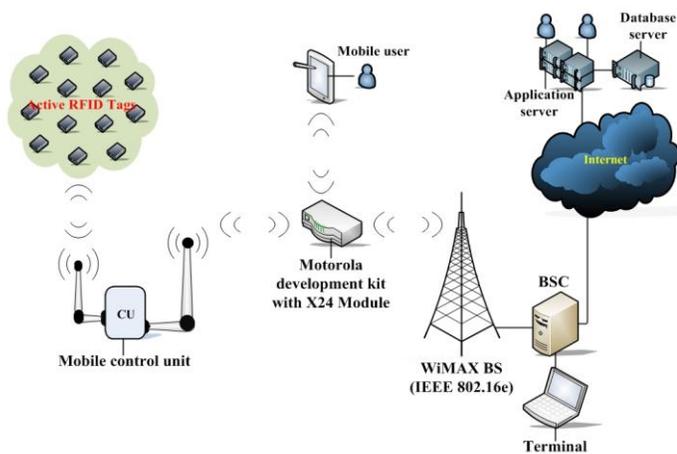


Figure 6. Integrating RFID Systems with Mobile WiMAX

## VI. RESULTS

RFID systems are the best choices if the area that will be monitored is small. If the requirement is to monitor a small number of locations, these systems can be instrumented with RFID readers at a low cost. In case of long ranges, it is

essential to have high coverage as possible. An important disadvantage of RFID readers is that these devices do not have a communication network to enable exchange of information. Therefore using wireless communication technologies will be good choices in terms of infrastructure costs since costs associated with Wi-Fi are minimal.

Because Wi-Fi is widely available, more standardized and used across a broad variety of mobile devices, we chose to integrate it with our system. During the integration of Wi-Fi to our systems, we had challenges. These challenges were related to the operating system that we are using in our control unit.

As we discussed in previous section, we used TinyOS as our base operating system. Although TinyOS has extensive networking support, it does not support Wi-Fi technology. To overcome this challenge, we used TinyWifi [24] as an extension of TinyOS [12]. In TinyWifi, applications from highly resource constrained sensor networks can easily be compiled for resource rich Wi-Fi based networks [24]. This solution opens the rich protocol repository of TinyOS to the researchers who are studying in wireless communication area.

As a result, TinyWifi enabled us to develop necessary TinyOS applications and protocols and execute them directly on Linux by compiling for the TinyWifi platform. Using TinyWifi as a TinyOS replacement, we successfully evaluated the wireless protocols that are originally designed for sensor networks. Our evaluation was based on the test applications of TinyOS that demonstrate the functioning of radio communication, serial messaging. By using TinyWifi, we were able to receive demo sensor measurements, display them and build routing trees, which are the variations of examples presented in [25].

In our future work, we want to evaluate the Zigbee protocol stack in our system once Zigbee protocol is fully supported by TinyOS. Zigbee protocol stack for TinyOS is currently under development and [26] provides open source tools for IEEE 802.15.4 and ZigBee.

## VII. CONCLUSION

Combining RFID systems and wireless communication technologies is a promising solution that can increase the productivity and give a competitive advantage to those that begin to use it first. However, there will be some standardization issues. These issues need to be solved before the technology sees wide acceptance.

Even though there are issues with combining RFID and wireless technologies, we successfully evaluated the wireless protocols that are originally designed for sensors in our system. In our study, we proposed a system architecture that is designed to provide RFID users flexibility of wireless networks at lowest cost. Additionally, this paper has discussed the ways to combine RFID systems with wireless communication technologies and its advantages such as cost reduction. As a future goal, we want to evaluate the integration of Zigbee technology with our system, when it is fully supported by the operating system we are using.

## REFERENCES

- [1] K. Finkenzeller, "RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication", 3rd Edition, John Wiley & Sons Ltd., West Sussex, UK, 2010, ISBN:978-0-470-69506-7.
- [2] M. E. Yüksel, "Integration Of The RFID Systems Over Wireless Communication Technologies", Master of Science Thesis, Department Of Computer Engineering, Istanbul University, Istanbul, Turkey, Feb. 2010, pp. 4-30.
- [3] M. E. Yüksel and A. S. Yüksel, "RFID Technology in Business Systems and Supply Chain Management", Journal of Economic and Social Studies, Vol.1, No. 1, 2011, pp. 53-71.
- [4] <http://www.seattle.intel-research.net/wisp/>, Last accessed: April 28, 2011.
- [5] K. P. Fishkin, M. Philipose, and A. Rea, "Hands-On RFID: Wireless Wearables for Detecting Use of Objects", The 9th IEEE International Symposium on Wearable Computers (ISWC'05), Osaka, Japan, Oct. 2005, pp. 38-43.
- [6] A. Kansal, J. Friedman, P. L. Aghera, B. Vasu, A. Dixit, D. C Lee, M. B. Srivastava, W. J. Kaiser, D. Estrin, and G. Sukhatme, "CENS Tech Report: Real Action Gaming Robots (Ragobots)", <http://nesl.ee.ucla.edu/document/show/32>, Last accessed: April 28, 2011.
- [7] J. Collins, "Lost and Found in Legoland", RFID Journal, April 28, 2004, <http://www.rfidjournal.com/article/view/921>, Last accessed: April 28, 2011.
- [8] L. Ho, M. Moh, Z. Walker, T. Hamada, and C.-F. Su, "A Prototype on RFID and Sensor Networks for Elder Healthcare: Progress Report", Proceedings of the ACM SIGCOMM Workshops, PA, USA, August 2005, pp. 70-75.
- [9] M. Roberti, "Navy Revs Up RFID sensors", RFID Journal, June, 2004, <http://www.rfidjournal.com/article/articleview/990> Last accessed: April 28, 2011.
- [10] R. Wessel, "Cargo-Tracking System Combines RFID, Sensors, GSM and Satellite", RFID Journal, January 25, 2008, <http://www.rfidjournal.com/article/articleview/3870>, Last accessed: April 28, 2011.
- [11] C. Swedberg, "ZigBeef Offers Ranchers a Long-Distance Cattle Head Count", RFID Journal, February 21, 2008, <http://www.rfidjournal.com/article/articleview/3935>, Last accessed: April 28, 2011.
- [12] <http://www.tinyos.net/>, Last accessed: April 28, 2011.
- [13] [http://www.atlasrfidstore.com/Alien\\_9900\\_Developer\\_Kit\\_p/alr-9900-devc.htm](http://www.atlasrfidstore.com/Alien_9900_Developer_Kit_p/alr-9900-devc.htm), Last accessed: April 28, 2011.
- [14] <http://www.impinj.com/products/SubOneCol.aspx?id=5002&erms=kit>, Last accessed: April 28, 2011.
- [15] [http://www.ccsinfo.com/product\\_info.php?products\\_id=RFID\\_kit](http://www.ccsinfo.com/product_info.php?products_id=RFID_kit), Last accessed: April 28, 2011.
- [16] <http://www.synapse-wireless.com/?mainID=3&subID=6&type=product&prodID=6>, Last accessed: April 28, 2011.
- [17] [http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+JAVA\\_US-EN](http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+JAVA_US-EN), Last accessed: April 28, 2011.
- [18] [http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+Lite\\_US-EN](http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+Lite_US-EN), Last accessed: April 28, 2011.
- [19] [http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+Edge\\_US-EN](http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+Edge_US-EN), Last accessed: April 28, 2011.
- [20] [http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+JAVA\\_US-EN](http://www.motorola.com/Business/US-EN/Business+Product+and+Services/M2M+Wireless+Modules/G24+JAVA_US-EN), Last accessed: April 28, 2011.
- [21] <http://developer.motorola.com/docstools/sdks/>, Last accessed: April 28, 2011.
- [22] <http://nescc.sourceforge.net/>, Last accessed: April 28, 2011.
- [23] [http://docs.tinyos.net/index.php/NESCDT-An\\_editor\\_for\\_nesC\\_in\\_Eclipse](http://docs.tinyos.net/index.php/NESCDT-An_editor_for_nesC_in_Eclipse), Last accessed: April 28, 2011.
- [24] <http://www.comsys.rwth-aachen.de/research/projects/tinywifi/> Last accessed: April 28, 2011.
- [25] M. H. Alizai, B. Kirchen, J. A. B. Link, H. Wirtz, and K. Wehrle, "TinyOS Meets Wireless Mesh Networks", The 8th ACM Conference on Embedded Networked Sensor Systems (SenSys 2010), Zurich, Switzerland, Nov. 2010, pp. 429-430.
- [26] <http://open-zb.net/index.php>, Last accessed: April 28, 2011.

## The Selected Propagation Models Analysis of Usefulness in Container Terminal Environment

Slawomir J. Ambroziak, Ryszard J. Katulski

Department of Radiocommunication Systems and Networks  
Gdansk University of Technology, Faculty of Electronics, Telecommunications and Informatics  
Narutowicza Street 11/12, 80-233 Gdansk, Poland  
sj\_ambroziak@eti.pg.gda.pl; rjkat@eti.pg.gda.pl

**Abstract**— It is particularly important to determine which propagation model is the most suitable for designing mobile radio systems in container terminal environment. The selected propagation models have been investigated. Results of the models' usefulness verification in terms of signal loss determination in such environment have been analyzed and discussed. The applied research methodology has been described too. The analysis have proved mismatching of these models to experimental data.

**Keywords:** radio waves propagation; measuring research; container terminal environment; propagation models verification

### I. INTRODUCTION

Container port area should be treated as a very difficult radio waves propagation environment, because lots of containers made of steel are causing very strong multipath effect and there is time-varying container arrangement in stacks of different height. There are a number of propagation models, mainly for urban, suburban or rural environments [1, 2]. There is also propagation model destined for container port environment, but this model has been developed for designing only fixed radio links [3, 4]. Modeling of basic transmission loss in mobile radio links is more complicated, so it is particularly important to determine which propagation model is the most suitable for designing such links. This issue is very important and there is a lack of an analysis in this field. To solve this problem there is a need to verify existing models based on results of measuring research. Such tests have been carried out by authors in Deepwater Container Terminal Gdansk SA. Nearly 290 thousand data of propagation cases were collected according to normative requirements [5, 6], which have to be met during the research. The analysis contained in [7] has been taken into account too.

At the outset of the paper (section II) the applied research methodology have been presented. This part describes both the measuring equipment and procedures.

Next, in section III, the selected propagation models have been shortly characterized. These models are: ITU-R P.1411 models for NLoS1 situations (in cases of propagation over roof-tops for urban and suburban areas), COST231 – Walfisch-Ikegami model and the above mentioned multi-variant empirical model for designing fixed radio links in container terminal for LOS (line-of-sight) and NLOS (non-line-of-sight) situations [3, 4].

The main part of the paper (section IV) presents results of verification of the models' usefulness in terms of designing the mobile radio networks in container terminal environment. This verification is based on mean error and standard error of estimate, which are commonly being used to verify accuracy of the path loss models.

At the end of the paper, in section V, the results have been summarized and discussed. Additionally, authors shortly present future research aimed at developing new propagation model for designing mobile radio links in container terminal environment.

### II. APPLIED RESEARCH METHODOLOGY

The propagation research have been carried out in the years 2008-2009 in DCT Gdansk. The structure and power description of the measuring radio link have been presented in [8]. This link was built with fixed transmitting section, mobile receiving section and the propagation environment, which was the subject of research.

As is known, basic transmission loss  $L_b$  of this environment may be expressed using following formula [8]:

$$L_b [dB] = P_t [dBW] + G_t [dBi] - P_{MR} [dBW] - F_c [dB], \quad (1)$$

on the basis of the power gain  $G_t$  of the transmitting antenna, the power  $P_t$  on input of the transmitting antenna – set during calibration process of the transmitting section, the power  $P_{MR}$  on input of measuring receiver and the correction factor  $F_c$  – calculated during calibration process of the receiving section. It may be expressed by:

$$F_c [dB] = L_{rc} [dB] - G_r [dBi], \quad (2)$$

where  $L_{rc}$  means losses in the receiving section feeder lines and  $G_r$  is the power gain of receiving antenna.

The fixed transmitting section (Fig.1) of the test equipment consisted of signal generator connected to transmitting antenna through the RF amplifier.

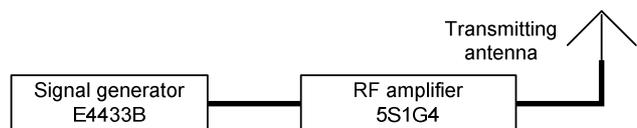


Figure 1. The block diagram of the fixed transmitting section

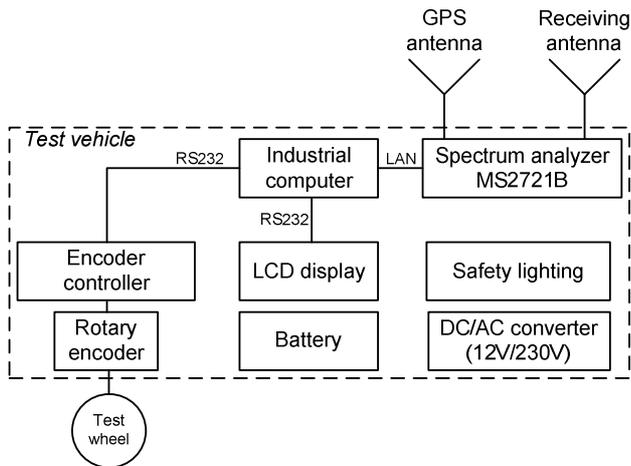


Figure 2. The block diagram of the mobile receiving section

The transmitting antenna was a monopole vertical antenna with electrical length of one-quarter of a wavelength. It has been developed and implemented in a manner, that allows to change its linear length, so it may be used to research on various frequencies.

The mobile receiving section (Fig. 2) consisted of a spectrum analyzer (with built-in GPS receiver), an industrial computer, a rotary encoder with its controller and a test wheel, a LCD display, a safety lighting and a battery with a DC/AC converter. The receiving antenna was the same type as the transmitting antenna. Whole receiving section has been carried by test vehicle (hand-cart).

The measurement results should include information about slow and fast changes of the power flux density of electromagnetic field (slow and fast fading, respectively) [7]. For obtaining 1 dB confidence interval around the real mean value, the test points have been chosen at each  $0.8 \lambda$  (wavelength), over  $40 \lambda$  averaging interval [6].

During the research in the DCT Gdansk nearly 290 thousand data of propagation cases have been collected. These cases concern propagation routes with various lengths, various frequencies of test signal and various heights of transmitting antenna installation.

### III. CHARACTERISTICS OF SELECTED PROPAGATION MODELS

The container terminal is a non-typical radio wave propagation environment. Due to its structure, consisting of containers' stacks placed on a flat surface and cut by a uniform grid of routes, it seems to be similar to urban areas [4]. However, fact that the containers are made of corrugated steel is the reason to suppose that the conditions of radio waves propagation in such environment might be quite different. It is also important that both the layout of containers' stacks, as well as their height are variable in time. After considering above mentioned issues, four well-known propagation models have been selected, namely:

- Walfisch-Ikegami for NLOS situations [1],
- ITU-R P.1411 for NLoS1 situations (propagation over roof-tops for urban and suburban areas) [2],

- empirical model for fixed radio links in container terminal (for LOS1 and NLOS1 situations) [3, 4].

These models are going to be evaluated in terms of their usefulness for designing of mobile radio links in container terminal environment.

#### A. The COST231 Walfisch-Ikegami model

The COST231 Walfisch-Ikegami model allows for good path loss estimation by consideration of a number of parameters to describe the character of the urban environment, namely: average height of buildings, widths of roads, building separation and road orientation with respect to the direct radio path. Obviously, the model also takes into account such parameters as propagation path length and signal frequency. The model distinguishes between LOS and NLOS situations. The second one was selected to be evaluated. In this case the basic transmission loss is composed of free space loss, multiple screen diffraction loss and roof-top-to-street diffraction and scatter loss. Formulas used to calculate basic transmission loss are explained in detail in [1].

#### B. The ITU-R P.1411 model

Recommendation [2] includes propagation models destined for designing short-range outdoor radio systems for different types of environments. There have been selected two models for typical cases (NLoS1), where base station antenna is mounted above roof-top level.

The first one is the model described in section 4.2.1 of Rec. [2]. This model should be used for estimating the basic transmission loss in a highly urbanized city centers, medium-sized cities and suburban areas, where the roof-tops are all about the same height. It is a modified and extended version of the Walfisch Ikegami model. In addition, this model describes situations where the length of path covered by buildings is less than the so called "settled field distance". This situation hasn't been taken under consideration in Walfisch-Ikegami model. Mathematical formulas describing this model have been omitted due to their high complexity.

The second model has been characterized in section 4.2.2 of Rec. [2]. It may be used to calculate the basic transmission loss in suburban environment. Depending on the distance between base station and mobile station this model distinguishes three regions in terms of the dominant arrival waves at the mobile station, namely:

- direct wave dominant region, when the distance between antennas of the wireless link is very short,
- reflected wave dominant region, when the separation between base station and mobile station is relatively short,
- diffracted wave dominant region, when antennas separation is long [2].

Mathematical formulas describing the basic transmission loss have been omitted because of their complexity.

#### C. The model for fixed radio links in container terminal

In context of this paper, particularly noteworthy is empirical model for designing fixed radio links in container terminal. It was developed upon the results of almost five

thousand of propagation path measurements in real container terminal environment. This model makes the basic transmission loss dependent on the following parameters: frequency  $f$ , propagation path length  $d$ , path type qualification: line of sight or non-line of sight condition, difference between transmitter antenna height  $h_T$  above terrain level and average height  $h_{av}$  of container stack, but two possible cases are investigated separately:  $h_T \geq h_{av}$  and  $h_{av} > h_T$  [3, 4].

From among four model variants, two describes the propagation situations that have occurred during the tests in the DCT Gdansk SA, namely:

- LOS1, for  $h_T \geq h_{av}$ :

$$L_{LOS1} [dB] = 55.2 + 20 \log f [MHz] + 5.8 \log d [km] - 22.1 \log (h_T [m] - h_{av} [m]), \quad (3)$$

- NLOS1, for  $h_T \geq h_{av}$ :

$$L_{NLOS1} [dB] = 32.6 + 20 \log f [MHz] + 7.9 \log d [km] + 0.8 \log (h_T [m] - h_{av} [m]), \quad (4)$$

This model is valid for frequency range from 500 MHz up to 4 GHz. It should be highlighted that it is destined to estimate the basic transmission loss of fixed links.

#### IV. STATISTICAL EVALUATION OF SELECTED MODEL

Verification of selected models' usefulness in terms of designing the mobile radio networks in container terminal environment is based on two measures of matching experimental data to mathematical models, namely: mean error (ME) and standard error of estimate (SEE). These errors are commonly being used to verify accuracy of the path loss models and they are defined by following expressions [4]:

$$ME [dB] = \frac{1}{N} \sum_{i=1}^N (L_{m,i} [dB] - L_{c,i} [dB]), \quad (5)$$

$$SEE [dB] = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (L_{m,i} [dB] - L_{c,i} [dB])^2}, \quad (6)$$

where  $L_{m,i}$  is the value of measured basic transmission loss in  $i$ -th position of receiver equipment ( $i=1, \dots, N$ ),  $L_{c,i}$  means basic transmission loss value computed using propagation model for  $i$ -th position, and  $N$  is the sample size. Mean error value reflects the expected average difference between path loss values obtained using proposed model and real path loss measurement results, while standard error of estimate reflects dispersion of measured path loss values and describes how the propagation model matches to experimental data [3].

Table 1 summarizes values of mean error and standard error of estimate for selected propagation models. It may be seen that the smallest error values have been obtained for the COST231 Walfisch-Ikegami model (for medium sized city and suburban areas and for data from the range of applicability) and for the model for fixed radio links in container terminal (for NLOS1 scenario). In the first case mean error reached -2.18 dB, which means that this model overestimates basic transmission loss in relation to real values. For the second model, obtained mean error is positive and equals 3 dB, which means underestimation of propagation loss. In both cases SEE exceeds the value of 7 dB. Although the maximum acceptable standard deviation is 8 dB [9], it is expected that new propagation model for analyzed environment will allow to obtain much smaller error values. It also should be noted, that the SEE for Walfisch-Ikegami model, obtained for all data (from a range of 0.5÷4 GHz) is greater than acceptable value.

On the other hand the least matched to experimental data is the ITU-R P.1411 model for NLoS1 scenario (§4.2.1), designed to calculate path loss in a highly urbanized city centers. Mean error with the value of -9.36 dB and the standard error of estimate at the level of 14.3 dB make this model unsuitable to calculate basic transmission loss for mobile links in the container terminal environment.

TABLE I. VALUES OF MEAN ERROR AND STANDARD ERROR OF ESTIMATE FOR SELECTED PROPAGATION MODELS

Model	Scenario	Range of measurement data	Sample size	ME [dB]	SEE [dB]
COST 231 Walfisch-Ikegami	Medium sized city and suburban areas	All data	287582	-5.31	10.60
		Range of applicability	130968	-2.18	7.90
	Metropolitan centres	All data	287582	-10.12	15.96
		Range of applicability	130968	-3.93	9.26
ITU-R P.1411 NLoS1 situation (§4.2.1 of Rec. [2])	Medium sized city and suburban centres	All data	287582	-8.36	13.41
		Range of applicability	254184	-8.88	13.74
	Metropolitan centres	All data	287582	-8.77	13.90
		Range of applicability	254184	-9.36	14.30
ITU-R P.1411 NLoS1 situation (§4.2.2 of Rec. [2])	Suburban areas	All data	287582	-4.48	9.98
		Range of applicability	190581	-5.92	10.83
Model for fixed radio links in container terminal	LOS1	Range of applicability	287582	3.80	8.30
	NLOS1			3.00	7.57

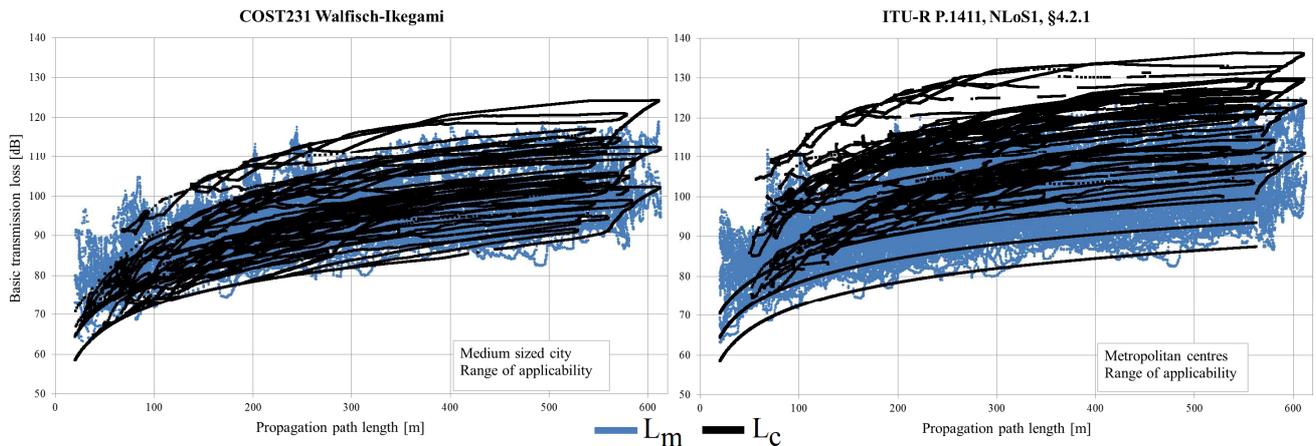


Figure 3. Basic transmission loss calculated on the basis of COST231 Walfisch-Ikegami model and ITU-R P.1411 model

In Fig. 3, the basic transmission loss graphs (including all propagation cases) for the Walfisch-Ikegami model and for the ITU R P.1411 model are presented. They have been drawn on the background of the measurement data for better illustration their matching to experimental data. The graph on the right is for the ITU-R model for NLoS1 scenario in metropolitan centers. This model is described in section 4.2.1. of Rec. [2]. As is seen the calculated values don't match well to experimental data. The graph on the left presents the results calculated on the basis of Walfisch-Ikegami model for medium sized city and suburban areas. It is the best fitted model, assuming that the parameters are from the range of applicability.

The results mentioned above allow to draw the conclusion that the propagation conditions occurring in the container terminal are different from the highly urbanized environments. They are more similar to suburban areas.

#### V. CONCLUSION

The analysis of the usefulness of selected propagation models to designing the mobile radio systems in the container terminal has been presented. This analysis has been done on the basis of the evaluation of selected propagation models in terms of their fit to data obtained during the tests. The research have been carried out in accordance with the recommendations [5, 6] and with taking into account the analysis contained in [7]. There are large differences in the results obtained for different propagation models. In addition the analysis have proved mismatching of these models to experimental data. Therefore, there is a need to increase accuracy of basic transmission loss estimation for mobile links in the container terminal environment. It may be done by modifying existing models or by developing new propagation model, taking into account additional independent variables, specific for the container terminal. Both methods are the goal of future research.

#### ACKNOWLEDGMENT

Described research have been funded by the Polish Ministry of Science and Higher Education, as a part of research and development project No R02 012 01. The authors express their sincere thanks for allocated funds for this research.

The authors also express thanks to the administration and the employees of the container terminal DCT Gdansk SA for allowing the measurements on the container terminal area.

#### REFERENCES

- [1] D.J. Cichon and T. Kürner: "Propagation Prediction Models", Digital Mobile Radio Towards Future Generation Systems, COST 231 Final Report
- [2] ITU-R, P.1411-4: "Propagation data and prediction methods for the planning of short-range outdoor radiocommunication systems and radio local area networks in the frequency range 300 MHz to 100 GHz", 2007
- [3] R.J. Katulski, J. Sadowski, and J. Stefanski: „Propagation Path Modeling In Container Terminal Environment”, Proceedings of VTC 2008-Fall: IEEE 68th Vehicular Technology Conference, Calgary 2008, Canada, pp. 1-4
- [4] R.J. Katulski: "The radio wave propagation in wireless telecommunication" (in Polish), Transport and Communication Publishers, Warsaw, 2009
- [5] ITU-R, SM.378-7: "Field-strength measurements at monitoring stations", 2007
- [6] ITU-R, SM.1708: "Field-strength measurements along a rout with geographical coordinate registration", 2005
- [7] W.C.Y. Lee: "Mobile communications design fundamentals", Wiley, 1993
- [8] S.J. Ambroziak: "Measuring research on radio wave propagation – normative requirements" (in Polish), Telecommunication review and telecommunication news, No. 2-3/2010, pp.84-89
- [9] E.O. Rozal and E.G.Pelaes: "Statistical Adjustment of Walfisch-Ikegami Model based in Urban Propagation Measurements", IMOC 2007, pp. 584-588

# Bandwidth Allocation Algorithms for VOIP Networks: An Experimentation System and Evaluation of Created Algorithm

Rafal Orłowski, Michal Chamow, Iwona Pozniak-Koszalka, Leszek Koszalka, Andrzej Kasprzak

Department of Systems and Computer Networks,  
Wroclaw University of Technology,  
Wroclaw, Poland

E-mail: {156669, 156594, iwona.pozniak-koszalka, leszek.koszalka, andrzej.kasprzak}@pwr.wroc.pl

**Abstract**—In this paper, we present how to allocate bandwidth in VoIP networks, how to set up a connection, and which links in the network should be used. Common are algorithms which find shortest path (SPF) – like OSPF in Internet network. We create a new algorithm which takes into consideration two parameters, including maximum throughput bound. Using the created experimentation system we show that the algorithm can reduce the number of rejected calls and in some cases, the total cost of the calls.

**Keywords**—bandwidth; algorithm; experimentation system; optimization; efficiency.

## I. INTRODUCTION

Nowadays, one can observe an increasing number of computer network users. Voice over Internet Protocol (VoIP) is a general term for a family of transmission technologies that deliver voice communications over IP networks. The Internet and other packet-switched networks are particular examples. In this paper, we address the problem of how VoIP applications should be allocated bandwidth in networks. The objective is to find, in real-time, an optimal path in the network when establishing a VoIP connection between particular nodes.

The rest of the paper is organized as follows: after introducing the necessary terminology in Section II, algorithms that solve the allocation problem are briefly described in Section III (basic algorithm) and Section IV (modified algorithm), next, the results of investigations are presented and discussed. The designed and implemented experimentation system is described in Section V and the analysis of results of simulations made along with design of experiments is in Section VI. The conclusions appear in Section VII.

## II. PROBLEM STATEMENT

Any network topology can be modeled as a graph in which vertices correspond to nodes and edges correspond to links (physical or logical) between the nodes. Two mainstream network architectures are considered. In the first considered case, nodes can be one of two possible types: *relay nodes* or *edge nodes*. Relay nodes transfer VoIP traffic, whereas edge nodes generate other network

traffic, and use IP telephone (wireless phone, soft-phone or hands-free phones) generate traffic [1] [2]. This paper considers both wired as well as wireless networks. A significant research has been conducted in the area of 802.11-based wireless solution [3] [4]. For the second type of network, all the nodes belong to the same category. It is no distinction between nodes. This conception can be used when we have large network and each node is treated as small sub-network. Wireless network can be also described in the same way. The following terminology is used:

*Constraints: cost and throughput.* For a given link, defined by nodes  $v$  and  $w$ , it is natural to assume there is an associated cost, denoted  $c_{vw}$ : here, the meaning of 'cost' is deliberately left open to interpretation, as it often depends on user or hardware specifications. For example, according to the P2P model described in [5], 'cost' can mean: delay in milliseconds, distance in kilometers, and number of ISPs between two nodes. Furthermore, every link, defined by nodes  $v$  and  $w$ , will have a limited throughput, denoted  $d_{vw}$  (typically measure in kbps).

*Connections.* When a connection is established, it is allocated a bandwidth  $b$ . Therefore, the value  $b$  must be subtracted from the throughput  $d_{vw}$  every time a connection is established, assuming nodes  $v$  and  $w$ , lie on the connection path. If  $d_{vw} < b$ , no new connection can be established over the link: in that case, a new connection path must be sought, avoiding the node-pair  $(v, w)$ . If no such path can be found, the connection request is rejected. To simplify calculations, the bandwidth  $b$  is chosen as the measurement unit, and the throughput values  $d_{vw}$  are normalized with respect to multiples of  $b$ . E.g., if  $b$  corresponds to 32 kbps (codec G.726 [6]) and  $d_{vw}$  corresponds to 2 Mbps, we set  $b' = 1$  and  $d_{vw}' = 128$ .

*Calls.* All the call data are stored in a connection schedule (an example in Table I), including: time  $t_s$ , when the connection starts, nodes  $v$  and  $w$  defining start/end of the connection, and time  $t_t$  when the call is terminated  $t_t$ . When a connection is established between  $v$  and  $w$ , all the nodes and links lying in the connection path, as well as the associated costs, are stored in the connection schedule.

**Objectives.** The main objective of bandwidth allocation is to try and ensure that all the connection requests are satisfied (equivalently, the percentage of rejected requests are minimized). Another objective is to minimize the total cost of connections, where cost is represented by  $v_{ab}$ . Yet another objective is to reduce the number of terminated calls while the network resources are decreasing, but that can have an affect on the call costs.

### III. THE ALGORITHMS

#### A. Basic Algorithm

Let us have a look at an example with a given connection schedule (composed of 5 rows - see Table I) and network topology as in Fig. 1 (with nodes enumerated from 1 to 9).

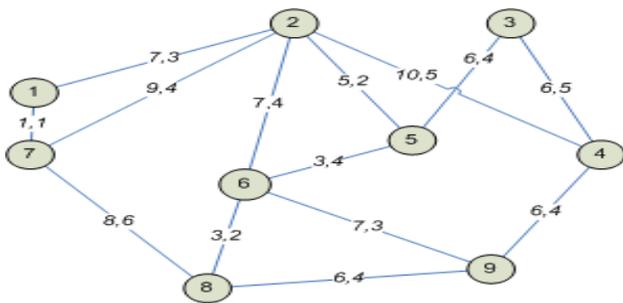


Figure 1. A network topology.

Each link is characterized by two attributes: the cost of the link,  $c_{vw}$ , and, the throughput,  $d_{vw}$  (it is a normalized value, so for e.g.,  $d_{vw} = 2$  means that two connections can be made simultaneously). It is assumed, that only one call is permitted per time slot  $[t_s, t_r]$ . We present, step by step, how the Basic Algorithm, i.e. Short Path Finding (SPF) algorithm works.

**Step 0.** In first iteration ( $t_s=0$ ) there is one connection to set up.

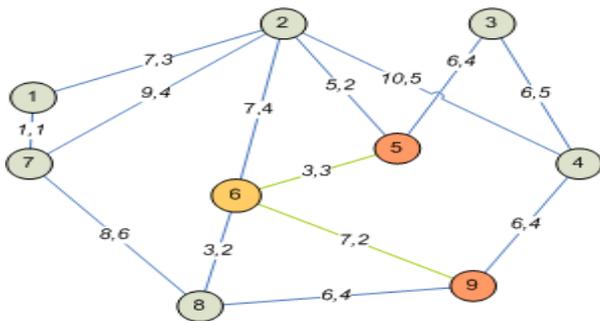


Figure 2. After first iteration (Step 0).

The shortest path (which can be verified using Dijkstra algorithm, for example) is shown in Fig. 2 - it may be observed that transit node is node 6. The total cost of the path is  $c_{5,6} + c_{6,9} = 3 + 7 = 10$ . Remark: The value  $d_{vw}$  must be decreased accordingly, for each link on the

connection path. At the end of any iteration, existing connections are required termination. In this particular example, no such action is required.

**Step 1.** Another connection is set up (see row 2 in Tab. 1). Once again it is the shortest path in the graph (Fig. 3) with transit nodes 5-2-1 and the total cost equal to 19. Is this satisfactory? No! The link between nodes 1 and 7 cannot be used for another connection. Two other paths are possible (3-5-6-8-7 and 3-5-2-7), each with total cost = 20 (i.e. just one unit more), but neither consumes the total throughput of any link in the network.

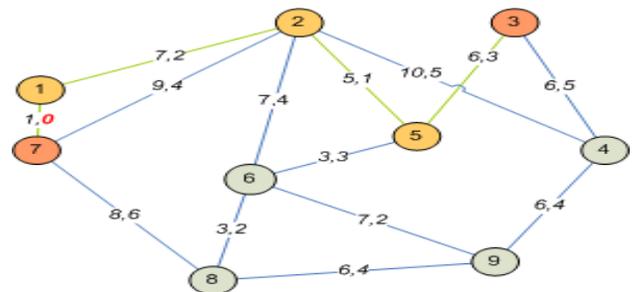


Figure 3. After second iteration (Step 1).

**Step 2.** The result of this iteration (path 4-9-3 with the total cost of 13) is shown in Fig. 4.

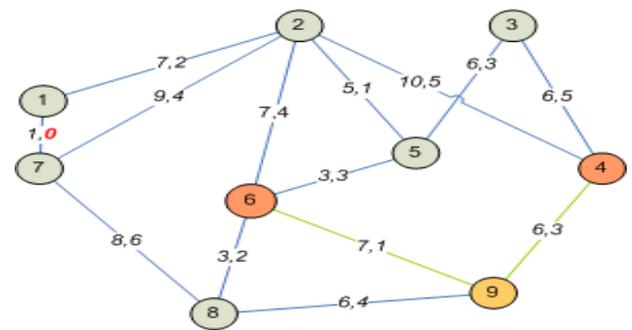


Figure 4. After third iteration (Step 2).

**Step 3.** The results of this iteration (path with no terminal node) is shown in Fig. 5

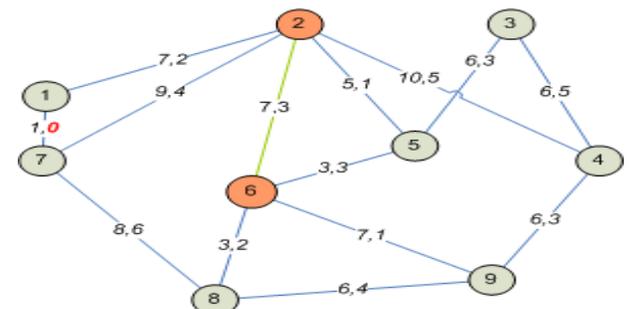


Figure 5. After fourth iteration (Step 3).

At the end of this iteration, only one connection is terminated (between nodes 5 and 9). In this case, we must increment  $d_{vw}$  for each link participating in this connection. The result is shown in Fig. 6.

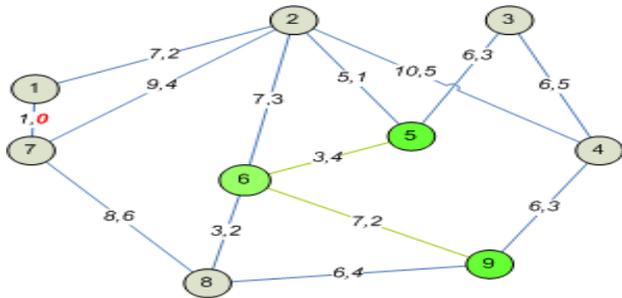


Figure 6. Disconnection.

**Step 4.** Here, a connection needs to be set up between node 1 and node 9. If we only consider the values  $c_{vw}$ , the optimal path is 1-7-8-9. But link 1-7 is already operating at full capacity. Thus, a different path needs to be found.

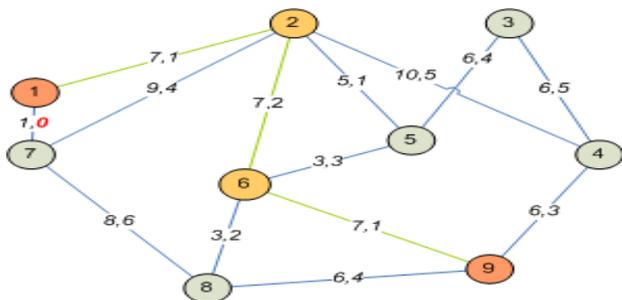


Figure 7. The result of Step 4.

The new path found (1-2-6-9) has a cost of 21. It is 6 units more than the cost of the 1-7-8-9 path, which can not be reached by SPF. The summary of the obtained results with SPF is given in Table I, there are also results given by the created algorithm described in Section III B.

TABLE I. ALGORITHMS PERFORMANCE

Step	$t_s$	$v$	$w$	$t_i$	Basic		Premaru	
					Transit nodes	Cost	Transit nodes	Cost
0	0	5	9	3	6	10	6	10
1	1	3	7	6	5,2,1	19	5,2	20
2	2	4	6	6	9	13	9	13
3	3	6	2	12	-	7	-	7
4	4	1	9	7	2,6	21	7,8	15
<b>Total Cost</b>						<b>70</b>		<b>65</b>

This simple example shows all the most important processes occurring in VoIP applications: setting up connections and/or making disconnection. It also

illustrates the difficulty of finding optimal paths: the problem is to find a path having a minimal cost, and such that other potential connections will not be blocked in subsequent iterations. In Step 1 we could have set up the path 3-5-2-7. Then the path cost in iteration 1 would be 20 (one more than it is now), but the advantage is that the connection 1-7-8-9 could then be set up in Step 4, thus, sparing five units of cost. If several more connection paths were to be badly chosen, the whole network itself might become completely saturated after just a few iterations, rendering future calls impossible. Moreover, poorly chosen connection paths severely compromise network reliability, leading to widespread user dissatisfaction.

*B. Algorithm Premaru*

The created algorithm, named Premaru, can compete with SPF. The idea of Premaru is based on introduction of two parameters denoted as  $p$  and  $q$ . The fundamental rule of the designed algorithm is following: “If the lowest value  $d_{vw}$  of all links in a found path is lower than  $q$ , moreover, there is any other path that connects the same nodes and its total cost is bigger than previous path by  $p$  or less, then the second (another) path will be chosen”. It would avoid blocking beneficial paths in further iterations. This procedure is described by the pseudo-code and the block-diagram in Fig. 8.

Function *find\_shortest\_path* ( $v, w, C$ ) finds the shortest path between nodes  $v$  and  $w$ . It uses the cost matrix  $C$  of the whole network. Function *throughput* ( $path, D$ ) returns the minimum throughput,  $d_{min}$ , of all edges in  $path$ . It exploits the throughput matrix,  $D$ , of the whole network. If the minimum throughput of the found path is lower than  $q$  we zeroes edge’s throughput. That prevents reuse of edges having a throughput lower than  $q$ .

After that, a search is made for another path, named  $path_{temp}$ . If  $path_{temp}$  satisfies the condition  $cost(path_{temp}) - cost(path) \leq p$ , then this path is chosen. It enables to have a minimum throughput larger than  $q$ . This avoids traffic-crowded edges -  $path_{temp}$  is calculated using a temporary matrix of throughputs  $D_{temp}$ , where all throughputs not greater than  $q$  are omitted.

```

path = find_shortest_path(v, w, C, D);
d_min = throughput(path, D);
if d_min <= q then
    D_temp = throughput_zeroing(q, D);
    path_temp = find_shortest_path(v, w, C, D_temp);
    if cost(path_temp) - cost(path) <= p
        then path = path_temp;
    end if
end if
set_up_connection(path, D);
    
```

Figure 8. Algorithm Premaru – pseudo-code description.

Function  $find\_shortest\_path(v, w)$  is left undefined. It can be the Dijkstra algorithm [7], as suggested earlier, but it can be the Bellman-Ford algorithm [7] or a heuristic algorithm (such as genetic algorithm), either. In our recent implementation, the modified Dijkstra algorithm was utilized - it finds the shortest path, taking into account  $d_{vw}$  (if  $d_{vw}$  is equal to 0, the edge  $v-w$  cannot be used).

#### IV. EXPERIMENTATION SYSTEM

The experimentation system has been designed and implemented by our research team following ideas presented in [8]. The core module of the system is simulation environment – a complex program which allows testing both considered algorithms (*Algorithms Performance Module*) and making multi-aspect investigations. The user can choose an experiment design which will be performed in automatic manner (*Experiment Design Module*) generating input parameters, including network matrices ( $C$  – cost matrix and  $D$  – throughput matrix) and connection schedule ( $con\_gen$ ). The values of Premaru parameters  $p$  and  $q$  can be also specified by the user. Output data is stored on a properly designed database (Data Acquisition Module). In Fig. 9, the experimentation system is shown as input – output plant.

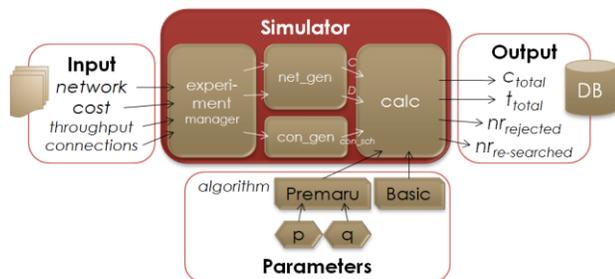


Figure 9. Experimentation system as input – output plant.

Terminology used in Fig. 9 is explained below:

- *network* – information about number of nodes and network density,
- *cost* – information about cost distribution (homogenous or Weibull) and its parameters (maximum cost for homogenous distribution),
- *throughput* – maximum throughput of a link (parameter of homogenous distribution),
- *connections* – the number of connections,  $\lambda$  is the parameter of the exponential distribution,
- *Basic algorithm* (also named Dijkstra) – a simple Dijkstra's algorithm, modified for our purposes.
- *Premaru algorithm* – the created algorithm that includes parameters  $p$  and  $q$ ,
- $C_{total}$  – the total cost of all connections,
- $t_{total}$  – the total time of simulation,
- $nr_{rejected}$  – the number of rejected calls,

- $nr_{re-searched}$  – the number of calls for which was done more than one path search.

The application software was written using C# and .NET 3.5 Windows Forms, because they constitute a highly flexible platform for the design of user-friendly interfaces using familiar components (check-boxes, text-boxes, buttons, etc.). An external library was used to produce the charts. The database was compiled and maintained using SQLite. The system (simulator) only requires a PC that runs Windows OS and .NET Framework 3.5 or higher (freely downloadable from the Microsoft corporation website).

An important issue is experiment design, i.e., creating input data that models real-life scenarios. According to aspects specified in Section II the details about such scenarios may be described as follows.

*The Network.* The network is represented by the graph described by the pair of matrices  $c_{vw}$  and  $d_{vw}$ . The entries of the first matrix represent costs between nodes  $v$  and  $w$ . The entries of the second matrix represent the throughputs between nodes  $v$  and  $w$ . Those matrices are highly correlated, because if a non-zero entry in one of them corresponds to a non-zero entry in the other. Accordingly, for each link between nodes we have a pair of values: cost and throughput. The size of the network is also important. Tests can be made for 'small' (50 nodes) and 'large' networks (100 nodes). In the case of complete graphs (networks), each pair of nodes is connected, i.e. it represents network with 100% density. But such networks are too expensive and rarely used, in practice. The system allows generating networks with density from 3% to 80%. Each link between nodes is assigned a cost value, which is stored in matrix  $C$  as  $c_{vw}$ . This value might represent the time-delay (in time units - milliseconds, for example). Realistic cost values can be generated by exploiting two probability distributions: (i) homogeneous distribution, (ii) Weibull distribution. With a homogeneous distribution, cost values are drawn from 1 to  $k$ , where  $k$  determines the maximum value of cost (parameter  $k$  can be chosen in our system). The Weibull distribution provides a good mathematical model of delay in VoIP networks [9]. The Weibull distribution is characterized by:  $k$  – a shape parameter and  $\lambda$  – a scale parameter. Each link between nodes is assigned a throughput value, which is stored in matrix  $D$  as  $d_{vw}$ . Throughput for every link is drawn from homogeneous distribution. The throughput can be from 1 to  $p$ , where  $p$  represents the maximum throughput bound).

*Connection schedule.* In order to construct a good connection schedule, it is important to know how subscribers behave. In our research we assumed that the starting time of a connection can be accurately described by random variable with a homogeneous distribution. We

also assumed that the duration of a connection can be accurately described by a random variable having an exponential distribution. This assertion is based on the results presented in [10]. The probability density function of phone call duration described there can be modeled by an exponential distribution with  $\lambda=0.02$ . Our experimentation system allows changing the value of  $\lambda$ .

To generate a list of connections necessary parameters are: (i) Network size (number of nodes), (ii) Number of connections, (iii)  $\lambda$ -parameter in exponential distribution. The way of selecting nodes when a connection is being established is modeled by a homogenous distribution. There is only one condition: the initial node must not be the final node. The time-instant of establishing a connection is also modeled by a homogenous distribution.

V. INVESTIGATIONS

A. Design of experiment

The following input data were taken into consideration:

- Network size : 100 nodes.
- Network density: 3%, 6%, 9%, 12%, 15%, 18%, 21%, 24%, 27%, 30%, 35%, 40%, 45%, 50%, 60%, 70%, 80% .
- Number of connections: 1000, 2000, 3000, 4000, 5000, 7500, 10000.
- The cost values of links: generated using Weibull distribution with  $k = 2.09$  and  $\lambda = 7.5$ .
- Call duration: modeled using the exponential distribution with  $\lambda = 0.02$ .
- The Premaru parameters ( $p, q$ ): 16 different pairs – all combinations for  $q = 1, 2, 3, 4$  and for  $p = 1, 2, 3, 4, \dots$ , i.e. (1, 1), (1, 2), ... , (3, 4), (4, 4) .

B. Number of rejected calls

In Fig. 10, one can see the average results given by SPF Basic algorithm and Premaru algorithm.

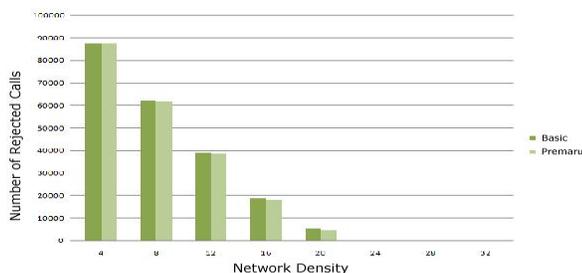


Figure 10. Number of rejected calls in relation to network density.

It may be observed, that number of rejected calls is almost the same when Premaru algorithm is used instead of Basic algorithm. We found that these differences were very small - less than 5% (Fig.11). Moreover, the network density do not influence on number of rejected calls rate.

It was only observed that the number of rejected calls was linearly proportional to the number of connections.

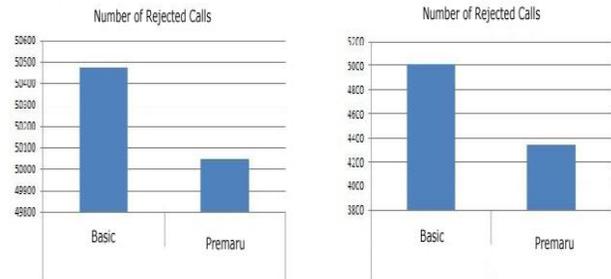


Figure 11. Number of rejected calls for network density 10% (on left) and 20% (on right).

C. The total cost

The average total cost given by Basic algorithm and Premaru algorithm in relation to the number of calls for various network densities is shown in Fig.12.

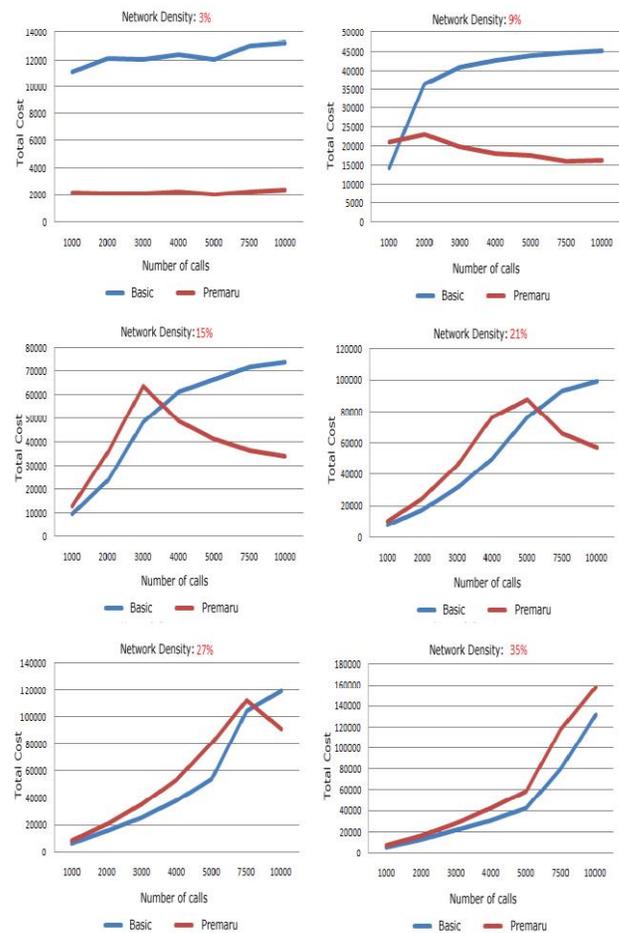


Figure 12. The total cost in relation to the number of connections for different network densities.



# On the Receiver Performance in MU-MIMO Transmission in LTE

Zijian Bai, Stanislaus Iwelski, Guido H. Bruck, Peter Jung  
 Department of Communication Technologies  
 University of Duisburg-Essen, Oststrasse 99, 47057 Duisburg, Germany  
 Email: {first.surname}@kommunikationstechnik.org

Biljana Badic, Tobias Scholand, Rajarajan Balraj  
 Intel Mobile Communications (IMC)  
 Düsseldorf Landstrasse 401, 47259 Duisburg, Germany  
 Email: {first.surname}@intel.com

**Abstract**—Multiuser multiple-input multiple-output (MU-MIMO) transmission scheme has drawn most attentions during the recent development of Long Term Evolution (LTE) systems. Based on the feedback information of the downlink channel, evolved NodeB may achieve multiple accesses via MIMO technology in MU-MIMO Transmission Mode and allow user equipments to share resources in frequency and time domain. In this paper, we review several signal detectors and evaluate their performance in MU-MIMO transmission. The review work aims at the feasibility study of receivers in LTE systems. Different scenarios have been considered in the evaluation progress, e.g. low and high spatial correlation, real channel estimation and feedback delay. Simulation results show that benefits can be obtained by MU-MIMO transmission in the spatial correlated MIMO channel due to the higher condition number in the channel. Besides, reviewed receivers with the co-channel interference-aware signal detection yield good performance compared with single user MIMO receivers. The interference aware receivers are also robust in MU-MIMO transmission with imperfect working conditions, including channel estimation errors and precoding matrix index feedback delay.

**Index Terms**—LTE (Long Term Evolution), MU-MIMO (Multiuser Multiple Input Multiple Output), Max-Log-MAP (Maximum A Posteriori Probability), IRC (Interference Rejection Combiner)

## I. INTRODUCTION

LTE (Long Term Evolution) is the trademarked project name of a high performance air interface for cellular telephone. It is a project of the 3rd Generation Partnership Project (3GPP) for a set of enhancements to the Universal Mobile Telecommunications System (UMTS), hence is termed as E-UTRA (Evolved UMTS Terrestrial Radio Access). The LTE specifications [1] provide downlink peak rates of at least 300 Mbit/s, an uplink of at least 150 Mbit/s associated with 20 MHz channel bandwidth and RAN round-trip time of less than 10 ms. Scalable carrier bandwidths, from 1.4 MHz to 20 MHz, and both frequency division duplexing (FDD) and time division duplexing (TDD) are supported in LTE.

The key technologies in LTE systems are Orthogonal Frequency Division Multiple Access (OFDMA) [2] in the downlink and Single Carrier Frequency Division Multiple Access (SC-FDMA) [3] for uplink transmission. Furthermore, Multi-Input Multi-Output (MIMO) is considered for increasing system capacity in LTE, facilitating spatial multiplexing (SM) and Alamouti based transmit diversity (TxD) schemes.

Besides single user MIMO (SU-MIMO) Transmission Modes, multiuser MIMO (MU-MIMO) transmission is supported as well. MU-MIMO allows base station (eNodeB) to communicate with different radio terminals simultaneously by means of space division multiple access, whereas SU-MIMO only considers access to multiple antennas at a single terminal. The benefit of MU-MIMO transmission is the naturally independent signals cross the antennas mounted at different UEs, which are physically distributed. This allows less restriction on the MIMO channel condition for applying MIMO technology in signal transmission, whereas SU-MIMO requires good uncorrelated signals between antennas at the same user equipment (UE) for SM transmission scheme. MU-MIMO transmission is realized with the precoding and UE pairing in LTE systems, where different precoding matrices or vectors [1] are reported by the paired UEs via feedback channel.

In this paper, we look at different signal receivers and evaluate their performance in MU-MIMO transmission in LTE systems. Related work has been carried out in e.g. [4] and [5] for performance evaluation in system-level. In contrast, we focus on the receiver structures and evaluation of their performance in link-level with realistic working conditions, e.g. channel estimation error, closed-loop feedback delay and different modulation and coding schemes. The aim of the evaluation is to review the reliability and robustness of different receivers in MU-MIMO transmission and to exploit the potential benefits of MU-MIMO transmission in LTE systems compared with SU-MIMO scheme.

In what follows, complex base band notation will be used, deploying matrix vector calculus, for describing the system structure and the signal processing. Discrete-time variables will be denoted by vectors which are given as lower case characters in bold face italics. Matrices will be denoted by upper case characters in bold face italics. Complex values will be underlined. Furthermore,  $(\cdot)^H$  denotes the Hermitian of a vector or a matrix,  $\mathbf{I}$  is the identity matrix and  $E\{\cdot\}$  gives the expectation value of given random variables.  $\Re\{\cdot\}$  and  $\Im\{\cdot\}$  denote the real and the image parts of the given complex value, respectively.

The reminder of this paper is organized as follows. Section II gives an overview of the LTE system with focus on the precoding matrix index feedback and UE pairing scheme in MU-MIMO transmission. Different MIMO receivers are

discussed in Section III. Section IV summarizes the performance of the receivers in MU-MIMO transmission in different working scenarios in LTE systems. Section V concludes this paper.

## II. LTE SYSTEM OVERVIEW

### A. System Model

We consider 3GPP 36-series specifications for LTE systems as the baseline for our following discussions. As presented in [6], the system constructed by the LTE physical layer technologies, namely the OFDMA and MIMO, approaches to a bit-interleaved coded modulation (BICM) system [7]. This yields transmissions of the source bits in LTE systems are independent and meet the channel with same quality. The mutual information between transmitted and received bits is maximized in such systems. In what follows, a general system concept of LTE is presented with  $N_T$  transmit (Tx) antennas,  $N_R$  receive (Rx) antennas and  $N_L$  transmission layers (the parallel spatial data streams). Owing to the orthogonality between the subcarriers in each OFDM symbol, we can represent the subcarrier specific SU-MIMO transmission in a single-tap channel scenario in LTE systems by

$$\mathbf{r} = \mathbf{H} \mathbf{P} \mathbf{d} + \mathbf{n}, \quad (1)$$

where  $\mathbf{r} \in \mathbb{C}^{N_R}$  represents the received signal vector,  $\mathbf{H} \in \mathbb{C}^{N_R \times N_T}$  the MIMO channel matrix,  $\mathbf{P} \in \mathbb{C}^{N_T \times N_L}$  the applied precoding matrix,  $\mathbf{d} \in \mathbb{C}^{N_L}$  the transmitted signal vector and  $\mathbf{n} \in \mathbb{C}^{N_R}$  the additive white Gaussian noise (AWGN) vector at the receiver. Without loss of generality, mutually uncorrelated signal elements are assumed in the transmitted signal vector with covariance matrix  $\mathbf{R}_d = E_s \mathbf{I}$  and  $\mathbf{n}$  is a zero mean circularly symmetric complex Gaussian (ZMCSCG) random vector with  $\mathbf{n} \sim CN(\mathbf{0}, N_0 \mathbf{I})$ . We further define the elements in  $\mathbf{d}$  are drawn from a quadrature amplitude modulation (QAM) constellation  $M = 2^Q$ , where  $Q$  is the number of bits per symbol.

We consider MU-MIMO transmission with  $N_{UE}$  UEs by adding  $(N_{UE} - 1)$  UE signals into the downlink transmission. The received signals at the  $k^{\text{th}}$  UE can be represented by,

$$\mathbf{r}_k = \sum_{n_{UE}=1}^{N_{UE}} \mathbf{H}_k \mathbf{P}_{n_{UE}} \mathbf{d}_{n_{UE}} + \mathbf{n}_k, \quad (2)$$

with mutually independent data from different UEs.

Different transmission modes are supported in LTE systems. Transmission Mode 2 and 3 are used for open-loop transmissions with TxD and SM [6] [8], respectively. Transmission Mode 4, 5 and 6 are related to the closed-loop transmission with precoding matrix index (PMI) feedback for SU-MIMO SM, MU-MIMO and SU-MIMO with single layer, respectively. In the context of this paper, we concentrate our discussion on Transmission Mode 5 and for the reference purpose Transmission Mode 6.

For Transmission Mode 6 in LTE systems, we represent the system function by setting in  $N_L = 1$  (1)

$$\mathbf{r} = \mathbf{H} \mathbf{p} \mathbf{d} + \mathbf{n}, \quad (3)$$

in which the precoding matrix is degraded to be a precoding vector  $\mathbf{p} \in \mathbb{C}^{N_T \times 1}$ .

For Transmission Mode 5, MU-MIMO transmission, two UEs with single layer to each UE are scheduled in LTE systems [8]. Using (2) and setting  $N_{UE} = 2$ , we can represent the system function for the  $k^{\text{th}}$  UE by

$$\mathbf{r}_k = \mathbf{H}_k \mathbf{p}_1 \mathbf{d}_1 + \mathbf{H}_k \mathbf{p}_2 \mathbf{d}_2 + \mathbf{n}_k, \quad k = 1, 2. \quad (4)$$

### B. Spatial Channel Model

For benchmarking, 3GPP introduced spatial channel model (SCM) in [9]. In particular, Section A1 and A2 of [9] provide guidelines for the SCM simulations by setting out from the general concept of [10]. In [9] four representative SCM test cases including antenna configurations and realistic antenna correlation properties are introduced, also explicitly considering antenna polarization. The SCM test case SCM-A refers to the suburban macro propagation, SCM-B and SCM-C consider the urban macro cellular environment with either low or wide angle spreads, and SCM-D reflects the case of an urban micro cellular scenario. In the context of this paper, we consider the SCM-B and SCM-D test cases as the baseline to evaluate the system performance of MU-MIMO transmission.

### C. PMI Selection

As specified in [6] [8], the closed-loop principle has been introduced in LTE systems downlink transmission by reporting back different information from UE to the eNodeB through the uplink channel in a periodic or aperiodic fashion [8]. The feedback information, the PMI, provides information about the preferred precoding matrix in a codebook based precoding [6]. The reported PMIs from UE help the eNodeB inquiring the knowledge of the downlink channel state information (CSI). Using the channel information a UE severed by eNodeB may find the most suitable precoding matrix which aligns own signals to the own downlink channel state. This procedure helps UEs in both Transmission Mode 5 and Transmission Mode 6 to improve the desired received signal energy. Considering a single transmission layer per UE in those two modes, we conclude that a matched filter (MF) based receiver with maximum likelihood (ML) decision rule will maximize the signal strength for the target UE without prerequisite of the knowledge of Transmission Mode [11]. The MF receiver can be represented by the filter vector

$$\mathbf{m}_k^H = \left( \mathbf{H}_k \mathbf{p}_k \right)^H, \quad k = 1, 2. \quad (5)$$

The corresponding post processing signal to noise ratio (Post-SNR) is given by

$$\gamma_{MF} = \left\| \mathbf{H}_k \mathbf{p}_k \right\|^2 \gamma, \quad k = 1, 2, \quad (6)$$

with  $\gamma$  being the SNR at each receiver antenna. The most suitable PMI will be selected to maximize the Post-SNR at the output of the MF receiver. Therefore, we may represent the

selection of the best precoding vector  $\underline{p}$  by using the following maximum Post-SNR criterion

$$\underline{p}_k = \arg \max_{\underline{p}_i \in \mathcal{P}} \left\| \underline{H}_k \underline{p}_i \right\|^2, \quad k = 1, 2, \quad (7)$$

with  $\mathcal{P}$  being the corresponding codebook for particular number of Tx antennas [6].

The above discussion gives the precoding matrix selection in a particular subcarrier. However, the more interesting scenario of PMI selection in LTE systems is the wideband selection due to the limitation of the uplink channel capacity for feedbacks. The wideband PMI selection requires one preferred precoding matrix being selected for a group of subcarriers in LTE systems [8]. The extension of (7) for the wideband precoding matrix selection in a bandwidth of  $B_{\text{PMI}}$  including  $N_{\text{subc}}$  subcarriers is

$$\begin{aligned} \underline{p}_k &= \arg \max_{\underline{p}_i \in \mathcal{P}} \sum_{j=1}^{N_{\text{subc}}} \left\| \underline{H}_{k,j} \underline{p}_i \right\|^2 \\ &= \arg \max_{\underline{p}_i \in \mathcal{P}} \left( \underline{p}_i^H \underline{R}_{\text{Tx}} \underline{p}_i \right), \\ k &= 1, 2. \end{aligned}$$

with  $\underline{H}_{k,j}$  being the subcarrier specific MIMO channel matrix in the  $j^{\text{th}}$  subcarrier at the  $k^{\text{th}}$  UE and  $\underline{R}_{\text{Tx}}$  being the estimated transmitter antenna correlation matrix based on the channel matrices in  $N_{\text{subc}}$  subcarriers [12].

#### D. UE Pairing in MU-MIMO

A good UE pairing scheme is required at the eNodeB to choose the correct pair of UEs for MU-MIMO transmission in LTE systems. This pairing scheme shall maintain less interference between scheduled UEs in MU-MIMO transmission. A proper pairing scheme can be designed by maximizing the Chordal distance between the feedback precoding matrices of the UEs. The Chordal distance between two matrices is given in [13] and represented by

$$d_{\text{chord}}(\underline{p}_i, \underline{p}_j) = \frac{1}{\sqrt{2}} \left\| \underline{p}_i \underline{p}_i^H - \underline{p}_j \underline{p}_j^H \right\|_{\text{F}} \quad (8)$$

with  $\|\cdot\|_{\text{F}}$  being the Frobenius norm of the matrix.

The Chordal distance generalizes the distance between points on the unit sphere through an isometric embedding from complex Grassmann manifold  $\mathcal{G}(N_T, N_L)$  to the unit sphere [13]. Assuming an infinite number of UEs served by the current eNodeB, the  $k^{\text{th}}$  UE with reported precoding matrix  $\underline{p}_k$  will be paired with the  $m^{\text{th}}$  UE, where the  $m^{\text{th}}$  UE reports precoding matrix  $\underline{p}_m$  and the Chordal distance between  $\underline{p}_k$  and  $\underline{p}_m$  is maximized. With the maximized Chordal distance criterion  $\underline{p}_m$  stays in the antipolar position of  $\underline{p}_k$  and hence  $\left\| \underline{H}_k \underline{p}_m \right\|^2$  is minimized yielding the minimized inter-UE interference in (4).

In summary, the UE pairing scheme for MU-MIMO transmission in LTE systems can be designed in the way of finding

the  $m^{\text{th}}$  UE for the  $k^{\text{th}}$  UE based on the reported precoding matrices and the following criterion,

$$\underline{p}_m = \arg \max_{\underline{p}_i \in \mathcal{P}_{\text{UE}}} d_{\text{chord}}(\underline{p}_i, \underline{p}_k). \quad (9)$$

with  $\mathcal{P}_{\text{UE}}$  being the pool containing all reported precoding matrices at a certain eNodeB.

### III. RECEIVER DESIGN FOR MU-MIMO

Typical receivers in a MIMO-OFDM system for spatial multiplexing transmission can be categorized with their signal processing styles, e.g. non-linear and linear receivers [14]. We extend the structure of those receivers for MU-MIMO transmission in this section. Considering the symmetric system function faced by both UEs in MU-MIMO transmission, we simplify our discussion on the receiver structure for the first UE, i.e.  $k = 1$  in (4). With the definitions of the effective channel matrix

$$\underline{g}_e = \underline{H}_1 \underline{p}_1, \quad (10)$$

and the interference channel matrix

$$\underline{g}_t = \underline{H}_1 \underline{p}_2, \quad (11)$$

we represent (4) by

$$\underline{r}_1 = \underline{g}_e \underline{d}_1 + \underline{g}_t \underline{d}_2 + \underline{n}_1. \quad (12)$$

#### A. Linear Block Receiver

The linear block receiver with zero-forcing (ZF) or minimum mean squared error (MMSE) criterion applies linear filter on the received signal vector  $\underline{r}_1$  to compensate the channel distortion and the inter-symbol interference. The output signal vector is the product of the linear filter matrix and the received signals

$$\tilde{\underline{d}}_1 = \underline{m}^H \underline{r}_1. \quad (13)$$

The filter  $\underline{m}$  is constructed as

$$\underline{m}_{\text{ZF}}^H = \left( \underline{g}_e^H \underline{g}_e \right)^{-1} \underline{g}_e^H, \quad (14)$$

with ZF criterion and

$$\underline{m}_{\text{MMSE}}^H = \left( \underline{g}_e^H \underline{g}_e + \frac{1}{\gamma} \right)^{-1} \underline{g}_e^H, \quad (15)$$

with MMSE criterion. Due to the single transmission layer per UE in MU-MIMO transmission, we conclude that a linear block receiver with either ZF or MMSE criterion yields the same system performance as the MF filter in MU-MIMO transmission. In (6) the Post-SNR at the output of linear block receivers is calculated. Using the linear block receiver, the UE assumes that interference coming from the paired UE via  $\underline{g}_t$  are fully suppressed. However with low resolution or size limited codebooks in LTE, the residual interference is still significant strong and causes an error floor of the signal detection at the first UE. Non-perfect working condition such as feedback delay and channel estimation error will cause more performance degradation.

### B. Interference Rejection Combiner

Taking the remaining interference in the received signal vector into account, the receiver can be built in a fashion of suppressing interference and improving the Post-SNR. The interference aware receiver, such as the interference rejection combiner (IRC) in [15] is designed in this manner. We represent the IRC receiver structure in LTE systems here. Following [15], the general IRC structure for the received signal is

$$\underline{\mathbf{m}}_{\text{IRC}}^{\text{H}} = \frac{\underline{\mathbf{g}}_{\text{e}}^{\text{H}} \underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}^{-1}}{\underline{\mathbf{g}}_{\text{e}}^{\text{H}} \underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}^{-1} \underline{\mathbf{g}}_{\text{e}}}, \quad (16)$$

with  $\underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}$  being the covariance matrix of the sum of the interference and noise terms,

$$\underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}} = \text{E} \left\{ (\underline{\boldsymbol{\eta}} - \text{E} \{ \underline{\boldsymbol{\eta}} \}) (\underline{\boldsymbol{\eta}} - \text{E} \{ \underline{\boldsymbol{\eta}} \})^{\text{H}} \right\}, \quad (17)$$

and

$$\underline{\boldsymbol{\eta}} = \underline{\mathbf{g}}_{\text{t}} \underline{d}_2 + \underline{\mathbf{n}}_1. \quad (18)$$

The IRC introduces a whitening filter  $\underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}^{\frac{1}{2}}$  before the modified MF  $\underline{\mathbf{g}}_{\text{e}}^{\text{H}} \underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}^{\frac{1}{2}}$  to whiten the interference plus noise with respect to sending signals. Hence the IRC receiver performs linear operations and improves the Post-SNR by suppressing interference and noise terms at the same time. The Post-SNR at the output of the IRC receiver can be represented by

$$\gamma_{\text{IRC}} = \underline{\mathbf{g}}_{\text{e}}^{\text{H}} \underline{\mathbf{R}}_{\underline{\boldsymbol{\eta}}}^{-1} \underline{\mathbf{g}}_{\text{e}}. \quad (19)$$

### C. Max-Log-MAP Receiver

A Max-Log-MAP (Max-Log maximum a posteriori) receiver is presented in [7] for calculating the bit metric of bits in transmitted signals based on the BICM concept and the Max-Log-MAP criterion. An extension of the bit metric calculation on LTE MU-MIMO transmission in (4) may be represented by

$$\Lambda^i(\mathbf{r}_1, c_i = b) = \min_{\underline{d}_1 \in \chi_{1,c_i}^b, \underline{d}_2 \in \chi_2} \left\| \mathbf{r}_1 - \underline{\mathbf{g}}_{\text{e}} \underline{d}_1 - \underline{\mathbf{g}}_{\text{t}} \underline{d}_2 \right\|^2, \quad (20)$$

$$i = 1, 2, \dots, Q,$$

with  $\chi_{1,c_i}^b$  being the subset of the signal set  $\chi_1$  for  $\underline{d}_1$  in constellation with  $b \in \{0, 1\}$  in bit position  $i$  and  $\chi_2$  being the signal set for  $\underline{d}_2$ . Based on bit metric values for a particular bit in symbols, the soft decision output or the log-likelihood ratio (LLR) of the bit is given as

$$\Lambda_{\text{LLR}}^i(\mathbf{r}_1, c_i) = \Lambda^i(\mathbf{r}_1, c_i = 1) - \Lambda^i(\mathbf{r}_1, c_i = 0). \quad (21)$$

Different to the linear operations given in linear receivers, the Max-Log-MAP receiver applies non-linear operations on the received signal vector and requires more computation effort when the constellation size increases. Besides, (20) requires full-bloom search in two different constellations, namely  $\chi_1$  and  $\chi_2$  for desired signals and interference signals, respectively.

Table I  
SIMULATION PARAMETERS

Parameter	Value
System Bandwidth	10 MHz
Data Subcarriers	600
FFT Size	1024
Cyclic Prefix Length	80 × 1, 72 × 6 in one slot
Subframe/Slot Length	14 OFDM / 7 OFDM in one slot
Channel Model	SCM-B, SCM-D [9]
MIMO	4 × 2 for $N_{\text{T}} \times N_{\text{R}}$
UE Velocity	3 km/h, 30 km/h
Channel Estimation	Ideal Channel Knowledge (ICK), Estimation with cascade one dimensional Wiener interpolator (1DW) [17]
Receiver	MF filter, IRC, Max-Log-MAP [16]
Transmission Modes	MU-MIMO and SU-MIMO
Channel Coding	Turbo codes
Modulation and coding scheme	CQI=4, 7, 10, 13 in [8]
PMI Selection	One PMI for system bandwidth with MaxSNR criterion in Section II-C
UE pairing	Infinite number of UEs available for pairing with criterion in Section II-D
Closed-loop Delay	1 or 8 subframes delay between selection and applying of PMI

## IV. SIMULATIONS

In this section, we present the performance of the receivers discussed in Section III. The performance evaluation is carried out by the simulation results obtained from an LTE specification compliant simulator and is presented in terms of coded block error ratio (BLER) versus SNR in all figures, where the SNR refers to the signal to noise ratio per receive antenna. This LTE compliant simulator has been implemented in C/C++ with all physical layer (PHY) processing and different Transmission Modes in [8] as well as a spatial channel model given in Section II-B.

The executed simulations have been configured following the parameters for LTE system PHY [6]. Table I gives the additional parameters used for obtaining the simulation results presented in this section with Table II containing corresponding transport block sizes, modulation orders and the code rates for the selected CQI values. We select the SCM-B and SCM-D test cases as the simulation scenarios to carry out the MU-MIMO transmission performance in high and low spatial correlated channels, respectively. To obtain the coded system performance with linear receivers, the soft bits after linear operations are generated by following the LLR value computation method in [16].

The performance of the different receivers under the SCM-D 4 × 2 channel model with the UE velocity of 3 km/h and 30 km/h are depicted in Figure 1 and Figure 2, respectively. Being a reference, system performances in Transmission Mode 6 with the MF receiver are also plotted. In both Figure 1 and Figure 2, the simple MF receiver gives the worst performance for the signal detection in MU-MIMO

Table II  
CQI VALUES AND THE CORRESPONDING TRANSPORT BLOCK SIZES

CQI	Modulation	Code Rate	Transport Block Size [8]
4	QPSK	0.301	3624
7	16QAM	0.369	8760
10	64QAM	0.455	16416
13	64QAM	0.754	27376

Transmission Mode. With CQI = 4, i.e. QPSK modulation, the MF receiver provides similar performance as the other receivers. However, for large CQI values, the MF receiver faces error floor. More than 50% blocks cannot be decoded successfully by the MF receiver with CQI = 10 and CQI = 13. The MF receiver is unaware of the co-channel interference given by the interferer UE in MU-MIMO transmission and hence suffers the strong degradation caused by the interference. Such co-channel interference increases when the modulation order increases. Similar to the MF receiver, the IRC receiver deploys linear operation as well and has quite similar performance as the Max-Log-MAP receiver. The maximum performance difference between the IRC and Max-Log-MAP receivers is about 1 dB and 2 dB at  $\text{BLER} = 10^{-2}$  in 3 km/h and 30 km/h scenarios, respectively. Compared with the lower UE velocity, the higher UE velocity shifts the receiver performance by 2 dB for both IRC and Max-Log-MAP receivers. This is caused by the mismatching between the reported optimal PMI and the real channel matrices when the precoding matrices are applied. The higher the UE velocity is, the stronger is the mismatching, and therefore the worse is the receiver performance. However, this impact is limited by 2 dB from 3 km/h to 30 km/h which is acceptable for real systems. Comparing the performance between the Transmission Mode 5 and Transmission Mode 6, we conclude that the co-channel interference causes significant performance loss in Transmission Mode 5. The increased modulation order (e.g. CQI = 10 or 13) generates stronger interference and therefore causes more performance loss in Transmission Mode 5. However, this loss can be compensated by serving more UEs in Transmission Mode 5 than in Transmission Mode 6, yielding a higher system throughput.

Figure 3 presents the performance of different receivers under the SCM-B  $4 \times 2$  scenario with the same system settings as in Figure 1. The MF receiver still suffers large performance degradation or error floor issue in SCM-B channel. Due to the higher spatial correlation in SCM-B than in SCM-D both the IRC and the Max-Log-MAP receivers obtain 1 to 2 dB gain at  $\text{BLER} = 10^{-2}$  in SCM-B channel. This can be explained that each UE in MU-MIMO transmission receives the desired signal only through single layer, which may be aligned to the better sub-channel associated with the dominated eigenvalue from the channel matrix. The stronger spatial correlation yields a larger condition number. That means the UE obtains more channel gain for the desired signals and has less interference from the co-channel interferer UE, which leads to higher signal to interference and noise ratio (SINR) per subcarrier. However,

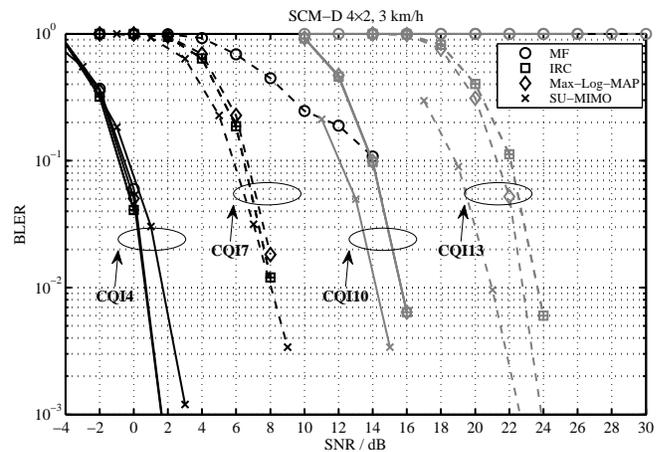


Figure 1. Performance in SCM-D  $4 \times 2$  with ICK and 3 km/h UE Velocity

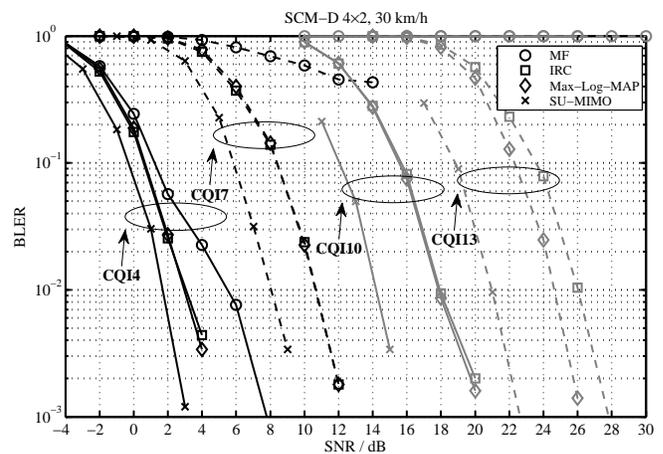


Figure 2. Performance in SCM-D  $4 \times 2$  with ICK and 30 km/h UE Velocity

this performance gain vanishes with CQI = 13. In this case, there is less coding gain due to the higher modulation order and higher code rate. The potential performance gain given by the improved SINR in higher correlation scenarios cannot compensate the loss given by strong frequency selectivity in SCM-B channel, which has longer power delay profile than SCM-D.

Instead of using ideal channel knowledge, performances of receivers in Figure 4 are presented by using the real channel estimator. Besides that, the delay between the PMI selection and applying the corresponding precoding matrices at the transmitter is set to be 8 subframes instead of 1 subframe. Under these real working conditions both the IRC and the Max-log-MAP receivers have 2 to 3 dB performance degradation compared to the results in Figure 3. This degradation is caused by the channel estimation error and the mismatching between the reported optimal PMI and the real channel matrices when the precoding matrices are applied (similar effects caused by high UE velocity). However, the IRC receiver still has quite similar performance as the Max-Log-MAP receiver under those non-perfect working conditions. Therefore we conclude

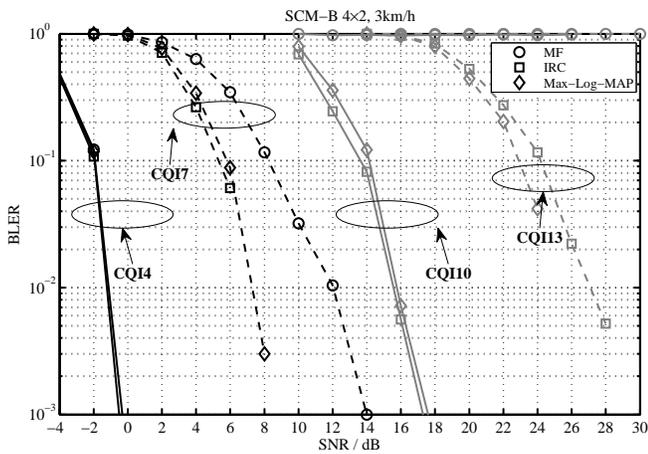


Figure 3. Performance in SCM-B 4x2 with ICK and 3 km/h UE Velocity

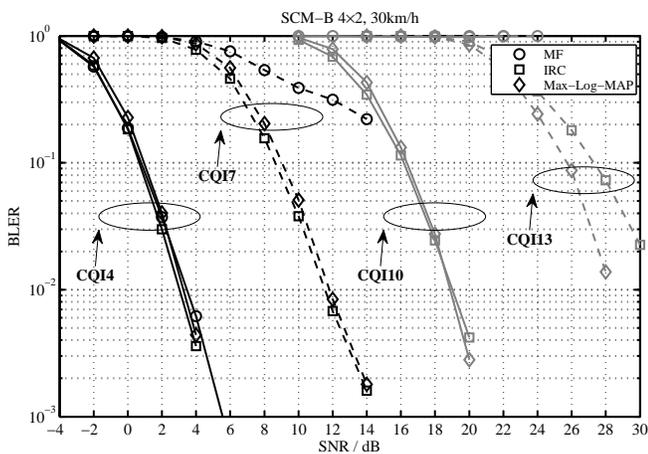


Figure 4. Performance in SCM-B 4x2 with 1DW and Feedback Delay of 8 Subframes

that the IRC receiver is robust for MU-MIMO transmission in LTE systems.

### V. CONCLUSION

In this paper, we presented the system performance in MU-MIMO transmission with different receivers in LTE systems. In addition, we presented the Max-SNR based simple PMI selection and the maximum Chordal distance based UE pairing criteria for MU-MIMO transmission. It has been shown that receivers with consideration of the co-channel interference, e.g. the IRC and Max-Log-MAP receivers, provide good performance in MU-MIMO transmission with perfect or non-perfect working conditions (channel estimation error and feedback delay), whereas the simple linear filter receiver faces large performance degradation or error floor issue. It has been demonstrated that a spatial correlated scenario helps the receivers to obtain the performance gain by eliminating the co-channel interference in MU-MIMO transmission.

### ACKNOWLEDGMENT

The authors wish to gratefully acknowledge the support of their colleagues at Intel Mobile Communications and at the Department of Communication Technologies of University of Duisburg-Essen. Parts of this work are carried out within the scope of the SAMURAI project which has been partially funded by the European Commission under FP7.

### REFERENCES

- [1] 3GPP36.201, "Evolved Universal Terrestrial Radio Access (E-UTRA); LTE physical layer; General description," 3GPP, Sophia Antipolis, Technical Specification 36.201 v8.3.0, Mar. 2009.
- [2] H. Yin and S. Alamouti, "OFDMA: A broadband wireless access technology," in *IEEE Sarnoff Symposium (SARNOFF)*, Mar. 2006, pp. 1-4.
- [3] H. G. Myung, J. Lim, and D. J. Goodman, "Single carrier FDMA for uplink wireless transmission," *IEEE Vehicular Technology Magazine*, vol. 1, no. 3, pp. 30-38, Sep. 2006.
- [4] C. Ribeiro, K. Hugl, M. Lampinen, and M. Kuehler, "Performance of linear multi-user MIMO precoding in LTE system," in *Proceedings of the 3rd International Symposium on Wireless Pervasive Computing (ISWPC 2008)*, Santorini, May 2008, pp. 410-414.
- [5] P. Frank, A. Müller, and J. Speidel, "Fair Performance Comparison between CQI- and CSI-based MU-MIMO for the LTE Downlink," in *Proceedings of European Wireless Conference*, Lucca, Apr. 2010.
- [6] 3GPP36.211, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation," 3GPP, Sophia Antipolis, Technical Specification 36.211 v8.9.0, Dec. 2009.
- [7] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 927-946, May 1998.
- [8] 3GPP36.213, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 3GPP, Sophia Antipolis, Technical Specification 36.213 v8.8.0, Sep. 2009.
- [9] 3GPP25.814, "Physical layer aspect for evolved Universal Terrestrial Radio Access (UTRA)," 3GPP, Sophia Antipolis, Technical Report 35.814 v7.1.0, Sep. 2006.
- [10] 3GPP25.996, "Spacial channel model for Multiple Input Multiple Output (MIMO) simulations," 3GPP, Sophia Antipolis, Technical Report 25.996 v8.0.0, Dec. 2008.
- [11] J. R. Barry, E. A. Lee, and D. G. Messerschmitt, *Digital Communication*, 3rd ed. New York: Springer, Sep. 2003.
- [12] M. Herdin, N. Czink, H. Ozelik, and E. Bonek, "Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels," in *Proceedings of IEEE 61st Vehicular Technology Conference (VTC 2005-Spring)*, Stockholm, May-Jun. 2005, pp. 136-140.
- [13] D. Love and R. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967-2976, Aug. 2005.
- [14] Z. Bai, J. Berkmann, C. Spiegel, T. Scholand, G. Bruck, C. Drewes, B. Gunzelmann, and P. Jung, "On MIMO with successive interference cancellation applied to UTRA LTE," in *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (ISCCSP 2008)*, Malta, Mar. 2008, pp. 1009-1013.
- [15] M. Escartin and P. Ranta, "Interference rejection with a small antenna array at the mobile scattering environment," in *Proceedings of the first IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Paris, Apr. 1997, pp. 165-168.
- [16] A. Viterbi, "An intuitive justification and a simplified implementation of the map decoder for convolutional codes," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 2, pp. 260-264, Feb. 1998.
- [17] C. Carbonelli and S. Franz, "Analytic performance evaluation of MIMO OFDM with pilot-aided channel estimation on doubly selective channels," *Telecommunications, European Transactions*, vol. 21, no. 7, pp. 589-602, 2010.

# A Hybrid Solution For Coverage Enhancement In Directional Sensor Networks

M. Amac Guvensan, A. Gokhan Yavuz

Department of Computer Engineering

Yildiz Technical University

Istanbul, Turkey

amac@ce.yildiz.edu.tr, gokhan@ce.yildiz.edu.tr

**Abstract**—In directional sensor networks (DSNs), motility capability of a directional sensor node has a considerable impact on the coverage enhancement. Motility capability may overcome overlapped field of views and occluded regions occurring during the initial deployment. On the other hand, adjusting working directions could not always heal coverage holes. Using mobility is the only solution under these circumstances. However, the high cost of mobile sensors and the high energy consumption of their physical movement are the two important constraints for their use. In this study, a hybrid solution has been proposed for the coverage improvement in DSNs. This hybrid solution increases the total coverage by up to 31% after the initial deployment. Besides, a new hybrid deployment model for DSNs has been discussed and the performance results of this model have been presented.

**Keywords**-Directional Sensor Networks; Coverage; Cost; Energy Efficiency; Motility; Mobility; Hybrid Deployment

## I. INTRODUCTION

The coverage problem in omni-directional sensor networks has been intensively studied in the past decade [1]. There are two main methods for coverage improvement in omni-directional sensor networks. *redployment/movement of sensor nodes*. On the other hand, coverage problems in directional sensor networks require more specific solutions since directional sensor nodes equipped with ultrasound, infrared and video sensors may work in several directions. Exploiting motility capability of those nodes is one of the basic solutions.

In a randomly deployed DSN, there might be overlapped areas and occluded regions after the initial deployment. A directional sensor node with motility capability could adjust its working direction along  $x$ ,  $y$ , and/or  $z$  axes. Thus, the node provides itself with a new field of view (FoV) without moving to a different location. This new direction possibly contributes to a better total coverage by (i) *minimizing overlapped areas* and (ii) *providing occlusion-free FoVs*. Available solutions to the coverage problem exploit motility rather than mobility due to its nominal cost and low energy consumption [2]. Since a directional sensor node maintains its geographical position while adjusting its FoV, the node does not require an additional driving mechanism and/or GPS device. This effectively decreases the total production cost of the sensor node. Nevertheless, mobility may heal coverage holes where motility is inadequate/insufficient.

However, mobility should be applied in a controlled manner because of its high energy consumption and the limited battery capacities of the sensor nodes.

To the best of our knowledge, existing solutions to the coverage problem in DSNs exploit either motility or mobility. In this paper, our main contribution is a distributed hybrid solution encouraging the use of both motility and mobility in a cascaded manner to improve the total coverage with minimum energy consumption. We discuss the idea of deploying hybrid directional sensor networks. Simulation results show that an optimum point for the deployment cost and the coverage improvement ratio could be achieved using a certain number of stationary, motile, and mobile nodes.

In Section II, existing solutions to the coverage problem in DSNs are discussed. Section III presents the directional sensing model and gives the details of the proposed hybrid solution. In Section III-B the idea of deploying hybrid sensor networks is addressed with respect to the coverage improvement and the cost of the network. Section IV presents the simulation results.

## II. RELATED WORK

Existing solutions to the coverage problem in DSNs are categorized into four groups [3]. We will only discuss the studies aiming at maximizing the whole coverage, as we focus on the improvement of area coverage. Enhancing area coverage is very important for DSNs to fulfill the specified sensing tasks. A small unmonitored sub-area defeats the whole purpose of the network. However, random deployment may cause several problems, such as overlapped and occluded regions, uncovered areas, and broken sensor nodes. Therefore, three solutions have been proposed by the research community to overcome these difficulties. First solution is to redeploy new sensors after the initial deployment. Second solution is to adjust the working directions of the directional sensor nodes to improve the field coverage [4] [5] [6] [7] [8] [9]. The last one is to relocate the sensor nodes with mobility capability [10].

The study [4] is one of the pioneer works on coverage enhancement. The authors present a new method based on a rotatable sensing model. To achieve less overlapping area, a directional node repositions itself on the reverse direction of the interior angle-bisector occurring between two neighboring

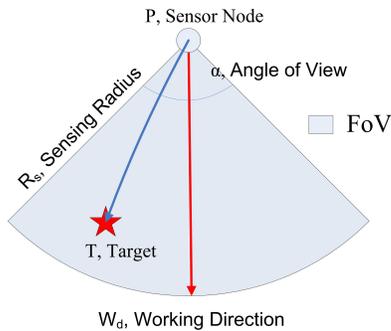


Figure 1. Directional Sensing Model

directional nodes. On the other hand, Cheng et al. describe the area-coverage enhancement problem as the Maximum Directional Area Coverage (MDAC) problem and prove the MDAC to be NP-complete [8]. The distributed solution for the MDAC problem, DGreedy algorithm, chooses the least overlapped direction as the new working direction. The authors observe that scarce sensors are highly critical to achieve maximal coverage, thus they utilize the number of sensing neighbors to differentiate node priorities.

Zhao and Zeng [5] propose an electrostatic field-based coverage-enhancing algorithm based on the Coulomb's Law to enhance the area coverage of wireless multimedia sensor networks by turning sensors to the correct orientation and decreasing the coverage overlaps of active sensors. They also aim at maximizing the network lifetime by shutting off as much redundant sensors as possible.

In [7], the authors name the above mentioned coverage problem as the optimal coverage problem in directional sensor networks (OCDSN). They propose a greedy approximation algorithm to the solution of the OCDSN problem, based on the boundary Voronoi diagram. By constructing the Voronoi diagram of a directional sensor network one could find the maximal breach path of this network. An assistant sensor traveling the edges of the Voronoi diagram determines which sensor to wake up in order to ensure the uncovered boundaries to be covered.

### III. A HYBRID SOLUTION TO THE COVERAGE PROBLEM IN DSNs

In this section, before presenting our hybrid solution, we will briefly explain the directional sensing model and the idea of hybrid deployment.

#### A. Directional Sensing Model

According to the binary detection model, a directional sensor node covers each point in its FoV. The common directional sensing capability for 2D spaces is illustrated in Figure 1. The sector covered by a directional sensor node  $S$  is denoted by a 4-tuple  $(P, R_s, \vec{W}_d, \alpha)$ , where  $P$

is the location,  $R_s$  is the sensing radius,  $\vec{W}_d$  is the working direction, and  $\alpha$  is the angle of view of the sensor node  $S$ .

Under ideal conditions without occlusion, a sensor node  $S$  covers an area with a size of  $\frac{\alpha}{2} R_s^2$  units. The special case of this model, where  $\alpha = 2\pi$  can be described as the omni-sensing model. For omni-directional sensors, there is only one possible working direction, whereas directional sensors have several possible working directions. However, they can work only at one direction at any given time  $t$ .

According to the binary detection model, a target point is covered when this point is located within a FoV of any directional sensor node. To find out as if this point is covered or not, the Target In Sector (TIS) test [11] needs to be applied to the related point. For area coverage problems, researchers opt for grid-based approach [12] to adapt this test model for indicating the (un)covered points in the observed area. Each point around the sensor node  $S$  is tested with the TIS test. The coverage map of the sensor node  $S$  is then created according to the test results.

#### B. Hybrid Deployment

We state that where the FoVs of two sensor nodes are overlapped, resolving overlapped region is still possible even only one of them has motility capability. Following this idea, we will elaborate on the effects of hybrid deployment in DSNs. There are three types of directional sensor nodes in DSNs. *stationary nodes*, *motile nodes*, and *mobile nodes*. Researchers have proved that motility and mobility significantly improve the total coverage in DSNs [3]. Nevertheless, networks consisting of motile/mobile directional sensor nodes require high budgets due to their considerable production cost. As an example, in [2], the cost of a stationary video camera, a pan-tilt-zoom video camera, and a mobile node are given as \$800, \$1300, and \$35000 respectively. The gap between the costs have definitely decreased and will continue to decrease in the future. However, there will always be a reasonable cost ratio between stationary, motile and mobile nodes. Thus, there is a necessity of hybrid directional sensor networks, where coverage improvement ratio and the cost of the network could be balanced. For example, the cost of a network built with mobile nodes is extremely high, whereas coverage improvement is impossible in DSNs consisting of only stationary nodes. Thus, we can assume that an optimum point for the cost and coverage improvement ratio could be achieved using a certain number of stationary, motile, and mobile nodes.

#### C. Cascaded Coverage Enhancement in DSNs

In randomly deployed directional sensor networks, the FoVs of two or more sensors might overlap and/or the FoVs of some sensors could be obscured by obstacles. In addition to these two problems, the working directions of some sensor nodes could be faced towards outside of the observed area.

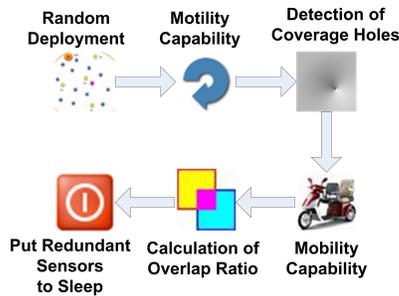


Figure 2. A block diagram of the proposed solution to the coverage problem in DSNs

As a consequence, the total coverage needs to be improved after the initial deployment. Most researchers exploit motility capability of directional sensor nodes for this purpose. To the best of our knowledge, only one study [10] proposes the use of motility. However, in this study the authors do not exploit motility capability. Both approaches have pros and cons. For example, it is impossible to heal some coverage holes with motile sensor nodes, where coverage holes are far enough from any sensor node. The only way (except redeployment) to resolve these holes is to benefit from the motility capability of the nodes. Nevertheless, motility is highly expensive because of the associated high production cost and high energy consumption. Therefore, after the initial deployment, we first exploit motility to minimize the overlapped areas and occluded regions. Then, we check for possible coverage holes. If there are redundant sensor nodes and coverage holes, we redirect these nodes to the holes. A block diagram of this hybrid solution is given in Figure 2.

The key idea of this hybrid solution is to utilize both the motility and mobility in a cascaded manner. For exploiting the motility, we have proposed the AFUP algorithm to adjust the working directions of the directional sensor nodes after the initial deployment. AFUP is a heuristic distributed algorithm which uses the repel forces of uncovered points around the nodes. Directional sensor nodes exchange their location ( $P$ ) and  $\vec{W}_d$  information with neighboring nodes. Then, each node marks the covered points in its map. Afterwards, using neighboring nodes' information, the nodes determine the overlapped regions in their map. If the number of overlapped points are more than a predefined threshold, the node sets its working direction as the center ( $C$ ) of the uncovered points in its map. Thus, the vector  $\vec{PC}$  gives the new working direction of the node. The aforementioned steps are repeated until each node reaches its equilibrium state. In some regions, where the node density is high, some nodes could not find an appropriate working direction. To account for such conditions, these nodes update their status as balanced. For exploiting mobility, we have designed a new Window-based Neighborhood Exploring (WNE) algorithm whose details will be discussed in a future work.

Table I  
AFUP ALGORITHM VS. RANDOM ALGORITHM

$R_s = 30m, \alpha = 60^\circ, Area = 250 \times 250m^2$			
Number of Sensor Nodes	Random Deployment(%)	AFUP(%)	Coverage Gain (%)
$N = 25$	15.65	18.48	18.08
$N = 50$	28.28	34.95	23.59
$N = 75$	39.69	49.28	24.16
$N = 100$	48.38	60.56	25.18
$N = 125$	56.28	69.80	24.02
$N = 150$	63.26	77.23	22.08

#### IV. PERFORMANCE EVALUATION

We have implemented a simulation environment using MATLAB 7.8. Several test scenarios have been run in this environment to show (i) the results of the AFUP algorithm, (ii) to find the node density where motility is inevitable, and (iii) to analyze the relation between the cost of the network and the coverage improvement ratio in hybrid directional sensor networks. We consider the total coverage and the overlap ratio as the two key metrics. The values for node density, sensing radius, and angle of view have been chosen with regard to the studies [13] and [5]. Accordingly, sensor nodes have been configured with an  $\alpha = 60^\circ$  and a  $R_s = 30m$ . Simulations have been performed for random deployment in a rectangular two-dimensional terrain of  $250 \times 250m^2$ . 15 different uniform random distributions have been generated for each individual scenario.

Simulation results show that motility significantly improves the total coverage after the initial deployment. Using only the local information of neighboring nodes, both overlapped areas are minimized and nodes near border are faced towards the observed area. Analyzing Table I shows that coverage gain varies from 18% to 25%. Especially, the coverage gain in dense networks is greater than in sparse networks. However, coverage gain starts to drop down when the network is saturated with sensor nodes.

Figure 3 shows the ratios of overlap minimization and coverage gain for different number of sensor nodes. Overlap minimization ratio shows by how much ratio the overlapped points are decreased after the initial deployment, whereas coverage gain indicates the rational increase of the total coverage after applying AFUP. As there are too many uncovered points in DSNs with low node density, directional sensor nodes could easily find appropriate working directions. Therefore, the overlap minimization ratio is considerably high in those networks. With increasing node density, the overlap minimization ratio starts to decrease. Facing the border nodes towards the observed area with increasing node density prevents the decrease of the overlapped areas. Thus, the overlapped areas might slightly increase above a certain threshold. This fact reveals that the use of motility is inevitable for directional sensor networks with high node densities. Another course of action would be to put the redundant nodes into sleep in the absence of mobile nodes.

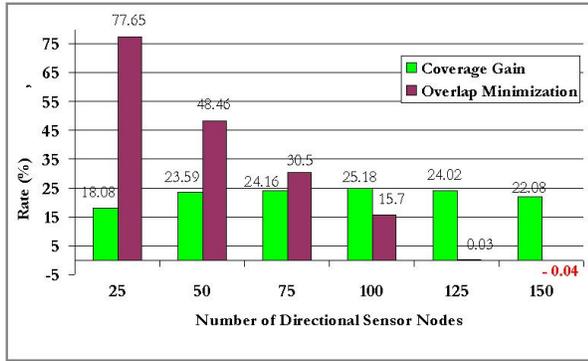


Figure 3. The relationship between the coverage gain ratio, the overlap minimization ratio and the number of sensor nodes

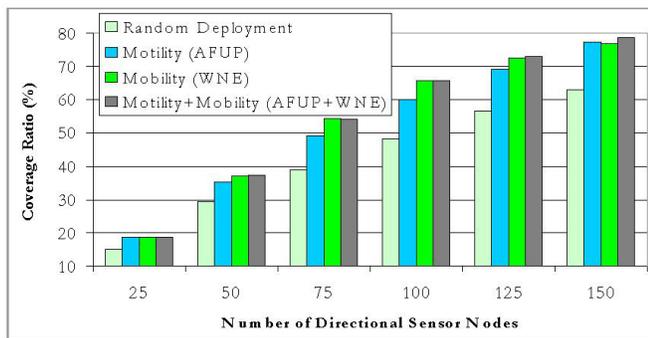


Figure 4. Coverage ratios of random deployment, AFUP, WNE, and AFUP+WNE

To explore the effect of mobility, we have utilized our WNE algorithm in the previously given scenarios. The preliminary results show that exploiting mobility after motility could increase the total coverage up to 6%. In the scenarios directional sensor nodes were assumed to exchange their location information only with nodes within their communication radius ( $R_c = 2R_s$ ). Thus, this resulted in exploring coverage holes only within three sensing radii and therefore, mobility has a limited impact on the coverage gain. Exchanging information with nodes located within two-hop communication distance would definitely improve the coverage gain.

Solving the coverage problem only with mobility is expensive, whereas motility is inadequate to cope with coverage holes. Our experimental results, given in Figure 4, show that motility+mobility in DSNs provides a substantial coverage improvement. Furthermore, according to Table II

Table II  
NUMBER OF MOVED DIRECTIONAL SENSOR NODES. MOBILE VS. MOTILE/MOBILE SENSOR NETWORK

Total Number of Nodes	25	50	75	100	125	150
Mobility	9	23	40	44	41	33
Motility+Mobility	1	8	19	17	10	4

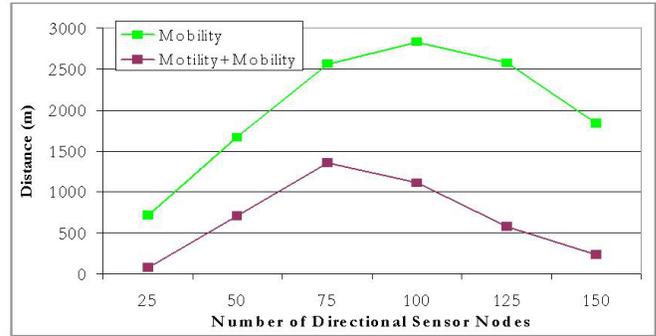


Figure 5. Total travel distance of directional sensor nodes for mobile and motile/mobile DSNs.

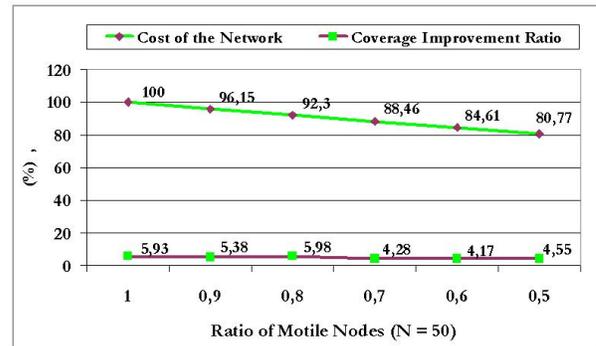


Figure 6. The correlation between the coverage improvement and the cost of the network in hybrid sensor networks

and Figure 5 motility+mobility is also highly energy efficient compared to a mobility only solution. A mobility alone needs 2 to 9 times more sensors to change their physical location to achieve same coverage gain as in a motility+mobility solution. Moreover, the total travel distance in motile/mobile DSNs is substantially less than the total travel distance in mobile DSNs, as shown in Fig. 5

Figure 6 demonstrates the importance of hybrid deployment. Following Figure 6, even if a large number of deployed directional sensor nodes are stationary, the coverage gain ratio does not change too much if all the nodes were motile. As an example, in the scenario where the ratio of motile nodes is 60%, the total coverage ratio increases by 4.17%. Deploying the same scenario with all the sensor nodes being motile, causes the total coverage to increase by only 5.93%. Thus, with hybrid deployment, by acknowledging a coverage gain drop of 1%-2%, the deployment cost could be reduced by 20%.

### V. CONCLUSION

In this study, we have presented the preliminary results of a Ph.D. thesis, which examines the effect of motility and mobility capabilities of the directional sensor nodes on the coverage improvement and the cost of the network. The proposed hybrid solution aims at maximizing the total

coverage with minimum energy consumption after the initial deployment. Simulation results also show that deploying hybrid DSNs could balance the ratio between the deployment cost of network and the total coverage improvement.

## VI. ACKNOWLEDGEMENT

This research has been supported by Yildiz Technical University Scientific Research Projects Coordination Department under the grant number 2011-04-01-DOP03.

## REFERENCES

- [1] Y. Mohamed and K. Akkaya, "Strategies and techniques for node placement in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 6, no. 4, pp. 621–655, 2008.
- [2] A. Kansal, W. Kaiser, G. Pottie, M. Srivastava, and G. Sukhatme, "Reconfiguration methods for mobile sensor networks," *ACM Transactions on Sensor Networks*, vol. 3, no. 4, 2007.
- [3] M. A. Guvensan and A. G. Yavuz, "On coverage issues in directional sensor networks: A survey," *Elsevier Ad Hoc Networks*, <http://dx.doi.org/10.1016/j.adhoc.2011.02.003>, February 2011.
- [4] D. Tao, H. Ma, and L. Liu, "Coverage-enhancing algorithm for directional sensor networks," in *Lecture Notes in Computer Science: Mobile Ad-hoc and Sensor Networks*, vol. 4325, November 2006, pp. 256–267.
- [5] J. Zhao and J.-C. Zeng, "An electrostatic field-based coverage-enhancing algorithm for wireless multimedia sensor networks," in *Proc. of IEEE Intl. Conf. on Wireless Communications, Networking and Mobile Computing (WiCom'09)*, Beijing, China, September 2009, pp. 1–5.
- [6] H. Ma, X. Zhang, and A. Ming, "A coverage-enhancing method for 3d directional sensor networks," in *Proc. of IEEE Intl. Conf. on Computer Communications (INFOCOM'09)*, Rio de Janeiro, Brazil, April 2009, pp. 2791–2795.
- [7] J. Li, R.-C. Wang, H.-P. Huang, and L.-J. Sun, "Voronoi based area coverage optimization for directional sensor networks," in *Proc. of Intl. Symposium on Electronic Commerce and Security (ISECS'09)*, vol. 1, Nanchang, China, May 2009, pp. 488–493.
- [8] W. Cheng, S. Li, X. Liao, S. Changxiang, and H. Chen, "Maximal coverage scheduling in randomly deployed directional sensor networks," in *Proc. of Intl. Conf. on Parallel Processing Workshops (ICPPW'07)*, China, October 2007.
- [9] C. Kandoth and S. Chellappan, "Angular mobility assisted coverage in directional sensor networks," in *Proc. of Intl. Conf. on Network-Based Information Systems (NBIS '09)*, Aug. 2009, pp. 376–379.
- [10] C.-K. Liang, M.-C. He, and C.-H. Tsai, "Movement assisted sensor deployment in directional sensor networks," in *Proc. of Intl. Conf. on Mobile Ad-Hoc and Sensor Networks*, Dec. 2010.
- [11] J. Ai and A. A. Abouzeid, "Coverage by directional sensors in randomly deployed wireless sensor networks," *Journal of Combinatorial Optimization*, vol. 11, no. 1, pp. 21–41, February 2006.
- [12] H. Chen, H. Wu, and N.-F. Tzeng, "Grid-based approach for working node selection in wireless sensor networks," in *Proc. Intl. Conf on Communications*, vol. 6, Paris, France, 20-24 2004, pp. 3673 – 3678.
- [13] N. Tezcan and W. Wang, "Self-orienting wireless multimedia sensor networks for occlusion-free viewpoints," *Computer Networks: Int. Journal of Computer and Telecommunications Networking*, vol. 52, no. 13, pp. 2558–2567, September 2008.

# Amplifying Matrix Design for Cooperative Relay Networks under Channel Uncertainty and Power Constraint

Kanghee Lee, Hyuck M. Kwon, Yanwu Ding, Zuojun Wang, Yu Bi, and Yazan Ibdah  
 Department of Electrical Engineering and Computer Science, Wichita State University, Wichita, Kansas 67260  
 Emails: {kxlee1, hyuck.kwon, yanwu.ding, zxwang3, yxbi, yxibdah}@wichita.edu

**Abstract**—This paper presents an optimum amplifying relay matrix based on minimum mean square error (MMSE) criteria for the cooperative distributed wireless relay networks under channel uncertainty and total power constraint at the relay nodes. A one-source-one-destination pair node and  $N$  relay nodes are considered, which can be extended for a general case. All relays are assumed to exchange their received signals with no errors for full cooperation. With the derived optimum amplifying relay matrix, bit error rate (BER) of the relay network under channel uncertainty is simulated and compared with that of the network under a certain channel condition to observe the degradation.

**Index Terms**—AF, MMSE, channel-uncertainty, power-constraint.

## I. INTRODUCTION

Due to the restricted power, relay nodes in wireless networks can only retransmit their received signal from a source node within a limited distance [1]. To overcome this limitation, various protocols were proposed. Relay protocols can be classified as amplify-and-forward (AF), decode-and-forward (DF), and compress-and-forward (CF) [2]–[8]. The AF relay protocol only retransmits a scaled version of their received signals from the source node according to their power constraint. Therefore, AF relay protocol employed in this current paper is a reasonable strategy when relay nodes have limited power. Compared to the DF and CF relay protocols, AF relay protocol results in a lower complexity. This is due to the fact it does not require signal processing at the relay for decoding and compressing. In addition, AF relay protocol shows a shorter delay at the relay node. This is also due to the same fact unlike the DF and CF protocols.

In practice, due to wireless communication characteristics during information transmission, all relay nodes can have an inaccurate knowledge of their local channels, either from the source node to relay nodes or from relay nodes to the destination node.

This work was partly sponsored by the Army Research Office under DEPSCoR ARO Grant W911NF-08-1-0256, and by NASA under EPSCoR CAN Grant NNX08AV84A.

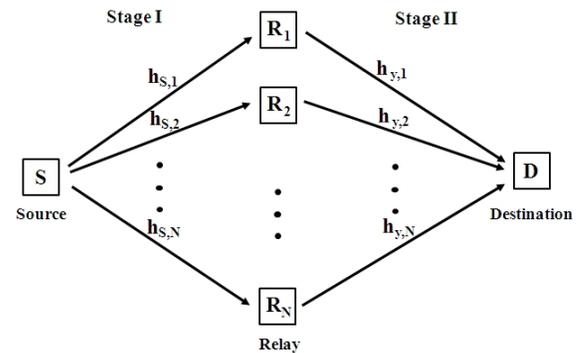


Fig. 1. Wireless relay network with one-source-one-destination pair and  $N$  relay nodes under no channel uncertainty ( $\mathbf{h}_s$ ,  $\mathbf{h}_y$ ).

ation node. In [5], the DF relay protocol with single-input single-output was considered for a multibranch cooperative distributed wireless relay network with direct link between the source and the destination under no channel uncertainty. While, in [6], a multibranch noncooperative AF relay network without direct link between the source and the destination was considered for single-input single-output with  $N$  relays under channel uncertainty. AF relay protocol in [6] did not employ cooperation among distributed relay nodes even though authors considered channel uncertainty. Furthermore, total power constraint at relay nodes was not included in [6]. In [7], [8], relay nodes exchanged their received signals for cooperation, and channels were assumed to be perfectly known at each relay node. While, in [9], [10], total power constraint at the destination node instead of relay nodes was considered for noncooperative AF wireless relay networks under certain channel condition to minimize mean square error. In addition, channel uncertainty for noncooperative AF wireless relay networks under the received power constraint at the destination node was studied in [11]. Namely, recently, cooperative AF wireless relay networks under channel uncertainty with power constraint at relay nodes was not considered. Therefore, the objective of this current paper is to find an optimum relay amplifying matrix for the cooperative distributed wireless

relay networks under channel uncertainty and total power constraint at relay nodes. The minimum mean square error (MMSE) criterion is used to optimize the relay amplifying matrix. Since cooperation is employed, the optimum relay amplifying matrix can be non-diagonal. On the other hand, it will be a diagonal matrix if a non-cooperative relay network is required as in [6]. The diagonal relay amplifying matrix under total power constraint at the destination node is designed for the case where the direct link between a source and a destination is considered weak and can be negligible. Using the derived optimum relay amplifying matrix, effects of three different kinds of channel uncertainty both cooperative case and noncooperative case on BER was studied.

The remaining paper is organized into four sections. Section II describes the system model and information transmission scheme. Section III presents a cooperative distributed MMSE relay scheme based on the AF strategy. Three different imperfect channel information cases for both cooperative and noncooperative distributed wireless networks are considered. Section IV presents simulation results. Finally, Section V concludes the paper.

**Notation:** Matrices and vectors are denoted, respectively, by uppercase and lowercase boldface characters (e.g.,  $\mathbf{A}$  and  $\mathbf{a}$ ). The transpose, complex conjugate, inverse, and trace of  $\mathbf{A}$  are denoted, respectively, by  $\mathbf{A}^T$ ,  $\mathbf{A}^*$ ,  $\mathbf{A}^{-1}$ , and  $\text{tr}(\mathbf{A})$ . The Hermitian of  $\mathbf{A}$  is denoted by  $\mathbf{A}^H$ , i.e., the complex conjugate and transpose of  $\mathbf{A}$ . An  $n \times n$  identity matrix is denoted by  $\mathbf{I}_n$ .  $\mathbf{A} \in \mathbf{C}^{m \times n}$  is an  $m \times n$  complex matrix. The expectation operator is  $E[\cdot]$ . Notations  $|a|$ ,  $\|\mathbf{a}\|$ , and  $\|\mathbf{A}\|_F$  denote the absolute value of  $a$  for any scalar, 2-norm of  $\mathbf{a}$ , and Frobenius-norm of  $\mathbf{A}$ , respectively. An  $n \times n$  diagonal matrix is denoted by  $\text{diag}(a_1, \dots, a_n)$ . An  $\arg \min_{\mathbf{A}} J(\mathbf{A})$  denotes the value of  $\mathbf{A}$  at which  $J(\mathbf{A})$  is minimized.

## II. SYSTEM MODEL

Figure 1 shows a wireless relay network with  $N$  cooperative distributed relay nodes between a source node and a destination node. As shown in Fig. 1, there are two stages for information transmission where a source node transmits a signal symbol  $s$  in Stage I and the relay nodes retransmit their received signals to a destination node in Stage II using the amplifying matrix. Certain channel and uncertain channel conditions are described in the remaining of this section.

### A. Certain Channel Condition

In a certain channel condition, all the relay nodes have perfect information of their local channel coefficients from the source node to the relay nodes and from the relay nodes to the destination node. Let  $\mathbf{h}_s \in \mathbf{C}^{N \times 1}$  denote the perfect

channel coefficient complex column vector from the source node to the relay nodes as

$$\mathbf{h}_s = [h_{s,1}, h_{s,2}, \dots, h_{s,N}]^T$$

where  $h_{s,i}$ ,  $i = 1, \dots, N$ , is the  $i$ -th element of  $\mathbf{h}_s$ , representing the channel coefficient from the source node to the  $i$ -th relay node. The received signal complex column vector  $\mathbf{r} \in \mathbf{C}^{N \times 1}$  at the relay nodes is written as

$$\mathbf{r} = \mathbf{h}_s s + \mathbf{v}_s \quad (1)$$

where  $\mathbf{v}_s \in \mathbf{C}^{N \times 1}$  is a zero mean complex additive white Gaussian noise column vector with covariance matrix  $\sigma_{v_s}^2 \mathbf{I}_N$ . Each channel  $h_{s,i}$  is assumed to be independent identically distributed with a zero mean circular complex Gaussian of unit variance and quasi-static Rayleigh fading so that they stay fixed during an observation interval. For the cooperation of relay nodes, it is assumed that all the relay nodes exchange their received signals each other with negligible errors. The amplified signal complex column vector  $\mathbf{x} \in \mathbf{C}^{N \times 1}$  at the relay nodes is given by

$$\mathbf{x} = \mathbf{F} \mathbf{r} \quad (2)$$

where  $\mathbf{F} \in \mathbf{C}^{N \times N}$  is an optimum relay amplifying matrix at the relay nodes. Let  $\mathbf{h}_y \in \mathbf{C}^{1 \times N}$  denote the perfect channel coefficient complex row vector from the relay nodes to the destination node as

$$\mathbf{h}_y = [h_{y,1}, h_{y,2}, \dots, h_{y,N}]$$

where  $h_{y,i}$ ,  $i = 1, \dots, N$ , is the  $i$ -th element of  $\mathbf{h}_y$ , representing the channel coefficient from the  $i$ -th relay node to the destination node. Each channel  $h_{y,i}$  is also assumed to be independent identically distributed with a zero mean circular complex Gaussian of unit variance and quasi-static Rayleigh fading. The received complex signal  $y \in \mathbf{C}^{1 \times 1}$  at the destination node can be represented as

$$y = \mathbf{h}_y \mathbf{x} + v_y \quad (3)$$

where  $v_y \in \mathbf{C}^{1 \times 1}$  is a zero-mean complex additive white Gaussian noise variable with variance  $\sigma_{v_y}^2$ . Substituting (1) and (2) into (3), the received complex signal  $y \in \mathbf{C}^{1 \times 1}$  at the destination node can be written as

$$y = \mathbf{h}_y \mathbf{F} \mathbf{h}_s s + \mathbf{h}_y \mathbf{F} \mathbf{v}_s + v_y. \quad (4)$$

### B. Uncertain Channel Condition

In reality, due to channel estimation errors, the estimates of the imperfect channel coefficient complex vectors  $\bar{\mathbf{h}}_s$  and  $\bar{\mathbf{h}}_y$  should be used instead of perfect  $\mathbf{h}_s$  and  $\mathbf{h}_y$ , respectively, under uncertain channel conditions.

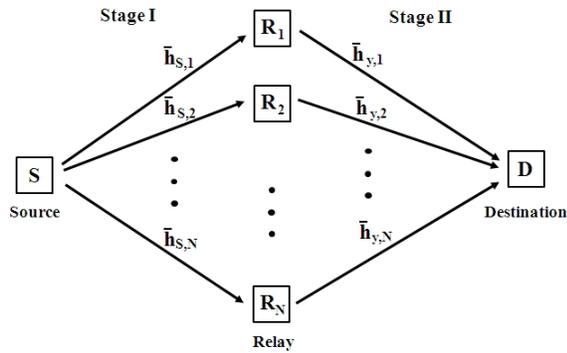


Fig. 2. Wireless relay network with one-source-one-destination pair and  $N$  relay nodes under channel uncertainty for both source-relay links and relay-destination links ( $\bar{\mathbf{h}}_s$ ,  $\bar{\mathbf{h}}_y$ ).

As shown Fig. 2, the estimated channel coefficient vectors can be written, respectively, as

$$\bar{\mathbf{h}}_s = \mathbf{h}_s - \phi_s \quad \text{or} \quad \bar{\mathbf{h}}_s = \mathbf{h}_s + \phi_s \quad (5)$$

and

$$\bar{\mathbf{h}}_y = \mathbf{h}_y - \phi_y \quad \text{or} \quad \bar{\mathbf{h}}_y = \mathbf{h}_y + \phi_y \quad (6)$$

where  $\phi_s \in \mathbf{C}^{N \times 1}$  and  $\phi_y \in \mathbf{C}^{1 \times N}$  are, respectively, the channel estimation error column and row vectors consisting of complex independent identically distributed zero mean Gaussian random variables with covariance matrix  $\sigma_{\phi_s}^2 \mathbf{I}_N$  and  $\sigma_{\phi_y}^2 \mathbf{I}_N$ . The estimated channel coefficient vectors are applied for both source-relay links and relay-destination links. Therefore, the received signal complex column vector  $\mathbf{r}_1 \in \mathbf{C}^{N \times 1}$  at the relay nodes is represented as

$$\mathbf{r}_1 = \bar{\mathbf{h}}_s s + \phi_s s + \mathbf{v}_s \quad (7)$$

where  $\mathbf{v}_s \in \mathbf{C}^{N \times 1}$  is a zero-mean complex additive white Gaussian noise vector with covariance matrix  $\sigma_{v_s}^2 \mathbf{I}_N$ . The subscript 1 denotes case 1 where  $\bar{\mathbf{h}}_s$  and  $\bar{\mathbf{h}}_y$  are used. The processed signal complex column vector  $\mathbf{x}_1 \in \mathbf{C}^{N \times 1}$  at the relay node outputs is written as

$$\mathbf{x}_1 = \mathbf{F}_1 \mathbf{r}_1 \quad (8)$$

where  $\mathbf{F}_1 \in \mathbf{C}^{N \times N}$  is a relay amplifying matrix employed by the relay nodes to improve performance at the destination node. The received complex signal  $y_1 \in \mathbf{C}^{1 \times 1}$  at the destination node can be written as

$$y_1 = \mathbf{h}_y \mathbf{x}_1 + v_y = \bar{\mathbf{h}}_y \mathbf{x}_1 + \phi_y \mathbf{x}_1 + v_y \quad (9)$$

where  $v_y \in \mathbf{C}^{1 \times 1}$  is a zero mean complex additive white Gaussian noise with variance  $\sigma_{v_y}^2$ . Substituting (7) into (8) and using (9), the received complex signal  $y_1 \in \mathbf{C}^{1 \times 1}$  at the destination node can be written as

$$y_1 = \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s s + \phi_y \mathbf{F}_1 \bar{\mathbf{h}}_s s + \bar{\mathbf{h}}_y \mathbf{F}_1 \phi_s s + \phi_y \mathbf{F}_1 \phi_s$$

$$+ \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{v}_s + \phi_y \mathbf{F}_1 \mathbf{v}_s + v_y. \quad (10)$$

In the next section, the optimum relay amplifying matrix  $\mathbf{F}$  will be determined using the MMSE criteria for the channel uncertainty conditions. Note that in [6], Khajehnouri et. al. considered a noncooperative distributed wireless relay network. Hence, the relay amplifying matrix was a diagonal matrix in [6]. However, the relay amplifying matrix in this current paper would be non-diagonal because cooperation is made among the relay nodes by exchanging their received signals.

### III. COOPERATIVE DISTRIBUTED MMSE RELAY SCHEME UNDER CHANNEL UNCERTAINTY

The MMSE will try to minimize the mean square error between the signal component at the destination node and the transmitted signal  $s$  at the source node. Total power constraint at the relay nodes will be included in the minimization. Assume both channel estimation vectors for both source-relay and relay-source links are not perfect, i.e., ( $\bar{\mathbf{h}}_s$ ,  $\bar{\mathbf{h}}_y$ ). The optimum relay amplifying matrix to minimize mean square error between the signal component at the destination node and the transmitted signal at the source node in (10) can be found from

$$\begin{aligned} \mathbf{F}_1^\dagger &= \arg \min_{\mathbf{F}_1} J(\mathbf{F}_1) \\ \text{s.t.} \quad &E[\|\mathbf{x}_1\|_2^2] = P_1 \end{aligned} \quad (11)$$

under the relay power constraint where the objective function  $J(\mathbf{F}_1)$  is written using (10) as

$$\begin{aligned} J(\mathbf{F}_1) &= E[|\mathbf{h}_y \mathbf{x}_1 - s|^2] = E[|\bar{\mathbf{h}}_y \mathbf{x}_1 + \phi_y \mathbf{x}_1 - s|^2] \\ &= E[|\bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s s + \phi_y \mathbf{F}_1 \bar{\mathbf{h}}_s s + \bar{\mathbf{h}}_y \mathbf{F}_1 \phi_s s + \phi_y \mathbf{F}_1 \phi_s s \\ &\quad + \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{v}_s + \phi_y \mathbf{F}_1 \mathbf{v}_s - s|^2] \\ &= \text{tr}(\sigma_s^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \mathbf{F}_1^H \bar{\mathbf{h}}_y^H + \sigma_s^2 \sigma_{\phi_s}^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \mathbf{F}_1^H \\ &\quad + \sigma_s^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \mathbf{F}_1^H + \sigma_s^2 \sigma_{\phi_s}^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{F}_1^H \bar{\mathbf{h}}_y^H \\ &\quad - \sigma_s^2 \bar{\mathbf{h}}_s^H \mathbf{F}_1^H \bar{\mathbf{h}}_y^H + \sigma_{v_s}^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{F}_1^H \bar{\mathbf{h}}_y^H \\ &\quad + \sigma_{v_s}^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \mathbf{F}_1^H - \sigma_s^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s) + \sigma_s^2 \end{aligned} \quad (12)$$

where  $E[\phi_y] = E[\phi_s] = \mathbf{0}$  is used in (12). The total power constraint at the relay nodes can be represented from (8) as

$$P_1 = E[\|\mathbf{x}_1\|_2^2] = \sigma_s^2 \|\mathbf{F}_1 \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\mathbf{F}_1\|_F^2. \quad (13)$$

Because the total power at the relay nodes is constrained  $P_1$ , the constrained optimization in [12] is to be performed as

$$\begin{aligned} L(\mathbf{F}_1, \lambda_1) &= J(\mathbf{F}_1) + \lambda_1 (E[\|\mathbf{x}_1\|_2^2] - P_1) \\ &= \text{tr}(\sigma_s^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \mathbf{F}_1^H \bar{\mathbf{h}}_y^H + \sigma_s^2 \sigma_{\phi_s}^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \mathbf{F}_1^H \\ &\quad + \sigma_s^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \mathbf{F}_1^H + \sigma_s^2 \sigma_{\phi_s}^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{F}_1^H \bar{\mathbf{h}}_y^H \end{aligned}$$

$$\begin{aligned}
 & -\sigma_s^2 \bar{\mathbf{h}}_s^H \mathbf{F}_1^H \bar{\mathbf{h}}_y^H + \sigma_{v_s}^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \mathbf{F}_1^H \bar{\mathbf{h}}_y^H \\
 & + \sigma_{v_s}^2 \sigma_{\phi_y}^2 \mathbf{F}_1 \mathbf{F}_1^H - \sigma_s^2 \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s + \sigma_s^2 \\
 & + \lambda_1 (\sigma_s^2 \|\mathbf{F}_1 \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\mathbf{F}_1\|_F^2 - P_1)
 \end{aligned} \quad (14)$$

where  $\lambda_1$  is a Lagrangian multiplier. Take the derivative of (14) in terms of complex conjugate of  $\mathbf{F}_1$ , i.e.,  $\mathbf{F}_1^*$ , and  $\lambda_1$ , respectively, using the properties of the derivative matrices [13], [14]. Then,

$$\begin{aligned}
 \frac{\partial L(\mathbf{F}_1, \lambda_1)}{\partial \mathbf{F}_1^*} &= \bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_y \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \sigma_s^2 + \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \sigma_s^2 \sigma_{\phi_y}^2 + \mathbf{F}_1 \sigma_s^2 \sigma_{\phi_y}^2 \\
 & + \bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_y \mathbf{F}_1 \sigma_{v_s}^2 - \bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_s \sigma_s^2 + \mathbf{F}_1 \sigma_s^2 \sigma_{\phi_s}^2 \sigma_{\phi_y}^2 \\
 & + \bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_y \mathbf{F}_1 \sigma_s^2 \sigma_{\phi_s}^2 + \lambda_1 \mathbf{F}_1 \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \sigma_s^2 \\
 & + \lambda_1 \mathbf{F}_1 \sigma_s^2 \sigma_{\phi_s}^2 + \lambda_1 \mathbf{F}_1 \sigma_{v_s}^2 = \mathbf{0}
 \end{aligned} \quad (15)$$

and

$$\frac{\partial L(\mathbf{F}_1, \lambda_1)}{\partial \lambda_1} = \sigma_s^2 \|\mathbf{F}_1 \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\mathbf{F}_1\|_F^2 - P_1 = 0. \quad (16)$$

Therefore, the optimal amplifying relay matrix  $\mathbf{F}_1^\dagger$  can be written as

$$\begin{aligned}
 \mathbf{F}_1^\dagger &= \sigma_s^2 (\mathbf{h}_y^H \mathbf{h}_y + \sigma_{\phi_y}^2 \mathbf{I}_N + \lambda_1 \mathbf{I}_N)^{-1} (\mathbf{h}_y^H \bar{\mathbf{h}}_s^H) \times \\
 & (\bar{\mathbf{h}}_s \bar{\mathbf{h}}_s^H \sigma_s^2 + \sigma_s^2 \sigma_{\phi_s}^2 \mathbf{I}_N + \sigma_{v_s}^2 \mathbf{I}_N)^{-1}.
 \end{aligned} \quad (17)$$

After applying the matrix inversion lemma [15]  $(\mathbf{A} + \mathbf{BTD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1} \mathbf{B} (\mathbf{T}^{-1} + \mathbf{D} \mathbf{A}^{-1} \mathbf{B})^{-1} \mathbf{D} \mathbf{A}^{-1}$ , the optimal amplifying relay matrix  $\mathbf{F}_1^\dagger$  can be written as

$$\mathbf{F}_1^\dagger = \frac{\sigma_s^2 \bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_s^H}{\left( \lambda_1 + \sigma_{\phi_y}^2 + \|\bar{\mathbf{h}}_y\|_2^2 \right) \left( \sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2 + \sigma_s^2 \|\bar{\mathbf{h}}_s\|_2^2 \right)}. \quad (18)$$

Substituting (18) into (13), the total power  $P_1$  at the relay nodes can be written as

$$P_1 = \frac{\sigma_s^6 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + \sigma_s^4 (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2}{\left[ \left( \lambda_1 + \sigma_{\phi_y}^2 + \|\bar{\mathbf{h}}_y\|_2^2 \right) \left( \sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2 + \sigma_s^2 \|\bar{\mathbf{h}}_s\|_2^2 \right) \right]^2}. \quad (19)$$

Using (19), the Lagrangian multiplier  $\lambda_1$  can be written as

$$\begin{aligned}
 \lambda_1 &= \frac{\pm \sigma_s^2 \sqrt{P_1^{-1} (\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2)}}{\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2 + \sigma_s^2 \|\bar{\mathbf{h}}_s\|_2^2} \\
 & - \left( \|\bar{\mathbf{h}}_y\|_2^2 + \sigma_{\phi_y}^2 \right).
 \end{aligned} \quad (20)$$

The same BER performance is accomplished regardless of the sign of  $\lambda_1$ . Hence, only the positive sign of  $\lambda_1$  is considered, i.e.,  $\bar{\omega}$ , where  $\bar{\omega} = \max(0, \omega)$ . Substituting (20) into (18), the optimum relay amplifying matrix  $\mathbf{F}_1^\dagger$  can be written as

$$\mathbf{F}_1^\dagger = \frac{\bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_s^H \sqrt{P_1}}{\sqrt{\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2}}. \quad (21)$$

Note that  $\mathbf{F}_1^\dagger$  does not depend on  $\sigma_{\phi_y}^2$ , but  $\bar{\mathbf{h}}_y$  does. The MMSE can be computed from  $J(\mathbf{F}_1)$  in (14) by replacing  $\mathbf{F}_1$  with  $\mathbf{F}_1^\dagger$ .

From (14) and (17), the constrained optimal solutions  $\bar{\lambda}_i$  and  $\mathbf{F}_i^\dagger$ ,  $i = 2, 3$ , of other two special cases under channel uncertainty and power constraints at the relay nodes are derived, respectively, as

$$\bar{\lambda}_2 = \frac{\sigma_s^2 \sqrt{P_2^{-1} (\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2)}}{\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2 + \sigma_s^2 \|\bar{\mathbf{h}}_s\|_2^2} - \|\bar{\mathbf{h}}_y\|_2^2 \quad (22)$$

$$\mathbf{F}_2^\dagger = \frac{\bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_s^H \sqrt{P_2}}{\sqrt{\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + (\sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2) \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2}} \quad (23)$$

for only source-relay links when  $\sigma_{\phi_y}^2 = 0$ , and

$$\begin{aligned}
 \bar{\lambda}_3 &= \frac{\sigma_s^2 \sqrt{P_3^{-1} (\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + \sigma_{v_s}^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2)}}{\sigma_{v_s}^2 + \sigma_s^2 \|\bar{\mathbf{h}}_s\|_2^2} \\
 & - \left( \|\bar{\mathbf{h}}_y\|_2^2 + \sigma_{\phi_y}^2 \right)
 \end{aligned} \quad (24)$$

$$\mathbf{F}_3^\dagger = \frac{\bar{\mathbf{h}}_y^H \bar{\mathbf{h}}_s^H \sqrt{P_3}}{\sqrt{\sigma_s^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s \bar{\mathbf{h}}_s\|_2^2 + \sigma_{v_s}^2 \|\bar{\mathbf{h}}_y \bar{\mathbf{h}}_s\|_F^2}} \quad (25)$$

for only relay-destination links when  $\sigma_{\phi_s}^2 = 0$ .

For performance comparisons, the diagonal relay amplifying matrices under total power constraint at relay nodes are derived in the rest of this section. Hence, the diagonal relay amplifying matrix  $\mathbf{F}_4^\dagger$  corresponding to both imperfect channel estimation vectors  $\bar{\mathbf{h}}_s^H$  and  $\bar{\mathbf{h}}_y^H$  in noncooperative wireless networks can be written as

$$\mathbf{F}_4^\dagger = \text{diag}(f_1, f_2, \dots, f_N), \quad i = 1, \dots, N \quad (26)$$

where

$$f_i = \frac{\bar{h}_{y,i}^* \bar{h}_{s,i}^* \sqrt{P_4}}{\sqrt{\sum_{i=1}^N |\bar{h}_{y,i}|^2 |\bar{h}_{s,i}|^2 (\sigma_s^2 |\bar{h}_{s,i}|^2 + \sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2)}}. \quad (27)$$

Similarly, the diagonal relay amplifying matrices  $\mathbf{F}_5^\dagger$  and  $\mathbf{F}_6^\dagger$  corresponding to only imperfect channel estimation column vector  $\bar{\mathbf{h}}_s^H$  and only imperfect channel estimation row vector  $\bar{\mathbf{h}}_y^H$  in noncooperative wireless networks can be written, respectively, as

$$\mathbf{F}_5^\dagger = \text{diag}(f_1, f_2, \dots, f_N), \quad i = 1, \dots, N \quad (28)$$

where

$$f_i = \frac{h_{y,i}^* \bar{h}_{s,i}^* \sqrt{P_5}}{\sqrt{\sum_{i=1}^N |h_{y,i}|^2 |\bar{h}_{s,i}|^2 (\sigma_s^2 |\bar{h}_{s,i}|^2 + \sigma_s^2 \sigma_{\phi_s}^2 + \sigma_{v_s}^2)}} \quad (29)$$

and

$$\mathbf{F}_6^\dagger = \text{diag}(f_1, f_2, \dots, f_N), \quad i = 1, \dots, N \quad (30)$$

where

$$f_i = \frac{\bar{h}_{y,i}^* h_{s,i}^* \sqrt{P_6}}{\sqrt{\sum_{i=1}^N |\bar{h}_{y,i}|^2 |h_{s,i}|^2 (\sigma_s^2 |h_{s,i}|^2 + \sigma_{v_s}^2)}}. \quad (31)$$

#### IV. SIMULATION RESULTS

The Monte-Carlo simulation results is performed to evaluate BER performance of the cooperative or noncooperative distributed MMSE relay scheme with imperfect channel information. All simulations are performed for one-source-one-destination pair  $N = 2$  cooperative or noncooperative distributed relay nodes. It is assumed that all relay nodes are located at equidistance between the source node and the destination node. The perfect channel coefficient vectors  $\mathbf{h}_s$  and  $\mathbf{h}_y$  are generated from independent Gaussian random variables with zero mean and unity variance. All nodes with only one antenna have the same noise power, i.e.,  $\sigma_{v_s}^2 = \sigma_{v_y}^2$ . It is also assumed that the transmitted signal at the source node is modulated by 4-ary quadrature amplitude modulation (4QAM). Perfect/imperfect channel coefficients with unity power are used. And, the total power constraints are set to  $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1$ .

For three different channel uncertainty conditions, i.e., 5%, 10%, and 20% of channel uncertainty are, respectively, modeled as additive white Gaussian noise and added to the perfect channel coefficient values, i.e., the variances of the channel estimation errors are chosen to satisfy  $10 \log_{10}(\sigma_{\phi_s}^2 / \sigma_{h_s}^2) = 10 \log_{10}(\sigma_{\phi_y}^2 / \sigma_{h_y}^2) = -13$  dB,  $-10$  dB, and  $-7$  dB whenever the links are imperfect. For comparison, the simulation results with perfect channel information are also included, i.e.,  $\sigma_{\phi_s}^2 = \sigma_{\phi_y}^2 = 0$ .

TABLE I

RELAY STRATEGIES AND CHANNEL UNCERTAINTY CONDITIONS IN FIG. 3 THROUGH FIG. 6 WITH  $N=2$  RELAY NODES

Fig. No.	Cooperative/noncooperative	Channel uncertainty
Fig. 3	cooperative	$(\bar{\mathbf{h}}_s, \bar{\mathbf{h}}_y)$
Fig. 4	cooperative	$(\bar{\mathbf{h}}_s, \mathbf{h}_y)$ or $(\mathbf{h}_s, \bar{\mathbf{h}}_y)$
Fig. 5	noncooperative	$(\bar{\mathbf{h}}_s, \bar{\mathbf{h}}_y)$
Fig. 6	noncooperative	$(\bar{\mathbf{h}}_s, \mathbf{h}_y)$ or $(\mathbf{h}_s, \bar{\mathbf{h}}_y)$

Figures 3 through 6 show the BER versus SNR for the cases listed in TABLE 1, where  $\text{SNR} \triangleq \sigma_s^2 \sigma_h^2 / \sigma_v^2 = \sigma_s^2 / \sigma_v^2$  represents the signal-to-noise power ratio. For example, Fig. 3 shows BER versus SNR for  $N = 2$  cooperative distributed relay networks under a  $(\bar{\mathbf{h}}_s, \bar{\mathbf{h}}_y)$  environment, i.e., both imperfect links  $\bar{\mathbf{h}}_s$  and  $\bar{\mathbf{h}}_y$ . Also, both perfect channel case  $(\mathbf{h}_s, \mathbf{h}_y)$

is included for comparisons. Due to the effect of imperfect channel information on both source-relay and relay-destination links in Fig. 3, the worst BER is observed, compared to the other cases in Figs. 4-6. In particular, it is observed that increasing the variances of the channel estimation errors, worse BER can be observed.

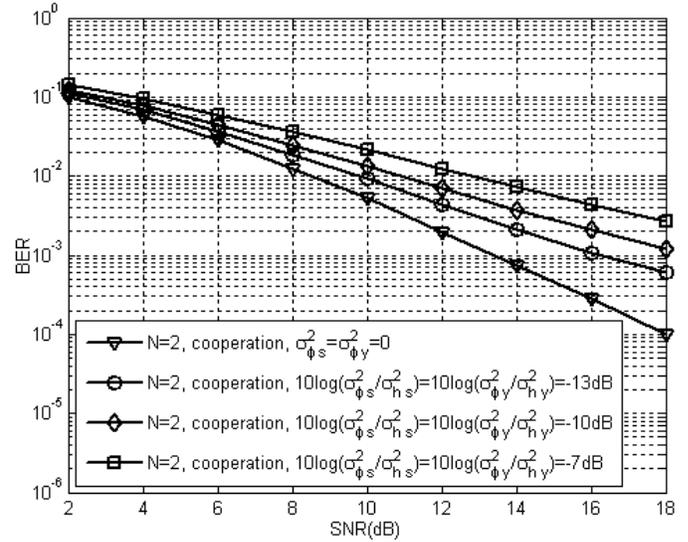


Fig. 3. BER performance of  $N = 2$  cooperative distributed relay networks under certain/uncertain channel conditions using both imperfect channel vectors  $(\bar{\mathbf{h}}_s, \bar{\mathbf{h}}_y)$ .

It is observed from Fig. 5 that the case  $(\mathbf{h}_s, \bar{\mathbf{h}}_y)$  shows a better performance than that of the  $(\bar{\mathbf{h}}_s, \mathbf{h}_y)$  case because the

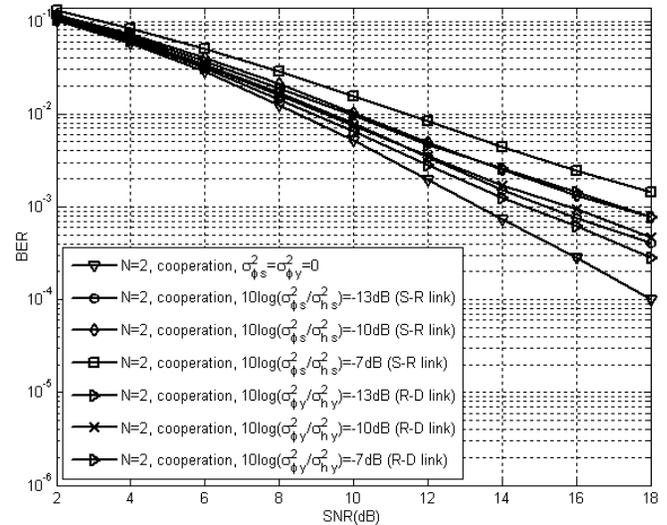


Fig. 4. BER performance of  $N = 2$  cooperative distributed relay networks under the cases of  $(\bar{\mathbf{h}}_s, \mathbf{h}_y)$  or  $(\mathbf{h}_s, \bar{\mathbf{h}}_y)$ .

diagonal relay amplifying matrix  $\mathbf{F}_3^\dagger$  in (24) is independent of

the variance of the channel estimation error  $\phi_y$ .

As observed from Fig. 3 and Fig. 5, the cooperative distributed relay scheme provides a better BER performance than the noncooperative one, e.g., 2 dB at BER =  $10^{-3}$ . As with the case of the cooperative channel uncertainty, BER of the noncooperative relay under the  $(\mathbf{h}_s, \hat{\mathbf{h}}_y)$  case shows a

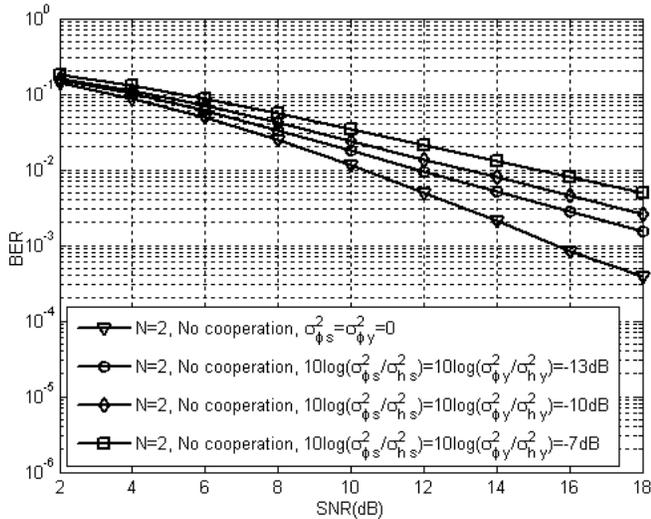


Fig. 5. BER performance of  $N = 2$  noncooperative distributed relay networks under certain/uncertain channel conditions using both imperfect channel vectors  $(\hat{\mathbf{h}}_s, \hat{\mathbf{h}}_y)$ .

better performance than that of the  $(\hat{\mathbf{h}}_s, \mathbf{h}_y)$  case because the diagonal relay amplifying matrix  $\mathbf{F}_6^\dagger$  in (31) is independent of the variance of the channel estimation error  $\phi_y$ . Finally, it

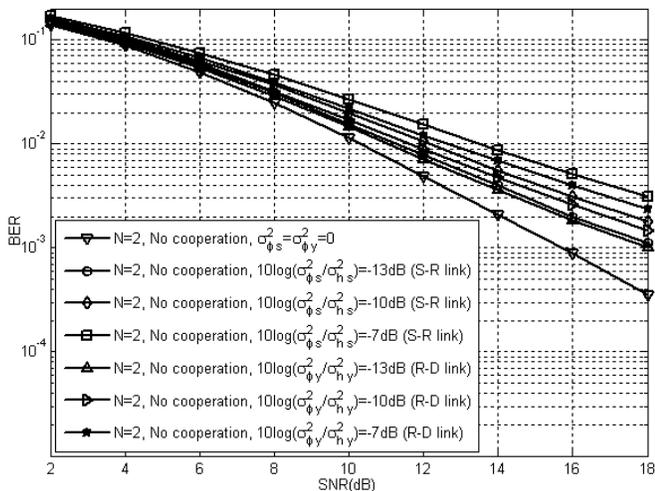


Fig. 6. BER performance of  $N = 2$  noncooperative distributed relay networks under the cases of  $(\hat{\mathbf{h}}_s, \mathbf{h}_y)$  or  $(\mathbf{h}_s, \hat{\mathbf{h}}_y)$ .

is observed in general from Figs. 4 and 6 that the effect of the noncooperative channel uncertainty is larger than that of

the cooperative one. And if both links are uncertain, then the performance is worse than the other cases.

## V. CONCLUSION

This paper considered distributed MMSE AF relay schemes in wireless relay networks under channel uncertainty and power constraints at relay nodes. It used one-source-one-destination pair and  $N$ -relay cooperative as well as noncooperative relay networks. Under either certain or uncertain channel condition, BER performance of cooperative distributed relay network is better than that of noncooperative case. Under channel uncertainty, it has been observed that increasing variances of the channel estimation error results in loss of diversity order. In addition, BER corresponding to the channel uncertainty before relays shows a worse performance than that of the channel uncertainty after relays.

## REFERENCES

- [1] Y. W. Hong, W. J. Huang, F. H. Chiu, and C. C. J. Kuo, "Cooperative communications in resource-constrained wireless networks," *IEEE Signal Process. Mag.*, vol. 24, no. 3, pp. 47-57, May 2007.
- [2] A. Saadani and O. Traoré, "Orthogonal or non orthogonal amplify and forward protocol: How to cooperate?" *IEEE WCNC 2008*, Las Vegas, NV, Mar. 2008, pp. 368-373.
- [3] J. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behaviour," *IEEE Transactions on Information Theory*, vol. 50, no. 12, Dec. 2004.
- [4] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037-3063, Sept. 2005.
- [5] T. Wang, A. Cano, G. B. Giannakis, and J. N. Laneman, "High performance cooperative demodulation with decode-and-forward relays," *IEEE Transactions on Communications*, vol. 55, no. 7, pp. 1427-1438, July 2007.
- [6] N. Khajehnouri and A. H. Sayed, "Distributed MMSE relay strategies for wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3336-3348, July 2007.
- [7] S. Berger and A. Wittneben, "Cooperative distributed multiuser MMSE relaying in wireless ad-hoc networks," *IEEE 39th Asilomar Conf.*, Oct. 2005, pp. 1072-1076.
- [8] R. Krishna, Z. Xiong, and S. Lambetharan, "A cooperative MMSE relay strategy for wireless sensor networks," *IEEE Signal Processing Letters*, vol. 15, pp. 549-552, 2008.
- [9] A. S. Behbahani and A. M. Eltawil, "Amplify-and-forward relay networks under received power constraint," *IEEE Transactions on Wireless Communications*, vol. 8, no. 11, pp. 5422-5426, Nov. 2009.
- [10] K. Lee, H. M. Kwon, Y. Ding, Y. Ibdah, Z. Wang, and Y. Bi, "Node geometry and broadband jamming in noncooperative relay networks under received power constraint," *IEEE SARNOFF 2011*, Princeton, NJ, May 3-4, 2011.
- [11] Y. Ibdah, H. M. Kwon, K. Lee, Z. Wang, Y. Bi, and M. Jo, "Broadband jamming and channel uncertainty for noncooperative wireless relay networks under received power constraint," *IEEE AMS 2011*, Manila, Philippines, May 23-27, 2011.
- [12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 1985.
- [13] R. A. Horn and C. R. Johnson, *Matrix Analysis*. 1st ed., Cambridge, MA: Cambridge University Press, 1985.
- [14] "The Matrix Cookbook," Online at: <http://matrixcookbook.com>.
- [15] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ: John Wiley & Sons, 2003.

# Finite-State Markov Chain Approximation for Geometric Mean of MIMO Eigenmodes

Ping-Heng Kuo

Information and Communication Laboratory  
Industrial Technology Research Institute (ITRI)  
Hsinchu, Taiwan  
pinghengkuo@itri.org.tw

Pang-An Ting

Information and Communication Laboratory  
Industrial Technology Research Institute (ITRI)  
Hsinchu, Taiwan  
pating@itri.org.tw

**Abstract**—Time-varying nature of mobile fading channels can affect the design of wireless systems in many different aspects. This paper investigates the dynamic behavior of intrinsic spatial links in multiple-input multiple-output (MIMO) transceivers based on Geometric mean decomposition (GMD). In particular, we suggest using a finite-state Markov chain (FSMC) to model the fluctuations of both eigenmodes geometric mean and capacity for MIMO-GMD schemes in Rayleigh fading channels. In order to compute the transition probabilities analytically, a number of related statistical distributions are developed and employed. Simulation results have illustrated that FSMC can provide accurate approximations.

**Keywords**—MIMO; Geometric mean decomposition; Rayleigh fading; Markov model.

## I. INTRODUCTION

Spatial multiplexing-based multiple-input multiple-output (MIMO) transceiver scheme has been regarded as one of the key technologies to increase the spectral efficiency of future wireless communication systems. By using multiple antennas at both transmitter and receiver side, multiple independent data streams can be transmitted within the same time-frequency resource with appropriate signal processing techniques [1]. In some common MIMO transceivers such as schemes based on singular value decomposition (SVD) [2], the gains of multiple parallel spatial links are different, which may lead to high complexity in power distribution and bit-allocation algorithms in spatial domain. In order to reduce the complexity, it is possible to use the same modulation format on all subchannels. However, this may result in the degradation in error performance since the BER (bit error rate) is dominated by the sub-channel with the lowest power gain. In general, the minimum singular value of the MIMO channel matrix can affect the system performance in many ways [3].

An alternative MIMO transceiver architecture termed as Geometric Mean Decomposition (GMD) has been proposed by Jiang *et al.* [4], which is able to realize multiple parallel spatial links with identical gains from a MIMO channel. Specifically, the gains of all these parallel links are equal to the geometric mean of channel eigenmodes. Thus, bit and

power allocation can be undertaken more conveniently as the modulation/coding scheme on all subchannels should be the same. In recent years, MIMO-GMD transceivers have received the attention from many researchers in both academia and industry. In particular, both LTE-A and WiMAX (802.16m), two of the most important standards for the next-generation wireless broadband, have considered MIMO-GMD in the standardization process [5], [6].

Nevertheless, to the best of our knowledge, statistical properties of the link gains realized by GMD in MIMO fading channels have not been characterized. In this paper, we are particularly interested in the dynamic behavior of the link gain and channel capacity in MIMO-GMD, which is important to the design of various adaptive transmission mechanisms. For instance, the adaptation/feedback rate for MIMO-GMD with adaptive modulation could be set in a more judicious manner if one could gauge the variation of link gain and capacity. On the other hand, successive interference cancelation (SIC) is usually used in the detection of a MIMO-GMD transceiver, in which the reliability of a data stream is dependent on previously-decoded data streams. If the receiver fails to detect a data stream successfully, error propagation is very likely to occur as the receiver is not able to correctly decode the data on the remaining streams. This might happen when the spatial link gain is too weak, and the characterization on channel variation may provide some crucial information on how frequently an error propagation event would occur.

Finite-state Markov chain (FSMC) is a commonly-used tool to model a time-varying process. It generally quantizes the process into multiple distinct states, and the time-varying behavior is characterized by the transition probabilities among these states. Some researchers have designed the feedback mechanisms for adaptive modulation systems with channels modeled by FSMC [7]. In this paper, FSMC is used to model both link gain and capacity processes of MIMO-GMD systems in fading channels. We provide several univariate and bivariate probability density functions (PDFs) of spatial links and capacities for a MIMO-GMD scheme in this paper, and demonstrate the methods of transition-

probability computations using these distribution functions. Note that the univariate PDFs have already been derived in [8]. In order to construct FSMC, this paper applies the results obtained in [8], as well as extending the derivations for bivariate PDFs.

The remainder of the paper is organized as following: In Section II, the channel model for our analysis is described, and the concept of MIMO-GMD architecture and FSMC are briefly reviewed. In Section III, univariate PDFs are developed based on Gamma approximations. Then, bivariate PDFs for eigenvalue geometric mean and channel capacity are given in Section IV. In Section V, we calculate transition probabilities for the geometric mean and capacity for the FSMC. Finally, conclusions are drawn in Section VI.

## II. BACKGROUND

### A. Channel Model and Assumptions

In this paper, we consider an  $(N_t, N_r)$  MIMO system with  $N_t$  and  $N_r$  antennas at transmitter and receiver respectively. The MIMO channel matrix,  $\mathbf{H}$ , is therefore an  $N_r \times N_t$  matrix. Here we denote  $m = \min(N_t, N_r)$ ,  $n = \max(N_t, N_r)$  and  $l = n - m$ . Each of the channel entries is modeled as a complex Gaussian random variable with zero mean and unit variance (Rayleigh fading). The eigenvalues of the channel correlation matrix ( $\mathbf{H}^\dagger \mathbf{H}$ ), represent the power gains of spatial links intrinsic to the SVD-based MIMO systems (the so called eigenmodes), and their joint statistics are governed by central Wishart distribution [9]:

$$f(\lambda_1, \dots, \lambda_m) = \frac{\exp(-\sum_{i=1}^m \lambda_i)}{m! \prod_{i=1}^m (n-i)!(m-i)!} \times \prod_{i=1}^m \lambda_i^l \prod_{i < j} (\lambda_i - \lambda_j)^2. \quad (1)$$

On the other hand, we presume that the channel evolves over time in accordance to:

$$\mathbf{H}(t+\tau) = J_0(2\pi f_D \tau) \mathbf{H}(t) + \sqrt{1 - J_0(2\pi f_D \tau)^2} \Psi, \quad (2)$$

where  $J_0(\cdot)$  is the zeroth-order Bessel function and  $f_D$  and  $\tau$  are the Doppler frequency and time displacement, respectively.  $\Psi$  is a  $N_r \times N_t$  random matrix with zero-mean complex Gaussian elements  $\mathcal{CN}(0, 1)$ . For simplicity, we denote the auto-correlation  $\zeta = J_0(2\pi f_D \tau)$  in the rest of the paper.

### B. Review of Geometric Mean Decomposition

The general input-output relationship of a MIMO transmission in flat-fading channel can be written as:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \quad (3)$$

where  $\mathbf{y}$  is an  $N_r \times 1$  vector of received symbol,  $\mathbf{x}$  is an  $N_t \times 1$  vector of transmitted data, and  $\mathbf{z}$  is a Gaussian noise vector with  $N_r$  elements. Based on (3), the authors of [4]

have suggested to design MIMO transceivers using GMD, in which the MIMO channel is decomposed as:

$$\mathbf{H} = \mathbf{Q}\mathbf{R}\mathbf{P}^\dagger \quad (4)$$

where  $\mathbf{P}$  and  $\mathbf{Q}$  are semi-unitary matrices and  $\mathbf{R}$  represents an  $m \times m$  upper triangular matrix with identical diagonal elements. By pre-coding the data signal  $\mathbf{x}$  by  $\mathbf{P}$  at the transmitter, and multiplying  $\mathbf{y}$  by  $\mathbf{Q}^\dagger$  at the receiver, the transmitted signal can be extracted and decoded using the *nulling and cancelation* procedure as in BLAST schemes [1]. Hence, without considering the effects of error propagation, the data signals are sent on  $m$  subchannels. We denote the diagonal elements of  $\mathbf{R}$  by  $g_\sigma$  and

$$g_\sigma = \left( \prod_{i=1}^m \sigma_i \right)^{\frac{1}{m}}. \quad (5)$$

where  $\sigma_i$  represents the  $i^{\text{th}}$  singular value of the MIMO channel  $\mathbf{H}$ . Hence, the sub-channel envelopes are identical and are equivalent to the geometric mean of MIMO channel singular values. Remarkably, as  $m$  subchannels have identical gains, the performance of MIMO-GMD schemes does not suffer from "worst subchannel" problems as in some other transceiver architectures. Presuming that the transmission power is uniformly allocated on all subchannels, the resultant channel capacity is

$$C = m \log_2(1 + \gamma g_s^2) = m \log_2(1 + \gamma g), \quad (6)$$

where  $\gamma$  denotes the SNR on each sub-channel and

$$g = g_s^2 = \left( \prod_{i=1}^m \lambda_i \right)^{\frac{1}{m}} \quad (7)$$

is the sub-channel power gain, which is equivalent to the geometric mean of eigenmodes.

### C. Finite-State Markov Chain

FSMC has been widely used for modeling time-varying processes in various contexts such as signal processing and wireless communications. In order to model the fluctuation of a random process using an FSMC, the process shall be firstly quantized into a number of discrete states. Then, the dynamic behavior is characterized by the transition probabilities among these states. In this paper, we only consider the first-order FSMC, in which the transition probability is solely dependent on the preceding state. In our case, we partition the processes of both eigenmode geometric mean and capacity based on their magnitudes. That is, we simply partition the random process of interest into  $S$  states by setting  $S - 1$  threshold levels. In the first state (denoted as  $\mathcal{S}_0$ ), the process has a value smaller than the lowest threshold level  $T_0$ . In the final state ( $\mathcal{S}_S$ ), the process is larger than the highest threshold level  $T_{S-1}$ . Otherwise, the process is said to be in the  $i^{\text{th}}$  state (denoted as  $\mathcal{S}_i$ ) if its value falls in the range between threshold levels  $T_{i-1}$

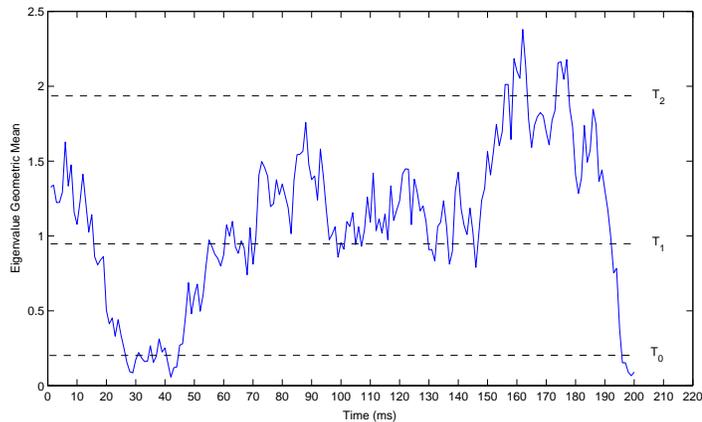


Figure 1. An example of a Markov state quantization. The time-varying eigenmodes geometric mean process in a (2,2) MIMO channel is partitioned into four states using three threshold levels:  $T_0$ ,  $T_1$  and  $T_2$ .

and  $T_i$ . Fig 1 shows an example in which the eigenmode geometric mean process is partitioned into four states. The main objective of this paper is to provide analytical methods of calculations for transition probabilities, which requires the univariate and bivariate statistical distributions for the target random processes (eigenmode geometric mean and capacity in MIMO-GMD schemes). We develop these distribution functions in subsequent sections.

### III. UNIVARIATE DISTRIBUTIONS

For MIMO-GMD schemes with two subchannels ( $m = 2$ ), the exact PDFs for the eigenmodes geometric mean can be written as described in [8]:

$$f(g) = \frac{4g^{2(n-1)}}{(n-1)!(n-2)!} K_1(2g), \quad (8)$$

where  $K_1(\cdot)$  represents the Modified Bessel Function of the Second Kind of order 1. Moreover, the closed-form PDF for the channel capacity in MIMO-GMD scheme with  $m = 2$  has also been derived as:

$$f(C) = \frac{2 \ln(2) \sqrt{2}^C (2^{\frac{C}{2}} - 1)^{2(n-1)}}{\gamma^{(n-0.75)} (n-1)!(n-2)!} K_1 \left[ \frac{2}{\gamma} (2^{\frac{C}{2}} - 1) \right]. \quad (9)$$

Nonetheless, for MIMO-GMD schemes with  $m > 2$  subchannels, the PDFs for both eigenmode geometric mean and channel capacity are difficult, if not impossible, to derive due to high computational complexity. Fortunately, some previous studies have concluded that MIMO eigenmodes  $(\lambda_1, \lambda_2, \dots, \lambda_m)$  can be very accurately approximated by a Gamma random variables [10]. In addition, [11] claims that the geometric mean of multiple independent Gamma random variables could be either a Gamma or a mixture of Gamma distributions. Although MIMO eigenvalues are not

Table I  
PARAMETERS OF GAMMA APPROXIMATION FOR EIGENVALUE GEOMETRIC MEAN,  $g$

	(4,4)	(4,8)	(8,8)
$E(g)$	1.8075	6.0279	3.2150
$\text{Var}(g)$	0.5280	1.5306	0.5507
$k$	6.1880	23.7393	18.7687
$\theta$	0.2921	0.2539	0.1713

independent processes, they are weakly correlated. So we may simply approximate the eigenvalue geometric mean  $g$  by Gamma distribution, the PDF which is widely-known.

Therefore, we may make a hypothesis stating that

$$f(g) \approx \text{Gamma}(k, \theta) = \frac{g^{k-1} \exp(-\frac{g}{\theta})}{\Gamma(k) \theta^k}, \quad (10)$$

where  $k = E(g)/\theta$  and  $\theta = \text{Var}(g)/E(g)$  are shape factor and scale factor of Gamma distribution, respectively. The computations for  $\text{Var}(g)$  requires the first two moments of  $g$  as  $\text{Var}(g) = E(g^2) - E(g)^2$ . Note that the  $v^{\text{th}}$  moment of  $g$  can be calculated as:

$$E(g^v) = \int_0^\infty \dots \int_0^\infty \left( \prod_{i=1}^m \lambda_i \right)^{\frac{v}{m}} f(\lambda_1, \dots, \lambda_m) d\lambda_1, \dots, d\lambda_m, \quad (11)$$

where  $f(\lambda_1, \dots, \lambda_m)$  has been given as (1). To make this paper self-contained, we have tabulated the numerical values of  $E(g)$ ,  $\text{Var}(g)$ ,  $k$  and  $\theta$  for (4,4), (4,8) and (8,8) cases in Table I. Hence, one may construct the PDFs for  $g$  in these cases accordingly.

Utilizing Gamma approximation of (10), we can extend it to find the results on capacity PDF. With simple transformations, we have

$$\begin{aligned} f(C) &= \left| \frac{\partial g}{\partial C} \right| \times f(g)|_{g \rightarrow g(C)} \\ &= \frac{\ln(2) 2^{\frac{C}{m}} (2^{\frac{C}{m}} - 1)^{k-1}}{m \gamma^k \Gamma(k) \theta^k} \exp\left(\frac{1 - 2^{\frac{C}{m}}}{\gamma \theta}\right). \end{aligned} \quad (12)$$

Here we justify the suitability of Gamma approximations for the eigenmode geometric mean via simulations. In Fig 2, we can see that Gamma distributions provide excellent approximation for eigenvalue geometric means. Although we did not show here, it has been found that Gamma distributions also fit the simulation data in  $m = 2$  cases. Hence, we can claim that, for practical MIMO systems ( $m \leq 8$ ) at least, the distributions of eigenmode geometric means can be approximated by Gamma random variables. For channel capacity, we compare the distributions of simulation samples with the capacity PDF (12), which was derived based on Gamma approximation of  $g$ . The comparisons are shown in Fig 3, and we can see that simulation data and computed results highly agree with each other.

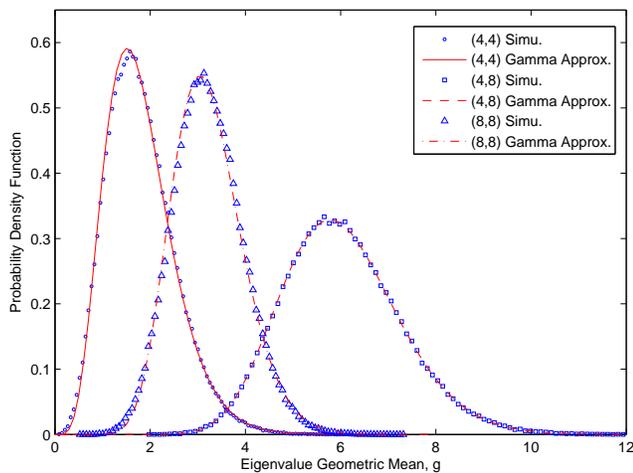


Figure 2. Comparison between simulation and Gamma approximation results for distributions of eigenvalue geometric mean (sub-channel power gain) in (4,4), (4,8) and (8,8) MIMO Rayleigh fading channels.

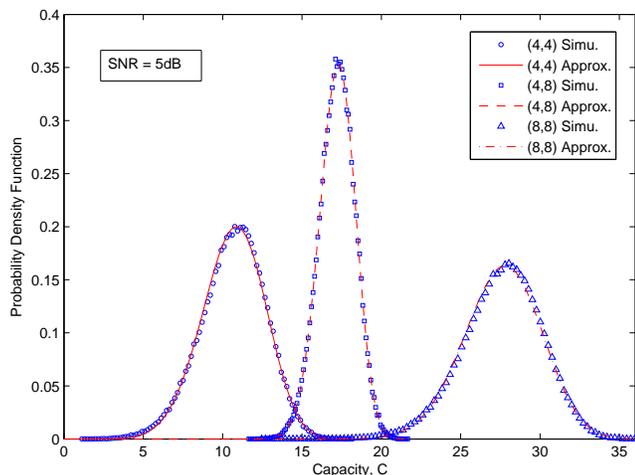


Figure 3. Comparison between simulation and calculation results for distributions of system capacity using GMD in (4,4), (4,8) and (8,8) MIMO Rayleigh fading channels. Assuming  $\gamma = 5dB$ .

#### IV. BIVARIATE DISTRIBUTIONS

In the previous section, we have developed PDFs for the eigenmodes geometric mean and channel capacity based on Gamma approximations. In order to compute the transition probabilities for FSMC, the joint probability density for the process at two adjacent time points is also required. Since we have observed that the eigenmode geometric mean can be accurately approximated by a Gamma random variable, the joint density of eigenmode geometric mean at two correlated time instants,  $g(t)$  and  $g(t + \tau)$ , could be described by the bivariate Gamma PDFs. Thus, by modifying the bivariate

Gamma PDF in [12], we have:

$$f(g, \hat{g}) \approx \frac{(g\hat{g})^{\frac{k-1}{2}}}{\theta^{k+1}\Gamma(k)(1-\rho)\rho^{\frac{k-1}{2}}} \exp\left\{\frac{-(g+\hat{g})}{\theta(1-\rho)}\right\} \times I_{r-1}\left(\frac{2}{\theta(1-\rho)}\sqrt{\rho g\hat{g}}\right), \quad (13)$$

where  $\hat{g} = g(t + \tau)$  and  $\rho$  represents the correlation coefficient between  $g$  and  $\hat{g}$ , which is defined as

$$\begin{aligned} \rho &= \frac{E[g(t)g(t+\tau)] - E[g(t)]E[g(t+\tau)]}{\sqrt{\text{Var}[g(t)]\text{Var}[g(t+\tau)]}} \\ &= \frac{E(g\hat{g}) - E(g)^2}{\text{Var}(g)}. \end{aligned} \quad (14)$$

In (14), both  $E(g)$  and  $\text{Var}(g)$  can be calculated using (1). On the other hand, the computation for the joint moment,  $E(g\hat{g})$ , requires the joint density for eigenmodes at two time instants [10]:

$$\begin{aligned} f(\lambda_1, \dots, \lambda_m, \hat{\lambda}_1, \dots, \hat{\lambda}_m) &= \frac{\prod_{i < j} [\frac{1}{\alpha^2}(\hat{\lambda}_i - \hat{\lambda}_j)] \prod_{i < j} (\lambda_i - \lambda_j) \prod_{i=1}^m \lambda_i^l}{\alpha^{2m} \prod_{i=1}^m [(n-i)!(m-i)!]} \\ &\times \exp\left(-\sum_{i=1}^m \lambda_i\right) \times G(\boldsymbol{\lambda}), \end{aligned} \quad (15)$$

where

$$G(\boldsymbol{\lambda}) = \det \begin{bmatrix} f(\hat{\lambda}_1|\lambda_1) & f(\hat{\lambda}_2|\lambda_1) & \cdots & f(\hat{\lambda}_m|\lambda_1) \\ f(\hat{\lambda}_1|\lambda_2) & f(\hat{\lambda}_2|\lambda_2) & & \vdots \\ \vdots & & \ddots & \vdots \\ f(\hat{\lambda}_1|\lambda_m) & \cdots & \cdots & f(\hat{\lambda}_m|\lambda_m) \end{bmatrix}$$

and

$$f(\hat{\lambda}|\lambda) = \frac{\alpha^2}{\beta^2} \left(\frac{\hat{\lambda}}{\alpha^2\lambda}\right)^{v/2} \exp\left[\frac{-\alpha^2\lambda - \hat{\lambda}}{\beta^2}\right] I_l\left(2\frac{\alpha}{\beta^2}\sqrt{\lambda\hat{\lambda}}\right)$$

with  $\alpha = \zeta$ ,  $\beta = \sqrt{1 - \zeta^2}$  and  $\zeta = J_0(2\pi f_D \tau)$  as before.

We can extend (13) to obtain the bivariate PDF for MIMO-GMD channel capacity process. Using Jacobian transform,  $f(C, \hat{C})$  can be written as

$$\begin{aligned} f(C, \hat{C}) &= \det \begin{vmatrix} \frac{\partial g}{\partial C} & \frac{\partial g}{\partial \hat{C}} \\ \frac{\partial \hat{g}}{\partial C} & \frac{\partial \hat{g}}{\partial \hat{C}} \end{vmatrix} \times f(g, \hat{g})|_{g \rightarrow g(C), \hat{g} \rightarrow \hat{g}(\hat{C})} \\ &= 0.4805 \frac{2^{\frac{C+\hat{C}}{m}} \left[\frac{C}{(2m-1)\gamma^2\rho}\frac{\hat{C}}{(2m-1)}\right]^{(k-1)/2}}{m^2 \gamma^2 \Gamma(k) \theta^{k+1} (1+\rho)} \\ &\times \exp\left[\frac{2 - 2^{C/m} - 2^{\hat{C}/m}}{\gamma \theta (1-\rho)}\right] \\ &\times I_{k-1}\left[\frac{2\sqrt{\rho}(2^{C/m}-1)(2^{\hat{C}/m}-1)}{\gamma \theta (1-\rho)}\right]. \end{aligned} \quad (16)$$

With PDFs given in Sections III and IV, the transition probabilities of FSMC can be computed.

## V. COMPUTATIONS OF TRANSITION PROBABILITIES

As aforementioned, we consider a first-order FSMC in this paper. Therefore, the transition probabilities from  $\mathcal{S}_i$  to  $\mathcal{S}_j$ , denoted as  $P_{i,j}$ , for the eigenmode geometric mean process can be computed as

$$P_{i,j} = \frac{\text{Prob}(g(t) \in \mathcal{S}_i, g(t+\tau) \in \mathcal{S}_j)}{\text{Prob}(g(t) \in \mathcal{S}_i)}, \quad (17)$$

where

$$\text{Prob}(g(t) \in \mathcal{S}_i, g(t+\tau) \in \mathcal{S}_j) = \int_{\mathcal{S}_j} \int_{\mathcal{S}_i} f(g, \hat{g}) dg d\hat{g}, \quad (18)$$

and

$$\text{Prob}(g(t) \in \mathcal{S}_i) = \int_{\mathcal{S}_i} f(g) dg. \quad (19)$$

Note that  $f(g)$  could be (8) or (10), depending on the value of  $m$ . For sake of simplicity, we assume that the channel variation is slow enough so the process would only transit to one of the adjacent states (from  $\mathcal{S}_i$  to  $\mathcal{S}_{i-1}$  or  $\mathcal{S}_{i+1}$ ) or stay in the same state (from  $\mathcal{S}_i$  to  $\mathcal{S}_i$ ) in this paper. Therefore, we have

$$P_{i,i} = 1 - P_{i,i+1} - P_{i,i-1}. \quad (20)$$

For channel capacity in MIMO-GMD scheme, the transition probabilities can be computed in a similar fashion. In order to verify that FSMC is an appropriate tool to model the fluctuation of MIMO-GMD channel, several Monte Carlo simulations have been carried out. In particular, simulation results on transition probabilities for both eigenmodes geometric mean and channel capacity are compared with our calculations by (17) and (20). Note that the threshold levels for state quantization are set arbitrarily in our simulations. In practical adaptive modulation schemes, for example, the threshold levels for state quantization could be set based on minimum-required channel gain for a target error performance. We set  $f_D = 30\text{Hz}$ ,  $\tau = 0.001\text{sec}$  and  $\gamma = 10\text{dB}$  in all scenarios. In Fig 4, the eigenmode geometric mean is modeled as a FSMC consisting of four states with  $\{T_0, T_1, T_2\} = \{1.5, 2.5, 3.5\}$ , and transition probabilities from both simulations and calculations are plotted. Similarly, we approximate the MIMO-GMD capacity process in a (2,2) system using a four-state FSMC with  $\{T_0, T_1, T_2\} = \{4, 7, 9\}$  in Fig 5. In both cases, it is apparent that our calculations can provide very accurate approximations.

## VI. CONCLUSION

In this paper, we have aimed to model the channel variation in MIMO-GMD scheme using a FSMC. Since FSMC simply quantizes the process into multiple discrete states, which naturally fits to various applications in channel-dependent adaptive transmission. For example, in link adaptation mechanisms, each modulation and coding scheme usually corresponds to a range of channel gains, which can

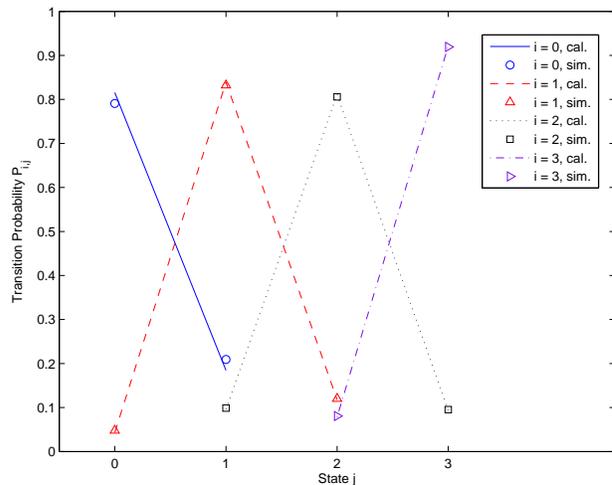


Figure 4. Comparison of transition probabilities for the eigenmodes geometric mean (with  $f_D = 30\text{Hz}$ ,  $\tau = 0.001\text{sec}$ ) in a (2,4) MIMO system obtained from calculations and simulations.

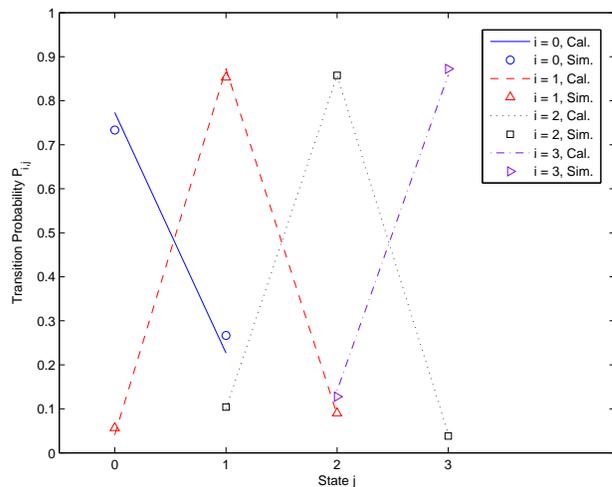


Figure 5. Comparison of transition probabilities for the MIMO-GMD channel capacity (with  $f_D = 30\text{Hz}$ ,  $\tau = 0.001\text{sec}$  and  $\gamma = 10\text{dB}$ ) in a (2,2) system obtained from calculations and simulations.

be deemed as a state in the FSMC. Additionally, we may also use FSMC of eigenmode geometric-mean to examine how likely the MIMO-GMD system would enter the state with high probability of error propagation. In order to calculate transition probabilities analytically, we have approximated the eigenmodes geometric-mean process by Gamma distribution, and developed associating bivariate PDFs based on such approximations. From the simulation results, it is apparent that our computations can give very accurate approximations. Hence, we can conclude that FSMC is

a simple but effective approach to capture the dynamic behavior of a MIMO-GMD channel.

#### REFERENCES

- [1] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. IEEE ISSSE*, Pisa, Italy, Sep. 1998, pp. 295–300.
- [2] G. Lebrun, J. Gao, and M. Faulkner, "MIMO transmission over a time-varying channel using SVD," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 757–764, Mar. 2005.
- [3] G. Burel, "Statistical analysis of the smallest singular value in MIMO transmission systems," in *Proc. WSEAS Int. Conf. on Sig. Speech and Img. Processing*, Skiathos, Greece, Sep. 2002.
- [4] Y. Jiang, J. Li, and W. W. Hager, "Joint transceiver design for MIMO communications using geometric mean decomposition," *IEEE Trans. Signal Process.*, vol. 53, pp. 3791–3803, Oct. 2005.
- [5] "GMD beamforming for LTE-Advanced SU/MU-MIMO," R1-093799, 3GPP TSG RAN WG1 Meeting 58, Oct. 2009.
- [6] "A joint transceiver design for MIMO precoder and antenna selection mechanism," C802.16m-08/314r2, IEEE Broadband Wireless Access Working Group, May 2008.
- [7] A. E. Ekpenyong and Y.-F. Huang, "Feedback-detection strategies for adaptive modulation systems," *IEEE Trans. Commun.*, vol. 54, pp. 1735–1740, Oct. 2006.
- [8] P.-H. Kuo, J.-H. Wu, Y.-T. Hsieh, and P.-A. Ting, "Statistical distributions for link gain and capacity of MIMO-GMD transceivers in Rayleigh fading channels," in *Proc. IEEE Veh. Tech. Conf. (VTC) Fall*, Ottawa, Canada, Sep. 2010.
- [9] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecomm. Related Technol.*, vol. 10, pp. 585–595, Nov. 1999.
- [10] P.-H. Kuo, *Fluctuations of MIMO Wireless Channels: Analysis based on Brownian motion theory*. Saarbrücken, Germany: VDM Verlag Dr. Muller, 2008.
- [11] R. S. Laurent, "The distribution of the geometric mean of independent F (or gamma) random variables," in *Proc. Joint Statistical Meetings (JSM)*, Atlanta, GA, USA, Aug. 2001.
- [12] G. E. Oien, H. Holm, and K. J. Hole, "Impact of channel prediction on adaptive coded modulation performance in Rayleigh fading," *IEEE Trans. Veh. Technol.*, vol. 53, pp. 758–769, May 2004.

# An Algorithmic Approach for Analyzing Wireless Networks with Retrials and Heterogeneous Servers

Nawel Gharbi

*Computer Science Department  
University of Sciences and Technology, USTHB  
Algiers, Algeria  
Email: ngharbi@wissal.dz*

Leila Charabi

*Computer Science Department  
University of Sciences and Technology, USTHB  
Algiers, Algeria  
Email: leila.charabi@gmail.com*

**Abstract**—Models with retrial phenomenon and heterogeneous servers arise in various wireless networks. This paper aims at presenting an approach for modeling and analyzing finite-source wireless networks with retrial phenomenon and heterogeneous servers using the Generalized Stochastic Petri Nets. This high-level formalism allows a simple representation of complex systems. Moreover, from the GSPN model, a Continuous Time Markov Chain can be automatically derived for the performance analysis. However, for important retrial networks, generating the Markov chain from the GSPN and solving it, require large storage space and long execution time. Hence, using the GSPN model as a support, we propose an algorithm for directly computing the infinitesimal generator of the GSPN model without generating neither the reachability graph nor the underlying Markov chain. In addition, we develop the formulas of the main stationary performance indices, as a function of the network parameters, the stationary probabilities and independently of the reachability set markings. Through numerical examples, we discuss the effect of the system parameters on performance.

**Keywords**-Retrial phenomenon; Heterogeneous servers; Wireless networks; Generalized Stochastic Petri nets; Performance indices.

## I. INTRODUCTION

Models with retrial phenomenon are characterized by the feature that a customer finding all servers busy or unavailable, is obliged to leave the service area, but he repeats his request after some random period of time. These models play an important role in wireless cellular networks [4], [11], [12]. Significant references reveal the non-negligible impact of repeated calls, which arise due to a blocking in a system with limited capacity resources or are due to impatience of customers. For a systematic account of the fundamental methods and results on this topic, we refer the readers to [2], [3], [7].

Most studies on retrial models with finite source (population), assume that the service station consists of homogeneous (identical) servers. However, retrial models with heterogeneous servers arise in various practical areas as telecommunications and cellular mobile networks. In fact, these models are far more difficult for mathematical analysis than models with homogeneous servers, and explicit results

are available only in few special cases and almost all studies are investigated only by means of queueing theory. In fact, we have found in the literature, only the few papers of Efrosinin [6] and Sztrik [10], [9] where heterogeneous servers case was considered using retrial queueing model, and the paper [8] where we have proposed the modeling and the analysis of multiclass retrial systems by means of colored generalized stochastic Petri nets.

The objective of this paper is to present an approach for modeling and analyzing performances of finite-source retrial networks with heterogeneous servers using the Generalized Stochastic Petri Nets (GSPNs) [1], [5]. From a modeling point of view, and compared to retrial queueing models, this high-level graphical formalism allows an easier description of the behavior of complex retrial networks, and it has shown to be a very effective mathematical model, appropriate for modeling and analyzing performance of parallel systems exhibiting concurrency and synchronization. Moreover, from the GSPN model, a Continuous Time Markov Chain (CTMC) can be automatically derived for the performance analysis. However, generating the Markov chain from the GSPN and solving it, still requires large storage space and long execution time, since the state space increases exponentially as function of the customers source size and servers' number. So, for important retrial networks, the corresponding models may have a huge state space. Hence, using the GSPN model as a support, we propose in this paper, an algorithmic approach for directly computing the infinitesimal generator without generating the reachability graph nor the underlying Markov chain. In addition, we develop the formulas of the main stationary performance indices, as a function of the number of servers of each class, the size of the customers source, the stationary probabilities and independently of the reachability set markings.

The paper is organized as follows. In Section 2, we describe the basic model of finite-source retrial networks with heterogeneous servers. In Section 3, the basic notions of GSPNs are reviewed. Next, we present the GSPN model describing retrial networks with heterogeneous servers. In Section 4, the proposed stochastic analysis approach is

detailed. The computational formulas for evaluating exact performance indices are derived in Section 5. Next, based on numerical examples, we validate the proposed approach and we discuss the effect of system parameters on the performability of the system. Finally, we give a conclusion.

## II. THE BASIC MODEL

We consider retrial networks with finite source (population) of customers of size  $L$  and a service station that consists of heterogeneous servers. Each customer is either free, under service or in orbit at any time. The input stream of primary calls is the so called quasi-random input. The probability that any particular customer generates a primary request for service in any interval  $(t, t + dt)$  is  $\lambda dt + o(dt)$  as  $dt \rightarrow 0$  if the customer is free at time  $t$ , and zero if the customer is being served or in orbit at time  $t$ .

The servers are partitioned in two classes: Class  $C_1$  and Class  $C_2$ , that is the servers of a given class have the same parameters. Each class  $C_j$  ( $1 \leq j \leq 2$ ) contains  $S_j$  identical and parallel servers. There are two possible states for a server: idle or busy (on service). If there is an idle server at the moment a customer request arrives, then the service starts immediately. The customer becomes "under service" and the server becomes "busy". Service times are independent identically-distributed random variables, whose distribution is exponential with parameter  $\mu_1$  if a class  $C_1$  server is selected and  $\mu_2$  for servers of class  $C_2$ .

Each customer request must be served by one and only one server. On the other hand, we consider, the *Random Server discipline*, which means that, the server to which a request is assigned is chosen randomly among all idle servers, whatever their class. After service completion, the customer becomes free, so it can generate new primary calls, and the server becomes idle again. Otherwise, if all servers of the two classes are busy at the arrival of a request, the customer joins the orbit and starts generation of a flow of repeated calls exponentially distributed with rate  $\nu$ , until he finds one free server. We assume that all customers are persistent in the sense that they keep making retrials until they receive their requested service.

As usual, we assume that the arrival, service and inter-trial times are mutually independent of each other.

## III. GSPN MODEL OF RETRIAL NETWORKS WITH HETEROGENEOUS SERVERS

In this section, we present the Generalized Stochastic Petri Net model describing finite-source retrial networks with two servers classes and random server discipline. To this aim, we give the basic notions of Generalized stochastic Petri nets (GSPNs).

A GSPN [1], [5] is a directed graph that consists of two kinds of nodes, called places (drawn as circles) and transitions that are partitioned into two different classes: timed transitions with rates of negative exponential distribution

(represented by means of white rectangles), which describe the execution of time consuming activities and immediate transitions (represented by black rectangles), which model logic activities as synchronization.

The system state is described by means of markings. A marking is a mapping from  $P$  to  $IN$ , which gives the number of tokens in each place after each transition firing. A transition is said to be enabled in a given marking, if and only if each of its normal input places contains at least as many tokens as the multiplicity of the connecting arc, and each of its inhibitor input places contains fewer tokens than the multiplicity of the corresponding inhibitor arc. Moreover, timed transitions can fire only after an exponentially distributed delay, while enabled immediate transitions have priority over timed transitions and fire in zero time.

The firing of an enabled transition creates a new marking (state) of the net. The set of all markings reachable from initial marking  $M_0$  is called the *reachability set*. The *reachability graph* is the associated graph obtained by representing each marking by a vertex and placing a directed edge from vertex  $M_i$  to vertex  $M_j$ , if marking  $M_j$  can be obtained by the firing of some transition enabled in marking  $M_i$ .

This graph consists of *tangible markings* enabling only timed transitions and *vanishing markings* in which at least one immediate transition is enabled. Since the process spends zero time in the vanishing markings, they are eliminated from the reachability graph by merging them with their successor tangible markings [1]. This elimination results in a *tangible reachability graph*, which is isomorphic to a continuous time Markov chain (CTMC).

In the following, we present the GSPN model describing finite-source retrial systems with two servers classes and random server discipline. In this model depicted in Figure 1, place *Cus\_Free* represents the free customers, *Orbit* contains the customers waiting for the service, *Ser\_Idle1* and *Ser\_Idle2* indicate respectively the number of free servers of class  $C_1$  and class  $C_2$ , while *Cus\_Serv1* and *Cus\_Serv2* model the busy servers of both classes. The arrival of a primary call causes the firing of the transition *Arrival*, which firing rate is marking dependent and equals  $\lambda.M(Cus\_Free)$  (*infinite service semantics*) which is represented by the symbol  $\#$  placed next to transition, because all free customers are able of generating calls, independently of each other.

The place *Choice* is then marked. Following the marking of both *Ser\_Idle1* and *Ser\_Idle2*, we have the following scenarios:

- If both places are empty (no free server), the immediate transition *Go\_Orbit* is enabled, and a token is deposited in the place *Orbit*, which means that the customer asking for service joins the orbit and becomes a source of a flow of repeated calls exponentially distributed with rate  $\nu$ .

The firing of the transition *Retrial* corresponds to the

generation of a repeated call from a customer in orbit. This transition has infinite servers semantics, since all customers in orbit can trigger repeated calls independently;

- If only one place (*Ser\_Idle1* or *Ser\_Idle2*) contains tokens and the other is empty, the immediate transition corresponding to the class with at least one free server (*Begin\_Serv1* for the class  $C_1$ , and *Begin\_Serv2* for class  $C_2$ ) is enabled. Hence the customer starts its service and the server becomes busy.
- When both places are marked (i.e., each class contains at least one free server), the two immediate transitions *Begin\_Serv1* and *Begin\_Serv2*, which are already in *structural conflict*, come into *effective conflict*. As there is no priority order between the two classes  $C_1$  and  $C_2$ , the same weight is assigned to both transitions as follows:

$$w(\text{Begin\_Serv1}) = w(\text{Begin\_Serv2}) = \omega$$

Thus, firing of one or other is probabilistic and the probability equals 1/2. Hence, servers of both classes have the same chance of being selected to serve the customers requests.

By the end of service of a customer under a server of class  $C_1$  ( $C_2$  respectively), the timed transition *Serv\_End1* (*Serv\_End2* respectively) fires. As several servers may be busy at the same time, the semantics of these two transitions is *infinite servers*. After completion of service, the customer returns to state free (one token in place *Cus\_Free*) and the server becomes available (one token is put in place *Ser\_Idle1* or *Ser\_Idle2*, according to the class to which the server belongs).

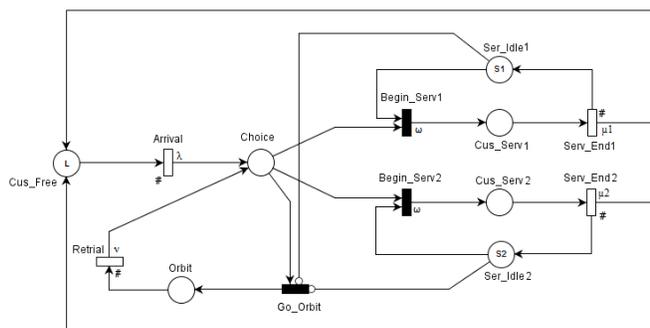


Figure 1. GSPN Model of finite-source retrial networks with two servers classes

#### IV. STOCHASTIC ANALYSIS

Initially, the orbit is empty, all customers are free and all servers are available. Thus the initial marking can be

expressed in this form:

$$\begin{aligned} M_0 &= \{M(\text{Cus\_Free}), M(\text{Choice}), M(\text{Orbit}), \\ &\quad M(\text{Ser\_Idle1}), M(\text{Cus\_Serv1}), M(\text{Ser\_Idle2}), \\ &\quad M(\text{Cus\_Serv2})\} \\ &= \{L, 0, 0, S_1, 0, S_2, 0\} \end{aligned}$$

Whatever the values of  $L$ ,  $S_1$  and  $S_2$ , the conservation of customers and servers of the two classes, gives the following equations:

$$\begin{cases} M(\text{Ser\_Idle1}) + M(\text{Cus\_Serv1}) = S_1 \\ M(\text{Ser\_Idle2}) + M(\text{Cus\_Serv2}) = S_2 \\ M(\text{Cus\_Free}) + M(\text{Cus\_Serv1}) \\ + M(\text{Cus\_Serv2}) + M(\text{Orbit}) = L \end{cases} \quad (1)$$

Observing these three equations, we note that the system state at steady-state can be described by means of three variables  $(i, j, k)$ , which we called a *micro-state* where :

- $i$  represents the number of customers being served by servers of class  $C_1$  (in place *Cus\_Serv1*);
- $j$  represents the number of customers being served by servers of class  $C_2$  (in place *Cus\_Serv2*);
- and  $k$  is the number of customers in orbit (in place *Orbit*).

Hence, having the micro-state  $(i, j, k)$ , the markings of all places can be obtained, since  $M(\text{Ser\_Idle1}) = S_1 - i$ ,  $M(\text{Ser\_Idle2}) = S_2 - j$  and  $M(\text{Cus\_Free}) = L - (i + j + k)$ .

On the other hand, applying (1), we can deduce:

$$\begin{cases} 0 \leq i \leq S_1 \\ 0 \leq j \leq S_2 \\ 0 \leq k \leq L - (S_1 + S_2) \end{cases} \quad (2)$$

The CTMC corresponding to the proposed GSPN contains  $n$  micro-states corresponding to the accessible tangible markings, where  $n$  equals  $(S_1 + 1) \cdot (S_2 + 1) \cdot [L + 1 - S]$  and  $S = S_1 + S_2$ .

Thus, the corresponding infinitesimal generator  $Q$  is a  $n \times n$  matrix, which is defined by :

$$\begin{cases} Q[(i, j, k), (x, y, z)] = \theta[(i, j, k), (x, y, z)] \\ Q[(i, j, k), (i, j, k)] = -\sum_{(l, m, n) \neq (i, j, k)} \theta[(i, j, k), (l, m, n)] \end{cases}$$

where  $\theta[(i, j, k), (x, y, z)]$  is the transition rate from state  $(i, j, k)$  to state  $(x, y, z)$ .

By analyzing the micro-states and the transitions rates of the CTMC, we obtain the following rates:

- $[0 \leq i \leq S_1 - 1, 0 \leq j \leq S_2 - 1] :$   
 $(i, j, k) \xrightarrow{\frac{1}{2}(L-i-j-k)\lambda} (i+1, j, k)$   
 and  $(i, j, k) \xrightarrow{\frac{1}{2}(L-i-j-k)\lambda} (i, j+1, k)$

- $[0 \leq i \leq S_1 - 1] : (i, S_2, k) \xrightarrow{(L-i-S_2-k)\lambda} (i+1, S_2, k),$
- $[0 \leq j \leq S_2 - 1] : (S_1, j, k) \xrightarrow{(L-S_1-j-k)\lambda} (S_1, j+1, k),$
- $[0 \leq k < L - (S_1 + S_2)] : (S_1, S_2, k) \xrightarrow{(L-S-k)\lambda} (S_1, S_2, k+1),$
- $[i > 0] : (i, j, k) \xrightarrow{i\mu_1} (i-1, j, k),$
- $[j > 0] : (i, j, k) \xrightarrow{j\mu_2} (i, j-1, k),$
- $[0 \leq i \leq S_1 - 1, 0 \leq j \leq S_2 - 1, k > 0] : (i, j, k) \xrightarrow{\frac{1}{2}k\nu} (i+1, j, k-1) \text{ and } (i, j, k) \xrightarrow{\frac{1}{2}k\nu} (i, j+1, k-1),$
- $[0 \leq i \leq S_1 - 1, k > 0] : (i, S_2, k) \xrightarrow{k\nu} (i+1, S_2, k-1),$
- $[0 \leq j \leq S_2 - 1, k > 0] : (S_1, j, k) \xrightarrow{k\nu} (S_1, j+1, k-1),$

As a consequence, the infinitesimal generator can be automatically calculated by means of Algorithm 1 given below.

### V. PERFORMANCE MEASURES

The aim of this section is to derive the formulas of the most important stationary performance. As the proposed models are bounded and the initial marking is a home state, the underlying process is ergodic. Hence, the steady-state solution exists and is unique. The infinitesimal generator  $Q$  corresponding to the GSPN model can be obtained automatically by applying the above algorithm. Then, the steady-state probability vector  $\pi$  can be computed by solving the linear system of equations:

$$\begin{cases} \pi.Q = 0 \\ \sum_i \pi_i = 1 \end{cases} \quad (3)$$

Where  $\pi_i$  denotes the steady-state probability that the process is in state  $M_i$ .

Having the probability distribution  $\pi$ , we can derive several exact stationary performance measures of finite-source retrial networks with two classes of servers applying the formulas given below. In following,  $M_i(p)$  indicates the number of tokens in place  $p$  in marking  $M_i$ ,  $A$  is the set of reachable tangible markings, and  $A(t)$  is the set of tangible markings reachable by transition  $t$ .

- Mean number of customers in the orbit: This corresponds to the mean number of tokens in *Orbit*,

$$n_{Orb} = \sum_{i: M_i \in A} M_i(Orbit) \cdot \pi_i = \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} k \cdot \pi_{i,j,k}$$

- Mean number of busy servers of class  $C_1$ : Note that this is also the mean number of customers under service by Class  $C_1$ , it corresponds to the mean number of tokens in place *Cus\_Serv1*,

$$n_{busyC_1} = \sum_{i: M_i \in A} M_i(Cus\_Serv1) \cdot \pi_i = \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} i \cdot \pi_{i,j,k}$$

---

### Algorithm 1 Infinitesimal Generator Construction

---

```

1: for k ← 0, L - S do
2:   for i ← 0, S1 - 1 do
3:     for j ← 0, S2 - 1 do
4:       Q[(i, j, k), (i+1, j, k)] ← 1/2(L-i-j-k)λ
5:       Q[(i, j, k), (i, j+1, k)] ← 1/2(L-i-j-k)λ
6:       Q[(S1, j, k), (S1, j+1, k)] ← (L-S1-j-
           k)λ
7:     end for
8:     Q[(i, S2, k), (i+1, S2, k)] ← (L-i-S2-k)λ
9:   end for
10: end for
11: for k ← 0, L - S - 1 do
12:   Q[(S1, S2, k), (S1, S2, k+1)] ← (L-S-k)λ
13: end for
14: for k ← 0, L - S do
15:   for i ← 1, S1 do
16:     for j ← 0, S2 do
17:       Q[(i, j, k), (i-1, j, k)] ← iμ1
18:     end for
19:   end for
20:   for j ← 1, S2 do
21:     for i ← 0, S1 do
22:       Q[(i, j, k), (i, j-1, k)] ← jμ2
23:     end for
24:   end for
25: end for
26: for k ← 1, L - S do
27:   for i ← 0, S1 - 1 do
28:     for j ← 0, S2 - 1 do
29:       Q[(i, j, k), (i+1, j, k-1)] ← 1/2.kν
30:       Q[(i, j, k), (i, j+1, k-1)] ← 1/2.kν
31:     end for
32:     Q[(i, S2, k), (i+1, S2, k-1)] ← kν
33:   end for
34:   for j ← 0, S2 - 1 do
35:     Q[(S1, j, k), (S1, j+1, k)] ← kν
36:   end for
37: end for
    
```

---

- Mean number of busy servers of class  $C_2$ :

$$n_{busyC_2} = \sum_{i: M_i \in A} M_i(Cus\_Serv2) \cdot \pi_i = \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} j \cdot \pi_{i,j,k}$$

- Mean number of busy servers: It corresponds to the sum of busy servers in both of the 2 classes

$$n_{busy} = n_{busyC_1} + n_{busyC_2} = \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} (i+j) \cdot \pi_{i,j,k}$$

- Mean number of customers in the network: Which is the total number of the mean number of customers in the orbit and those under service (by  $C_1$  and  $C_2$ ),

$$n = n_{Orb} + n_{busy} = \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} (i + j + k) \cdot \pi_{i,j,k}$$

- Mean number of free servers: This corresponds to the sum of the mean number of free servers of both classes:

$$n_{Free} = S - n_{busy}$$

- Effective customer arrival rate: This represents the throughput of the transition *Arrival*,

$$\begin{aligned} \bar{\lambda} &= \sum_{i: M_i \in A(Arrival)} \lambda \cdot M_i(Cus\_Free) \cdot \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} \lambda \cdot (L - i - j - k) \cdot \pi_{i,j,k} \\ &= \lambda \cdot n_{CusFree} \end{aligned}$$

- Effective customer retrial rate: It corresponds to the throughput of *Retrial* transition,

$$\begin{aligned} \bar{\nu} &= \sum_{i: M_i \in A(Retrial)} \nu \cdot M_i(Orbit) \cdot \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} \nu \cdot k \cdot \pi_{i,j,k} = \nu \cdot n_{Orb} \end{aligned}$$

- Mean service rate of class  $C_1$ : This corresponds to the throughput of the transition *Serv\_End1*,

$$\begin{aligned} \bar{\mu}_1 &= \sum_{i: M_i \in A(Serv\_End1)} \mu_1 \cdot M_i(Cus\_Serv1) \cdot \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} \mu_1 \cdot i \cdot \pi_{i,j,k} = \mu_1 \cdot n_{busyC1} \end{aligned}$$

- Mean service rate of class  $C_2$ : This corresponds to the throughput of *Serv\_End2*,

$$\begin{aligned} \bar{\mu}_2 &= \sum_{i: M_i \in A(Serv\_End2)} \mu_2 \cdot M_i(Cus\_Serv2) \cdot \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0}^{S_2} \mu_2 \cdot j \cdot \pi_{i,j,k} = \mu_2 \cdot n_{busyC2} \end{aligned}$$

- Availability of  $s$  servers in the system (among both classes):

$$\begin{aligned} A_s &= \sum_{i: M_i(Ser\_Idle1) + M_i(Ser\_Idle2) \geq s} \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0, i+j \leq S-s}^{S_2} \pi_{i,j,k} \end{aligned}$$

	Homogeneous case	Two servers classes system
Number of servers	4	$S_1=1, S_2=3$
Size of source	20	20
Primary call generation rate	0.1	0.1
Service rate	1	$\mu_1=1, \mu_2=1$
retrial rate	1.2	1.2
Mean number of busy servers	1.800 748	$C_1: 0.521 865$ $C_2: 1.278 883$ Total : 1.800 748
Mean number of source in the orbit	0.191 771	0.191 771
Mean primary call generation rate	1.800 748	1.800 748
Mean waiting time	0.106 495	0.106 495

Table II  
VALIDATION IN THE HOMOGENEOUS CASE

- Utilization of at least  $s$  servers: This corresponds to the probability that at least  $s$  servers among the two classes are busy:

$$\begin{aligned} U_s &= \sum_{i: M_i(Cus\_Serv1) + M_i(Cus\_Serv2) \geq s} \pi_i \\ &= \sum_{k=0}^{L-S} \sum_{i=0}^{S_1} \sum_{j=0, i+j \geq s}^{S_2} \pi_{i,j,k} \end{aligned}$$

- The mean waiting time: It's the mean period between the arrival of the customer and its service beginning. Using the Little's formula, the mean waiting time is given by:

$$\bar{W} = \frac{n_{Orb}}{\bar{\lambda}}$$

- The mean response time:

$$\bar{R} = \frac{n}{\bar{\lambda}}$$

## VI. VALIDATION AND NUMERICAL EXAMPLES

In order to test the feasibility of our approach, we developed a *C#* code to implement the above algorithm (1) and the performance indices formulas. Next, we tested it for a large number of examples. In particular, in the homogeneous case, the results were validated by the Pascal program given in [7]. From table (II), we can see that both models give exactly the same results up to the 6th decimal digit.

	$L$	$S_1$	$S_2$	$\lambda$	$\nu$	$\mu_1$	$\mu_2$
Figure 2	50	4	2	0.5	x axis	8	2
Figure 3	30	x axis	4	2	1	6	1
Figure 4	30	4	x axis	2	1	6	1

Table I  
INPUT SYSTEM PARAMETERS

In the following, we present sample numerical results to illustrate graphically the impact of different system parameters on the mean response time. The input parameters are collected in table I.

The figure (2) shows the sensitivity of the mean response time to the retrial generation rate. Indeed, the response time decreases with the intensity of the flow of repeated calls, particularly when the retrial intensity is low (between 0.01 and 0.3), beyond the value 0.3, the influence becomes less significant.

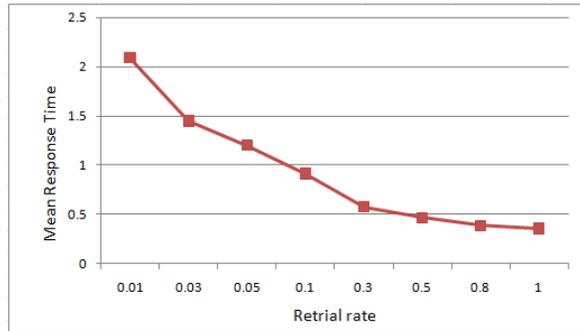


Figure 2. Mean response time versus retrial generation rate.

In figure 3, (4 respectively), we show the influence of the number of servers of  $C_1$  ( $C_2$  respectively) class, on the mean response time. We conclude that this last decreases with the increase of the servers number. However, the rate of influence of the number of servers in the  $C_1$  class is faster than the influence due to increasing the number of servers of  $C_2$  because the former is faster ( $\mu_1 = 6$  vs  $\mu_2 = 1$ ). In figure (3), the response time reached the optimum and stabilizes after a certain time (number of servers = 12). Hence, it is not interesting to invest in new servers in the  $C_1$  class.

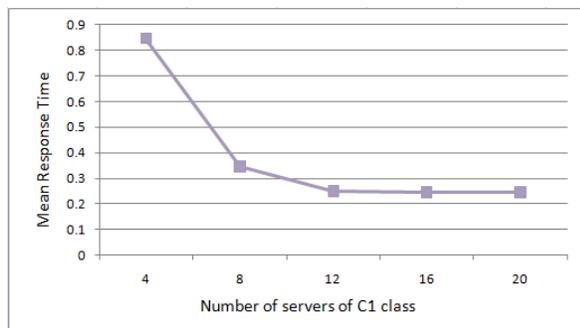


Figure 3. Mean response time versus  $C_1$  class servers number.

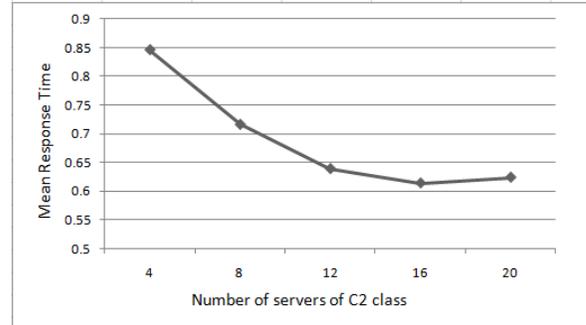


Figure 4. Mean response time versus  $C_2$  class servers number.

REFERENCES

[1] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis, *Modelling with Generalized Stochastic Petri Nets*, John Wiley & Sons, New York, 1995.

[2] J.R. Artalejo and A. Gómez-Corral, *Retrial Queueing Systems: A Computational Approach*, Springer, Berlin, 2008.

[3] J.R. Artalejo, *Accessible bibliography on retrial queues: Progress in 2000-2009*, *Mathematical and Computer Modelling*, vol. 51, pp. 1071-1081, 2010.

[4] J.R. Artalejo and M.J. LOPEZ-HERRERO, *Cellular mobile networks with repeated calls operating in random environment*, *Computers & operations research*, vol. 37, no7, pp. 1158-1166, 2010.

[5] M. Diaz, *Les réseaux de Petri - Modèles Fondamentaux*, Paris, Hermès Science Publications, 2001.

[6] D. Efosinin and L. Breuer, *Threshold policies for controlled retrial queues with heterogeneous servers*, *Annals of Operations Research*, vol. 141, pp. 139-162, 2006.

[7] G.I. Falin and J.G.C. Templeton, *Retrial Queues*, Chapman and Hall, London, 1997.

[8] N. Gharbi, C. Dutheillet, and M. Ioualalen, *Colored Stochastic Petri Nets for Modelling and Analysis of Multiclass Retrial Systems*, *Mathematical and Computer Modelling*, vol. 49, pp. 1436-1448, 2009.

[9] J. Roszik and J. Sztrik, *Performance analysis of finite-source retrial queues with nonreliable heterogeneous servers*, *Journal of Mathematical Sciences*, vol. 146, pp. 6033-6038, 2007.

[10] J. Sztrik, G. Bolch, H. de Meer, J. Roszik, and P. Wuechner, *Modeling finite-source retrial queueing systems with unreliable heterogeneous servers and different service policies using MOSEL*, *Proc. of 14th Inter. Conf. on Analytical and Stochastic Modelling Techniques and Applications ASMTA'07*, Pargue, Czech Republic, pp. 75-80, 2007.

[11] Tien Van Do, *A new computational algorithm for retrial queues to cellular mobile systems with guard channels*, *Computers & Industrial Engineering*, vol. 59, pp. 865-872, 2010.

[12] P. Tran-Gia and M. Mandjes, *Modeling of customer retrial phenomenon in cellular mobile networks*, *IEEE Journal on Selected Areas in Communications*, vol. 15, pp. 1406-1414, 1997.

## Performance Analysis of an Ant-based Routing Algorithm with Enhanced Path Maintenance for MANETs

Miae Woo

Dept. of Information and Communications Eng.  
Sejong University  
Seoul, Korea  
e-mail: mawoo@sejong.ac.kr

**Abstract**— Ant-based routing algorithms belong to a class of ant colony optimization which applies the behavior of ants in nature to routing mechanism. Since ant-based routing algorithms provide high adaptability to the dynamic network topology, it is suitable for routing in mobile ad-hoc network (MANET). In this paper, we introduce a routing method, namely EPMAR (ant-based routing algorithm using enhanced path maintenance), which enhances route selection method and the process upon link failure of EAR. We then compare the performance of ant-based routing algorithms, AntHocNet, EAR, and EPMAR, how they perform as the packet transmission rate varies. The simulation results show that EPMAR provided higher packet delivery ratio and less critical link failure than AntHocNet and EAR.

**Keywords**—Routing, Mobile Ad-Hoc Network, Ant Colony Optimization, Performance Analysis

### I. INTRODUCTION

An ant-based routing algorithm is inspired by the behavior of ants in nature, which uses swarm intelligence [1]. Ant-based routing is an application area of ant colony optimization (ACO) [2] which uses the behavior of an ant colony to optimize the given problem in a distributed fashion. Ants use chemical substance, called pheromone, to share information with other ants in their colony about the paths between the nest and food sources. Ants choose a path with highest pheromone deposit over other paths to the food source, and lay pheromone on the way back to the nest. Consequently, the shortest path would have the most deposited pheromone among paths to the food source. Such a problem solving method is applied to the routing problems in networks.

Characteristics of ant-based routing methods such as agent systems and capability of multi-path routing make ant-based routing methods suitable for mobile ad-hoc network (MANET) environments [3]. Routing methods based on agent systems allow high adaptability to the dynamic network topology. Since paths in MANETs fail frequently, ant-based routing methods can provide stability in the connections between the source and the destination using multiple paths.

Usually, ant-based routing algorithms consist of four phases: route setup, data transfer, route maintenance, and route recovery phases. Since mobile nodes move around, network topology in MANET changes as time passes. As a result, an optimal path may become non-optimal. To handle such situations, route maintenance is required to discover new or better paths [2].

There are several ant-based routing algorithms proposed especially for MANETs [3-10]. Among them, AntHocNet provides good performance compared to ad hoc on-demand vector routing (AODV) in terms of data delivery ratio and end-to-end delay [7]. Its performance in a realistic urban environment has been analyzed in [8]. However, the overhead of AntHocNet generated by the ants are quite high. To overcome such a drawback of AntHocNet, an efficient ant-based routing algorithm (EAR) was proposed in [9] and its performance was further investigated in [10]. EAR introduced several features in the route set-up phase to decrease the overhead introduced by ants and to efficiently update pheromone values in all the intermediate nodes along the path.

In this paper, we introduce an ant-based routing algorithm, namely EPMAR (ant-based routing algorithm using enhanced path maintenance) [11], which modifies data transfer phase and route recovery phase of AntHocNet and EAR. EPMAR uses procedures of EAR for route setup and route maintenance phases. EPMAR aims to cope with link failure more efficiently than AntHocNet and EAR. The performance of EPMAR is investigated and compared to those of EAR and AntHocNet to see how these algorithms perform when the transmission rates of constant bit rate (CBR) sources vary. The simulation results show that EPMAR provide higher packet delivery ratio and less critical link failure than AntHocNet and EAR.

The remainder of this paper is organized as follows. Section 2 describes the operation of EPMAR. In Section 3, the simulation environment to measure the performance of the EPAMR, AntHocNet and EAR is discussed. The simulation results are analyzed in Section 4. Finally, Section 5 gives the conclusion.

## II. DESCRIPTION OF EPMAR

EPMAR consists of four phases, which are route setup, data transfer, route maintenance, and route recovery phases. EPMAR uses same route setup and route maintenance procedures of EAR [9] and modifies data transfer phase and route recovery phase of EAR.

### A. Route Set-up Phase

The procedure of route set-up phase of EPMAR is same as that of EAR [9]. Route set-up phase starts when the source node wants to send a data packet to the destination node. If the source does not have any routing information to the destination, it broadcasts a reactive forward ant (RFA) to probe paths. A reactive forward ant contains several fields: source and destination addresses, generation number, trip time, a list of visited nodes, the number of visited nodes, and a flag for reactive backward ant generation at the intermediate node. One broadcasted ant can produce several ants because of the broadcasting mechanism along the route. Those ants have same source address, destination address, and generation number. Such ants are called as same generation ants. Same generation ants may have different values in the fields of trip time, list of visited nodes, number of visited nodes, and flag for reactive backward ant generation at the intermediate node. The values in these fields are updated as the reactive forward ant travels towards the destination.

If an intermediate node receives a reactive forward ant, it checks whether it already received any other same generation ant. If so, it discards the reactive forward ant which it just received. Otherwise, the intermediate node checks whether it has any routing information to the destination in its routing table. If it does not have any information, it saves the values of source address, destination address, generation number, trip time, and list of visited nodes in its routing table. It then updates trip time, list of visited nodes, and number of visited nodes in the reactive forward ant, and broadcasts the ant.

If the intermediate node has routing information to the destination which satisfies some criteria, it can generate a reactive backward ant. The criteria are as follows. First, any previously visited node en route to this intermediate node had not generated a backward ant. The routing information in this intermediate node should be fresh enough. We regard the routing information which was updated within 10 seconds as fresh. Finally, the intermediate node generating a backward ant should be close enough to the source and far enough from the destination. This condition is checked by looking at the hop distances from the source and to the destination. We use 10-hop distance is the maximum hop distance between the source and the intermediate node. Also the minimum hop distance between the intermediate node and the destination is set to 5 hops. If the intermediate node satisfies these conditions, it generates a reactive backward ant (RBA) towards the source. The intermediate node sets the flag for reactive backward ant generation to indicate that

the backward ant is not generated from the destination. When the source receives a reactive backward ant, it can decide who sends the backward ant by inspecting the flag for reactive backward ant generation.

The intermediate node unicasts the reactive forward ant to the next hop probabilistically. Before unicasting the ant, the intermediate node updates the fields and sets the flag for the reactive backward ant generation at the intermediate node in the reactive forward ant.

The destination node can accept several same generation ants and generate reactive backward ants as many as the accepted forward ants to form multi-path. There is a limit for the number of acceptable same generation ants. We only accept at most 3 ants among same generation ants.

The reactive forward ants accepted at the destination pass the list of visited nodes to the reactive backward ants. The reactive backward ants travel towards the sources by backtracking the nodes in the list of visited nodes.

The reactive backward ant calculates the overall trip times from the currently visited node to all the nodes on the path to the destination, and estimates transmission time required at the MAC layer by taking queue length and the average delay into account. It uses the calculated time to update pheromone values in the routing table [7]. Let  $T_{nj}^i$  be the pheromone value of the path from node  $i$  to node  $j$  through the neighbor node  $n$ .  $T_{nj}^i$  is calculated as a running average using the following equation.

$$T_{nj}^i = \gamma T_{nj}^i + (1 - \gamma) \left( \frac{1}{\hat{T}_j^i + h T_{hop}} \right)$$

In the equation,  $h$  is the hop distance between node  $i$  and node  $j$ .  $T_{hop}$  is the time required to deliver one packet to the next hop in unloaded condition.  $\hat{T}_j^i$  is the estimate of trip time from node  $i$  to node  $j$ . It is the sum of local estimates  $\hat{T}_{i+1}^i$  to reach next hop  $i+1$  in each node along the path from node  $i$  to node  $j$ .  $\hat{T}_{i+1}^i$  is calculated using following equation.

$$\hat{T}_{i+1}^i = \alpha \hat{T}_{i+1}^i + (1 - \alpha)(Q_{mac}^i + 1) \hat{T}_{mac}^i$$

where  $Q_{mac}^i$  is the number of packets in the queue to be sent at the MAC layer and  $\hat{T}_{mac}^i$  is the running average of estimate of the average time to send one packet at the MAC layer.

### B. Data Transfer Phase

After setting-up paths to the destination, data packets are forwarded based on the pheromone values in the routing table. AntHocNet and EAR select the next hop stochastically based on the pheromone value in the routing table to spread data load on multiple paths. However, in the MANET environment where link failures are frequent, it is not confirmed yet how load distribution has effect on the data transfer efficiency. Also, it is uncertain that routes with

low pheromone value can actually provide stable paths to the destination.

Thus, EPMAR chooses a path with highest pheromone value among multiple paths in the routing table. Let  $N_d^i$  be the set of neighbor nodes of node  $i$  through which paths exist to the destination  $d$  in the routing table. Then, a neighbor node which provides the highest pheromone value to the destination is selected as a next hop. In other words, the selection criteria of next hop is

$$\max_{n \in N_d^i} \{T_{nd}^i\}.$$

For the data transfer, multiple paths are used as backup paths when link failure along the best path occurs.

### C. Route Maintenance Phase

Route maintenance phase of EPMAR is same as AntHocNet [7] and EAR [9]. To maintain the established paths and to find better or alternative paths, a source node periodically dispatches proactive forward ants (PFAs) at the rate according to the data sending rate [7]. A proactive forward ant can be either unicasted probabilistically or broadcasted. Generally, a proactive forward ant chooses the next hop probabilistically to probe an established path. It collects up-to-date information about the established path and updates the pheromone values of the path by the corresponding proactive backward ants (PBAs). A proactive forward ant is broadcasted with a small probability at the intermediate node to explore a new or alternative path. If a node receives a proactive forward ant but it does not have any routing information to the destination, it broadcasts the proactive forward ant. Total number of broadcast allowed through the path toward the destination is limited to control the overhead.

### D. Route Recovery Phase

As in AntHocNet and EAR, EPMAR considers two situations as link failures. One is failure in receiving hello messages from a neighbor node and the other is failure in transmission of data packets.

If a node does not receive a hello message from its neighbor for a certain amount of time, the link is considered broken. In this case, the node removes the associated entries in its routing table, and broadcasts a link failure notification message. All the neighbors receiving the notification message update their routing table. If any one of them lost its best or the only path to the destination due to the link failure, it rebroadcasts the notification.

If transmission of a data packet is failed and there is no other path available for the data packet, then the node tries to repair the path locally by broadcasting a forward route repair ant (FRRA). If any backward route repair ant is not received within a certain time period, the node discards all the temporally buffered packets and broadcasts a link failure notification about the lost destination.

The forward route repair ant is perished at the intermediate node if it has any alternative path to the

destination. As a result, the source of data packets cannot recognize that the link failure happened in the path to the destination. According to the simulation results obtained in the previous work [9, 10], the ratio of route recovery by using forward route repair ant was low.

EPMAR augments the procedure in route recovery phase by sending unicast link failure message to the source of the data packet which cannot be forwarded due to link failure. The overhead incurred by this message is low because it is delivered by unicasting. When the source receives the unicast link failure message, it can recognize that the best path to the destination has been failed. To obtain up-to-date pheromone value of a path to the destination, the source dispatches a proactive forward ant. The procedure for the proactive forward ant in this case is same as that in the route maintenance phase. By doing this, the source node can update the pheromone value, and it can use the route with the highest pheromone value for the data transfer.

## III. SIMULATION ENVIRONMENT

To evaluate the performance of EPMAR, we ran simulations using Qualnet. In the area of  $3000 \times 1000 \text{ m}^2$ , 100 mobile nodes were randomly placed. A rectangular space was chosen to force the use of longer routes between nodes than those would occur in a square space with equal node density [12]. Simulation time is set to 300 seconds. By selecting traffic sources and destinations randomly, 30 connections were established. Four different transmission rates were used: 1 packet/sec, 2 packets/sec, 5 packets/sec, and 10 packets/sec. Each source generated 64-byte long CBR packets. Data transmission for each connection was started by selecting random delay from uniform distribution in [0, 60] seconds. Data transmissions were continued till the end of simulation.

For the physical layer, two-ray signal propagation model was used. The radio propagation range of each node was set to 300 meters. For the MAC layer, IEEE 802.11b protocol was used with 2 Mbps bandwidth.

The random waypoint mobility model was used for the node movement model [13]. Random waypoint model defines the mobility pattern of nodes by pause time and the maximum node speed. In Qualnet, each node began the simulation by selecting a random destination in the given space and moved to that destination at a speed distributed uniformly between 0 and some maximum node speed. Upon reaching the destination, the node paused for the specified pause time. It then selected another destination, and proceeded from there as previously described. Each node repeated this behavior for the simulation time. In the simulation, maximum node speed was set 20 m/sec. Also, 5 different pause times, 0, 30, 60, 120, and 300 seconds were used.

## IV. PERFORMANCE ANALYSIS

In order to compare the performance of EPMAR with AntHocNet and EAR, we investigated packet delivery ratio,

the number of critical link failures, the number of generated forward ants, number of forwarded ants and backward ants per node in average and path set-up time.

Figure 1 shows the average packet delivery ratio. As we can see in Figure 1, EPMAR gave the highest packet delivery ratio among three algorithms regardless of packet transmission rate. EPMAR provided about 16% better packet delivery ratio than AntHocNet and 4% better packet delivery ratio than EAR in average. Especially, as the packet transmission rate was increased to 10 packets per second, EPMAR provided 26% and 7% more packet delivery ratio than AntHocNet and EAR respectively.

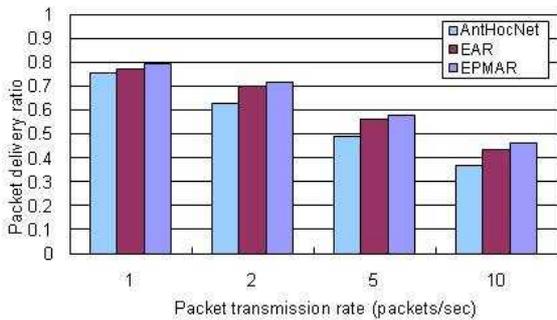


Figure 1. Average packet delivery ratio

The number of critical link failures occurred during the simulation time is shown in Figure 2. Clearly, we can see that least critical link failures occurred when EPMAR was used as a routing algorithm. EPMAR resulted 29% and 10% less critical link failures than in AntHocNet and EAR respectively.

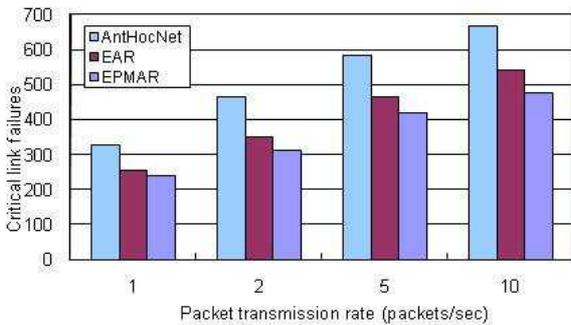


Figure 2. Number of critical link failures occurred

Through the results of the packet delivery ratio and critical link failure shown in Figure 1 and Figure 2, it is proven that EPMAR obtained the intended goals. In other words, packet delivery ratio was improved by choosing the path with the highest pheromone deposit. Also, informing

link failure situation by sending unicast link failure messages to the source could reduce additional link failures.

However, introduction of unicast link failure message increased overhead in the system. Clearly, delivery of unicast link failure message to the source adds up overhead. Upon receiving the unicast link failure messages, dispatching proactive forward ants is triggered at the source. The destination who receives proactive forward ants should respond with proactive backward ants. These proactive forward ants and backward ants are the added overhead in the EPMAR compared to EAR or AntHocNet. The related results are shown in Figure 3, 4, and 5.

Figure 3 depicted the number of forward ants generated at the 30 sources during the simulation. The counted forward ants include reactive forward ants (RFAs), proactive forward ants (PFAs), and forward route repair ants (FRRAs) which are generated due to link failure. Overall, as we can see in the figure, there is no big difference in the number of forward ants generated among three methods. EPMAR generated the least reactive forward ants and forward route repair ants. However, as we can expect, EPMAR generated the most proactive forward ants.

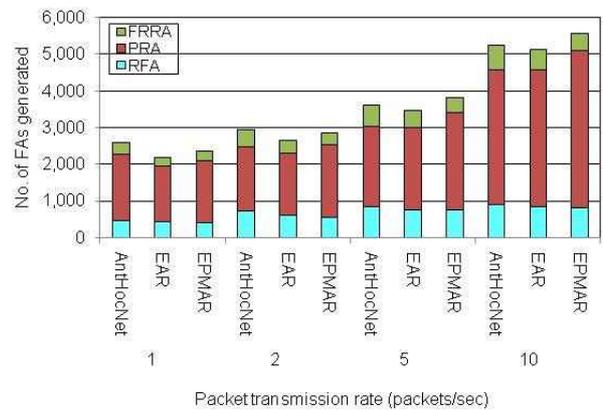


Figure 3. Number of generated forward ants

The generated forward ants are delivered to the destination through intermediate nodes either by unicasting or by broadcasting depending on the type of ants. Figure 4 shows the average number of forwarded ants per nodes. The shown number summed all the number of generations at the sources and forwards at the intermediate nodes and averaged out by the total number of nodes in the system. Although, EPMAR generated the most forward ants among three methods when the transmission rate were 5 packets/sec and 10 packets/sec, it provided the least forwarding in the overall system. The average number of forwarded ants per node of EPMAR was 31% of AntHocNet and 97% of EAR.

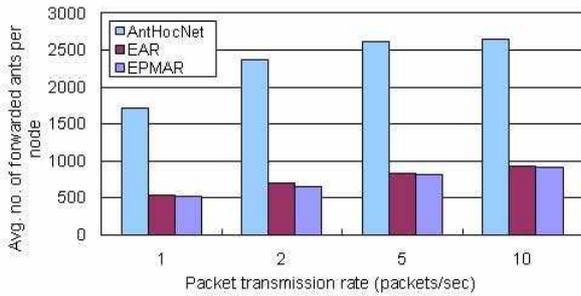


Figure 4. Average number of forwarded ants per node

On the other hand, EPMAR produced the most backwarded ants per node as shown in Figure 5. EPMAR forwarded 57% more backward ants than EAR and 21% more backward ants than AntHocNet. Reactive forward ants and proactive forward ants trigger dispatching the corresponding backward ants after they arrived at the destinations. Since the objective of forward route repair ants is recovery of local paths, the time-to-live value of the forward route repair ants is set to relatively smaller than those of reactive and proactive forward ants. As a result, there might be many cases which did not result the corresponding backward ants. Actually, local route repair ratio was less than 20% in the simulation results. Consequently, EPMAR produced more backward ants because it generated more proactive forward ants as shown in Figure 3. When compared with EAR, EPMAR forwarded at most 70 more backward ants and 20 less forward ants. Thus, the increased overhead was not big.

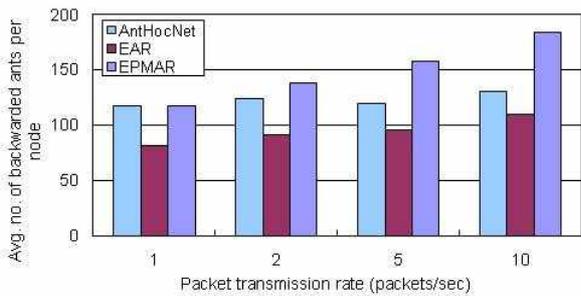


Figure 5. Average number of backwarded ants per node

Figure 6 shows the path set-up time of three routing methods. Since EPMAR and EAR uses the same route set-up procedure, path set-up time of these methods is almost same. Compared with the path set-up time of AntHocNet, EPMAR and EAR reduced the path set-up time about 3/10 by generating RBA at the intermediate nodes.

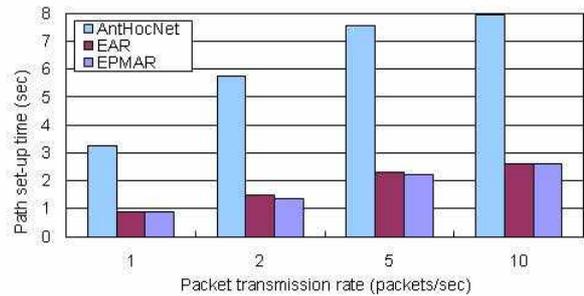


Figure 6. Path set-up time

V. CONCLUSION

In this paper, we introduced an ant-based routing algorithm with enhanced path maintenance, namely EPMAR. The objective of EPMAR was set to increase the performance by choosing the best path for the data delivery and to reduce the critical link failures. The performance of EPMAR was compared with that of AntHocNet and EAR. The simulation results showed that EPMAR provided better packet delivery ratio and less critical link failures than AntHocNet and EAR. Though EPMAR introduced a new message called unicast link failure message to reduce critical link failures, the control overhead was comparable to that of EAR and less than AntHocNet.

ACKNOWLEDGMENT

This research is supported by the ubiquitous Computing and Network (UCN) Project, the Ministry of Knowledge and Economy(MKE) Knowledge and Economy Frontier R&D Program in Korea as a result of UCN's subproject 11C3-C2-10M.

REFERENCES

- [1] M. G. Hinchey, R. Sterritt, and C. Rouff, "Swarms and Swarm Intelligence," *IEEE Computer*, Vol. 40, No. 4, pp. 111-113, April 2007.
- [2] K. M. Sim and W. H. Sun, "Ant Colony Optimization for Routing and Load-Balancing: Survey and New Directions," *IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 33, No. 5, pp. 560-572, Sep. 2003.
- [3] M. Güneş, U. Sorges, and I. Bouazzi, "ARA-the ant-colony based routing algorithm for MANETs," in *Proc. of IWAHN 2002*, pp. 79-85, August 2002.
- [4] H. Matsuo and K. Mori, "Accelerated Ants Routing in Dynamic Networks," *2nd International Conf. on Software Engineering, Artificial Intelligence, Networking & Parallel/Distributed Computing*, pp. 333-339, 2001.
- [5] S. Marwaha, C. K. Tham, and D. Srinivasan, "Mobile agents based routing protocol for mobile ad hoc networks," in *Proc. of IEEE ICON*, pp. 27-30, August 2002.
- [6] J. S. Baras and H. Mehta, "A Probabilistic Emergent Routing Algorithm for Mobile Ad Hoc Networks," in *Proc. of WiOpt03*, 2003.
- [7] G. Di Caro, F. Ducatelle, and L. M. Gambardella, "AntHocNet: An Adaptive Nature-Inspired Algorithm for Routing in Mobile Ad Hoc Networks," *Tech. Rep. No. IDSIA-27-04-2004*, IDSIA/USI-SUPSI, Sep. 2004.

- [8] F. Ducatelle, G. Di Caro and L. M. Gambardella, "An Evaluation of Two Swarm Intelligence MANET Routing Algorithms in an Urban Environment," in Proc. of IEEE Swarm Intelligence Symposium, Sep. 2008.
- [9] Miae Woo, Ngo Huu Dung, Woo Jong Roh, "An Efficient Ant-based Routing Algorithm for MANETs," in Proc. ICACT 2008, Feb. 2008.
- [10] Myungjun Oh, Ngo Huu Dung, and Miae Woo, "Effect of Transmission Rates on the Performance of Ant-Based Routing Algorithms for MANETs," in Proc. UBICOMM 2009, pp. 238-243, Oct. 2009.
- [11] M. Woo, "An Ant-based Routing Method using Enhanced Path Maintenance for MANETs," Journal of Korea Information and Communications Society, Vol. 35, No. 9, pp. 1281-1286, Sep. 2010.
- [12] Broch, D. A. Maltz, D. B. Johnson, Y. Hu and J. Jetcheva, "A Performance Comparison of Multi-Hop Ad Hoc Network Routing Protocols," in Proc. MobiCom'98, Oct. 1998.
- [13] D. B. Johnson and D. A. Maltz, "Dynamic source routing in ad hoc wireless networks," in Mobile Computing, edited by T. Imielinski and H. Korth, Kluwer Academic Publishers, pp. 153-181, 1996.

# Adaptive Cross Layer Approach for Video Transmission Over Cognitive UWB Network

Norazizah Mohd Aripin<sup>1,2</sup>, Norsheila Faisal<sup>1</sup>, Rozeha A. Rashid<sup>1</sup>, A.C.C Lo<sup>3</sup>

Faculty of Electrical Engineering, Universiti Teknologi Malaysia, Johor, Malaysia<sup>1</sup>

Department of Electronics & Communication Engineering, Universiti Tenaga Nasional, Selangor, Malaysia<sup>2</sup>

Wireless & Mobile Communication Groups, Delft University of Technology, Delft, The Netherlands<sup>3</sup>

[norazizahm@gmail.com](mailto:norazizahm@gmail.com), [sheila|rozeha}@fke.utm.my](mailto:sheila|rozeha}@fke.utm.my), [A.C.C.Lo@tudelft.nl](mailto:A.C.C.Lo@tudelft.nl)

**Abstract**— The demand of quality of service (QoS) for multimedia transmission over wireless network raises huge challenges such as time-varying channel conditions, limited resources, tight delay constraints, high bandwidth demand and complex protocol design. Therefore, there is need to efficiently utilize and manage the interactions among different layers of the protocol stack using cross layer design (CLD) approach in order to provide necessary support for video applications. This paper presents our proposed strategies for adaptive video transmission over Cognitive Ultra Wideband (C-UWB) network using MAC centric CLD approach. Generally, the proposed MAC centric CLD framework consists of sensing module, adaptive resource allocation module and adaptive quantization scale module. Two techniques are proposed namely; Basic CLD (B\_CLD) and Enhanced CLD (E\_CLD). In the B\_CLD, the decision strategies to stream video packets over C-UWB network are based on the pre-determined thresholds. On the other hand, decision strategies in E\_CLD are based on Lagrange optimization, which is implemented in the packet reception rate (PRR) based resource allocation scheme and adaptive quantization scale. Simulation results showed that the proposed E\_CLD scheme has significantly improved the video quality when compared to the Basic CLD scheme and the non-CLD scheme.

**Keywords**—cross-layer design; cognitive UWB network; video transmission

## I. INTRODUCTION

One of the major driving forces of a new technology is a combination of man's endless thirst for knowledge, unequal intuitive mind and limitless ambition to better his life. This is translated into an increasing demand for a low cost, higher speed and bandwidth hungry applications that run on many consumer electronic devices with seamless connectivity. However, the business models in emerging wireless systems are bounded with a concept of proportional end-user's cost to the volume of data transmitted but with limited bandwidth resources and transmission power. Due to this, industry players, scientists and academicians are urged to venture a new paradigm in offering the technical solutions. Cognitive radio (CR) is one of innovative solutions to address the issue of congested but inefficient spectrum utilization, introduced by Mitola [1].

In cognitive radio networks, it is the responsibility of cognitive users (CUs) to ensure that its existence will not cause any harmful interference to the primary user (PU). When the wireless systems that are potential candidates for cognitive radio are considered, Ultra Wideband (UWB) seems to be one of the tempting choices [2]. It is due to its

potential to fulfill some of the key cognitive radio requirements such as causing no interference to PUs, bandwidth, transmit power, supporting various throughputs and providing adaptive multiple access. In underlay mode, C-UWB user is allowed to co-exist in the same spectral and temporal domains with the PU by lowering the amount of transmit power. This is done by following the FCC rules [3], which authorized the use of UWB spectrum with spectral mask of  $-41.3\text{dBm/MHz}$ . Additionally, UWB is also known as a popular candidate for high rate data transmission over wireless personal area network.

Specifically on multimedia transmission over cognitive UWB (C-UWB) network, the key challenges come from the nature of multimedia application and cognitive radio itself as follows [1]:

- a) Resource constraint such as spectrum bandwidth, transmit power, data rate and time slot access.
  - In Time Division Multiple Access (TDMA) based MAC protocol, the main concerning issue is the sharing of time slots among wireless users. While in cognitive radio, data transmission and sensing activity are usually carried out separately at different time. Thus, appropriate time slots allocation and optimal scheduling are very important to guarantee accurate sensing information without causing any unintended delay.
- b) Dynamic network condition over time due to interference, shadowing and multipath fading.
  - During good channel condition, it is better to transmit data with higher rate to increase the throughput. However, it may cause higher bit error rate. Therefore, an optimal approach between the two conflicting objectives need to be addressed carefully.
- c) Heterogeneous video traffic
  - Due to frame dependency, loss of certain important frame will lead to indirect loss of other frames and may caused higher distortion impact. Hence, each video frame should be treated differently according to their frame priority, dependency and size.
- d) Stringent delay constraint
  - Delays of less than 200 milliseconds are required for interactive applications, such as videoconferencing, surveillance etc., while for multimedia streaming applications delays of 1-5s are tolerable. Packets that are arrived after their display time are discarded at the receiver side.

In cognitive networks, there is need for greater interaction among different layers of the protocol stack in order to achieve the end-to-end goals and performance in terms of resource management, security, QoS or other network goals. Therefore, cross layer design approach is needed. Cross layer design approaches can be categorized into application adaptation, application-centric adaptation, middle layer centric approach, middleware-based adaptation, and autonomous adaptation [8]. In this paper, we proposed a MAC centric cross layer design that is aware of MPEG-4 QoS requirements and PHY channel conditions to address the issue of multimedia transmission over C-UWB network. It is called a MAC centric CLD because the optimization and decisions are carried out by the MAC layer.

The rest of this paper is organized as follows. Section II presents the design concept of our proposed CLD which involve PHY, MAC and APP layer. In Section III, detail of the CLD functional components namely; video traffic module, adaptive resource allocation module, adaptive Q-scale module and sensing module are elaborated. Results and analysis are given in Section IV. Finally, conclusion and future recommendations are drawn in Section V.

## II. THE PROPOSED CROSS LAYER DESIGN CONCEPT

CLD plays a vital role at the design stage of CR system. The general key idea in CLD is to select the appropriate parameters which can be manipulated to gain a dramatic positive impact. Fig. 1 shows the layers involved, the parameters used and the adaptation actions performed in the proposed CLD design. The design goals are quantified in terms of PSNR, job failure rate (JFR) and user's perspective view. PSNR serves as an objective measure of the reconstructed video with respect to the uncompressed video frame. While JFR is defined as a ratio of total number of frames that are failed to be transmitted and total frames transmitted.

At the bottom, PHY layer is responsible in sensing the UWB wireless channel condition. Since sensing and data transmission are done separately at different time, optimal sensing period should be allocated at the MAC in order to obtain accurate sensing information while at the same time avoiding intolerance additional delay for video packet transmission. SNR of data packet is chosen as a link quality indicator, so as to detect spectrum availability. Practically, SNR is more reliable parameter because it is obtained after completion of demodulation process at the receiver [2]. The sensed information is then shared with the MAC and APP layer for optimal cross layer strategies.

At the APP layer, MPEG-4 video is encoded with different quantization scale during pre-process. The goal of pre-process video encoding is to prepare the video for adaptive quantization (Q) scale and hence allow rate adaptation during network simulation. The Q-scale is optimally adapted in accordance to the channel conditions. Detail of the Q-scale adaptation is elaborated in Section III-D. Heterogeneous video traffic is considered with attributes that are represented in terms of frame dependency, frame type, frame size and delivery deadline. At the MAC, optimal resource allocation is performed in accordance to UWB wireless channel

conditions, QoS target set by the APP layer and queue status.

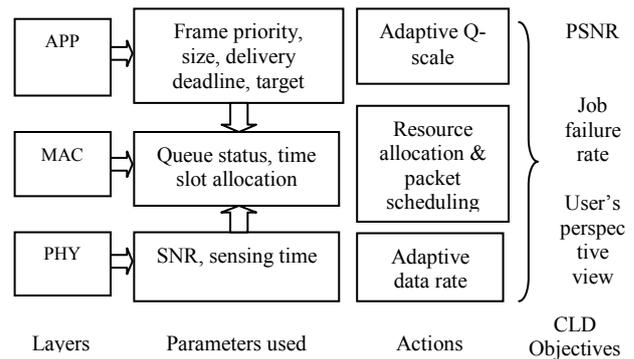


Figure 1. Layers involved, parameters used and action required in the proposed CLD

The optimal decisions are then forwarded to the respective layers for actions. At the APP layer, the Q-level is adjusted at every start of the GOP structure to maintain the synchronization and refresh of motion prediction algorithm. Due to adaptive Q-scale at the APP layer, the data rate at the PHY layer is also adaptive. Conversely, MAC layer schedules the sensing and data transmission in accordance to the varying channel conditions and the target packet reception rate (PRR). This will effectively reduce the unnecessary resource consumption and hence improve the system performance. Video frames are then scheduled for transmission in accordance to their attributes. The CLD design is simplified with no routing protocol and UDP is adopted at the transport protocol. This assumption is valid because UWB is targeted for WPAN with coverage within 10m and thus no hopping is required.

Centralized topology is adopted with one of the C-UWB nodes acts as a central controller surrounded uniformly by several others C-UWB nodes. The central controller is assigned as a common receiver while the other C-UWB nodes are assigned as transmitters with MPEG-4 video application.

## III. CLD FUNCTIONAL COMPONENTS

The proposed functional components of the CLD for multimedia transmission over cognitive UWB consists of video traffic module, resource allocation module, scheduler module, sensing module and adaptive Q-scale module. These components perform the actual cross layering task to support efficient wireless video transmission over cognitive UWB network. The state machine diagram and the algorithm are as shown in Figs. 2 and 3 respectively.

Initially, the target packet reception rate (PRR), optimal sensing time and delivery deadline are pre-determined for each C-UWB node. We assumed that the central controller knows the target QoS in advance. Referring to Fig. 2, the MAC centric CLD manager receives all required parameters from the CLD functional modules that involve APP, MAC and PHY layer for necessary parameter optimization. Once the optimized parameters are determined, it will be sent back to the respective layers for further actions.

At every video frame, C-UWB users trigger the central controller about its intention to perform local sensing. The C-UWB users report their sensing information to the central controller to be fused for the final decision of PU presence. The time of packet delivery is checked with the delay deadline to determine whether the packets in queue are expired or not. If expired, the packets are discarded from transmission. The Q-scale adaptation is carried out at the start of every GOP structure marked by the I-frame. Optimal resource allocation is performed based on the parameters passed from the APP and PHY layer. The optimized timeslot allocation, Q-scale and average of sensed SNR are then updated and passed back to the APP and MAC layer for packet transmission.

The MAC protocol is basically motivated from IEEE802.15.3 [10]. Each super frame starts with a beacon period (BP), during which the central controller sends the beacon containing network synchronization and control message. Then, C-UWB nodes can access the channel using contention-free with mechanisms that is based on slotted TDMA. However, the proposed MAC CLD eliminates the need of dedicated channel time slot request from each C-UWB nodes to the central controller and hence reduced the delay.

#### A. Sensing Module

The sensing module is responsible in monitoring the SNR level of each cognitive UWB users in the network. The SNR is represented by;

$$SNR = \frac{P_u L_{ij}}{\eta B} \quad (1)$$

Where  $P_u$  is average transmit power of node  $i$ ,  $L_{ij}$  is signal power attenuation,  $\eta$  is background noise energy and  $B$  is bandwidth. To calculate the signal power attenuation, UWB Tarokh's propagation model [7] is adopted as follows:

$$L_{ij} = [L_0 + 10\alpha \log_{10}(\frac{d_{ij}}{d_0})] + S; \quad d_{ij} > d_0 \quad (2)$$

where  $L_0$  is path loss at reference distance,  $\alpha$  is path loss exponent,  $S$  is shadowing,  $d_0$  is a reference distance and  $d_{ij}$  is the distance between the user <sub>$i$</sub>  and user <sub>$j$</sub> . From (2), the bit error rate (BER) and energy per bit can be calculated directly.

Local sensing is performed by each cognitive UWB user, while the central controller fuses the overall SNR information for cooperative sensing decision. In a cooperative spectrum sensing system using OR-Rule, the PU is considered to be present if any of the CU detects the presence of the PU. Assuming that there are  $N$  identical and independent cognitive radios in the cooperative spectrum sensing system, the cooperative probability of detection  $Q_d$  and probability of false alarm  $Q_f$  using OR-rule data fusion are given by:

$$Q_d = 1 - \prod_{i=1}^N (1 - P_{d,i}) \quad (3)$$

$$Q_f = 1 - \prod_{i=1}^N (1 - P_{f,i}) \quad (4)$$

where  $P_d$  and  $P_f$  are the probability of detection and probability of false alarm of a stand-alone cognitive radio respectively. To ensure almost accurate information of the instantaneous channel conditions, local sensing is triggered at every video frame interval. Assuming a worst case scenario for C-UWB, our previous findings in [11,12]

proposed 14 $\mu$ sec as the optimal sensing time for multimedia transmission over cognitive UWB. The results proved that only minimal overhead is introduced from sensing activity and the impact to video transmission is also minimal.

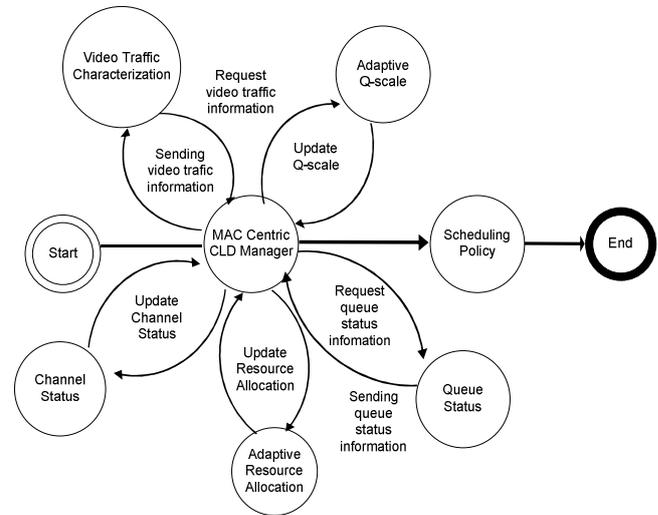


Figure 2. State machine diagram of the proposed CLD

#### ALGORITHM: Wireless video transmission basic-CLD

##### INITIALIZATION:

```

// To determine the frame type (I, P, B), size and
dependency
Video traffic characterization ();
SET target PRR; // The set target QoS for video
application
SET delivery deadline;
SET optimal sensing time;
REPEAT

FOR every video frame DO
//check delay delivery constraint

IF (T_delivery < T_deadline) THEN
Sensing (); // to obtain channel condition status
    IF (video frame is I-frame)
        Q-scale adaptation (); // Q-scale is
        changed at the start of new GOP
    ELSE
        // Q-scale unchanged
    END IF
Queue (); // to obtain queue status
Resource allocation (); //to determine time slot allocations
END IF
END FOR
Update number of time slots allocation
Update average sensed information
    Update Q-scale
    n = n + 1; // Update iteration
UNTIL n = max_n; // repeats algorithm until the last video
frame
    
```

Figure 3. The proposed CLD algorithm

### B. Heterogeneous Video Traffic Module

The role of video traffic module is to identify and classify the MPEG-4 video traffic in terms of frame type, frame size, dependency, delivery deadline and distortion impact. Delivery deadline,  $T_{deadline}$  is defined as the time by which the data units (DUs) must be decoded to be useful. It corresponds to the decoder timestamp in MPEG terminology. Each DU represents one of the I,P or B frames. Distortion impact is defined as the amount by which the distortion at receiver decreases if the DU is decoded on time at the receiver. Each  $DU_j$  has a distortion impact,  $\Delta d_j$  which is assumed to be constant for all the GOPs. The overall distortion can be computed as the initial distortion,  $d_0$  (ie: the distortion when no DUs are decoded) minus the sum of decrease  $\Delta d$  over all the DUs that have been decoded on time. Both sensing and video traffic module are responsible in providing appropriate sensing information and traffic classification to the MAC layer.

### C. Adaptive Resource Allocation Module

Since the network traffic is bursty, a fixed assignment of time or frequency slots to users is proven inefficient. In a case of basic CLD, linear adaptive resource allocation is adopted. In this case, the instantaneous SNR,  $SNR_i$  is compared with the two thresholds;  $SNR_{low}$  and  $SNR_{high}$ .  $SNR_{low}$  is obtained when the BER target of  $10^{-4}$  is achieved, while  $SNR_{high}$  is determined when the BER target of  $10^{-6}$  is achieved. Details on how the SNR threshold is determined can be found in [11]. Based on the works,  $SNR_{low}$  is set to -3dB while  $SNR_{high}$  is set to -1dB. When the channel condition falls within this allowable operating region, the time slots are assigned as follows;

$$M_i = 70 * SNR_i + 210 \quad (5)$$

$M_i$  is the allocated time slots for user  $i$ . The equation above is derived from linear correlation between SNR value and the allocated time slot.

Linear adaptive resource allocation is then further improved in the enhanced CLD scheme by also considering the instantaneous SNR and packet reception rate (PRR) of other cognitive users in the network. We named the technique as PRR based resource allocation and details of the algorithm can be found in [13-14]. To ensure efficient resource distribution among users, queue status is also considered before the actual time slot,  $M_{i\_actual}$  is assigned. Assuming the number of packets in queue is  $N_{queue}$  and  $L_{min}$  is the minimum data packet length allowed, the actual time slot assignment for PRR based resource allocation algorithm is as shown in Figure 4. Following the conditions, wastage of resources are reduced efficiently.

### D. Adaptive Quantization Scale Module

Distortion of the decoded video may originates from the quantization incurred at the decoder. In order to maintain the target BER for multimedia application, we proposed to adjust the Q-level in accordance to UWB wireless channel condition. The change in source rate may result in the fluctuation of the encoded video quality. To overcome this problem, source coding rate is only adjusted at the start of each GOP. It is known that larger step size (higher Q-scale) results in a lower bit rate and larger

amount of distortion. Hence, the optimal Q is also related to a well known rate-distortion (R-D) optimization problem; minimize distortion D, subject to a constraint  $R_c$  on the number of bits used, R. The constrained problem is as follows:

$$\text{Min } \{D\} \text{ subject to } R < R_c \quad (6)$$

```

//Calculate the initial resource using PRR based resource
allocation

$$M_i = \frac{K}{1 + \sum_{j=0}^{j=M} \frac{(1 - PRR_i)}{(1 - PER_{j,j \neq i})}}$$

//Check the queue status for actual resource allocation
IF ( $N_{queue} = 0$ )
     $M_{i\_actual} = L_{min}$ 
ELSE IF ( $M_i < N_{queue} * 210\mu\text{sec}$ )
     $M_{i\_actual} = M_i$ 
ELSE IF ( $M_i \geq N_{queue} * 210\mu\text{sec}$ )
     $M_{i\_actual} = N_{queue} * 210\mu\text{sec}$ 
    
```

Figure 4. Algorithm for the proposed adaptive resource allocation module

The above optimization problem can be elegantly solved using Lagrangian optimization and becomes;

$$\text{Min } \{J\}, \text{ where } J = D + \lambda R \quad (7)$$

where the Lagrangian R-D functional J is minimized for a particular value of the Lagrange multiplier,  $\lambda$ . Previous study had shown that [15];

$$\lambda = 0.85 (Q)^2 \quad (8)$$

At sufficiently high rates, the distortion impact can be approximated as;

$$D \cong (2Q)^2 / 12 \quad (9)$$

In the basic CLD scheme, the optimal Q-value is obtained through exhaustive search. While in Enhanced CLD scheme, the optimal Q-value is obtained by adopting the leaky bucket concept. For both schemes, the optimal  $\lambda$  is calculated using equation (8) in order to evaluate its impact to distortion.

## IV. SIMULATION PARAMETER

Simulations were divided into 3 stages namely; pre-process, network simulation and post-process. During pre-process, video samples are pre-encoded. The pre-process is performed only once and the generated trace files can be used over and over by new network simulations. While network simulations were carried out using NS-2. The post process is mainly deal with decoding the compressed video, calculating the performances and displaying the decoded received video. Tables 1, 2 and 3 show the parameter settings used. The target PRR and BER are set to 92% and  $10^{-6}$  respectively to meet the QoS requirement of multimedia application.

TABLE 1. MPEG-4 ENCODER SETTING

Parameter (unit)	Settings
Video input	Akiyo, Foremen, Coastguard
Group of Picture	12
Frame rate (frame/sec)	30
Frame deadline (sec)	1/30
Frame size (width x height)	176 x 144

Quantization scale	2-31
Number of frames	423

TABLE 2. MAC LAYER SETTING

Parameter (unit)	Settings
Super frame size ( $\mu$ sec)	1-65536 (adaptive)
Packet size (byte)	2000
PHY header time ( $\mu$ sec)	15
MAC header (byte)	16
Header check sequence, HCS (byte)	2
Frame check sequence, FCS (byte)	4
Short inter frame space, SIFS ( $\mu$ sec)	10
IP/UDP/RTP header (byte)	48
Acknowledgement	Immediate
Retransmission	3

TABLE 3: NETWORK SETTING

Parameter (unit)	Settings
Path loss exponent (dB)	1.7
Shadowing (dB)	2.8
Path loss at reference distance (dB)	50.5
Modulation	QPSK
Channel bandwidth (MHz)	528
Data rate	100Mbps

## V. RESULT & ANALYSIS

Table 4 summarizes the highest PSNR obtained at different UWB channel conditions and its corresponding optimal Q-value obtained through heuristic approach. As expected, the optimal Q-value moves towards higher value when channel condition becomes worse. Conversely, the Q-value is small when the channel condition is good. Meaning, smaller Q-value is chosen to reduce source coding distortion. When the assigned Q-value is higher than the optimal Q-value, the source distortion will increase (due to lower source rate). On the other hand, if the assigned Q-value is smaller than the optimal Q-value, more packets are generated per video frame and thus would lead to higher JFR. These explain why degradation in PSNR performance happened when non-optimal Q-value is used.

Akiyo, Foreman and Coastguard video sequences are used in the simulations to represent low, medium and high motion video respectively. In the non-CLD scheme, each user is assigned with fixed amount of time slot all the time regardless of their instantaneous channel conditions and the QoS. Alternatively, Basic CLD scheme adopted the linear resource allocation with Q-value adaptation. The video quality is further improved in the E\_CLD scheme by utilizing Lagrange optimization in the PRR based resource allocation and Q-scale adaptation. Table 5 shows the performance in terms of average PSNR over the whole video frames. It is evident that the proposed Basic CLD and E\_CLD scheme are more beneficial to high motion video (Coastguard) than the low motion video. Close observation of the PSNR value at each video frame shows that the proposed Basic CLD and E-CLD scheme give significant performance enhancement during bad and medium channel conditions. This really gives an advantage to cognitive UWB network, which is targeted for low SNR region to coexist with the PU. However, during good channel conditions, it is observed that the non-CLD scheme outperformed the proposed CLD schemes. E\_CLD also outperformed Basic CLD scheme

in terms of the average JFR. While the JFR for non-CLD scheme is the worst.

The performance can be much more appreciated if observed from the user perspective view as illustrated in Figs. 5, 6 and 7 below. However, due to page limitation, we present the video quality improvement difference between the non-CLD scheme and Basic CLD scheme. The figures obviously demonstrate that the received video quality is almost undecipherable when subjected to the non-CLD scheme during bad channel condition.

TABLE 4: OPTIMAL QP FOR VARIOUS CHANNEL CONDITIONS

Distance (m)	Average SINR (dB)	Optimal QP	Optimal PSNR (dB)
1	15.280	2	42.32
2	10.143	2	42.32
3	7.169	2	42.32
4	5.026	2	42.32
5	3.398	2	42.32
6	2.034	2	42.32
7	0.912	2	42.32
8	-0.090	2	42.32
9	-0.900	5	32.76
10	-1.737	13	24.92
11	-2.420	10	24.27
12	-3.080	15	22.57
13	-3.660	15	20.62

TABLE 5: AVERAGE PSNR FOR VARIOUS VIDEO SAMPLES

Video Sequence	Non-CLD [dB]	Basic CLD [dB]	E_CLD [dB]
Akiyo	28.68	29.08	30.21
Foreman	15.21	16.68	18.06
Coastguard	13.66	17.02	18.31

TABLE 6: AVERAGE JFR FOR VARIOUS VIDEO SAMPLES

Video Sequence	Non-CLD (%)	Basic CLD (%)	E_CLD (%)
Akiyo	3.5	1.0	0
Foreman	4.7	2.2	1.1
Coastguard	6.9	2.6	0.8



(a) No CLD, frame 59, PSNR= 12.76dB  
 (b) Basic CLD, frame 59, PSNR= 28.52dB  
 Figure 5. Subjective evaluation of the low motion video (Akiyo)



(a) No CLD, frame 64, PSNR= 12.78dB  
(b) Basic CLD, frame 64, PSNR= 18.19dB

Figure 6. Subjective evaluation of the medium motion video (Foreman)



(a) No CLD, frame 25, PSNR= 14.5dB  
(b) Basic CLD, frame 25, PSNR= 21.37dB

Figure 7. Subjective evaluation of the high motion video (Coastguard)

## V. CONCLUSION

This paper highlighted our proposed Basic CLD scheme and E-CLD scheme that is aware of the varying channel conditions and the target QoS at the APP layer in allocating sufficient time slots and in choosing the optimal quantization scale. We conclude that SNR is the main parameter to be considered in the CLD approach. Changing SNR channel conditions lead to different optimal Q-value. The proposed MAC centric CLDs had proven to be successfully improved the received video quality when. We believed that the CLD design can be further improved if more intelligent packet scheduling is adopted. Additionally, other optimization methods such as particle swarm optimization shall be considered to speed up the computation time for optimization and hence further reduce the end to end delay.

## ACKNOWLEDGMENT

The authors wish to express their gratitude to Research Management Center (RMC) of Universiti Teknologi Malaysia, Universiti Tenaga Nasional, Ministry Of Science, Technology & Innovations (MOSTI) of Malaysia and Islamic Development Bank (IDB) for the financial support of this project.

## REFERENCES

- [1] J. Mitola III. "Cognitive Radio an integrated agent architecture for software defined radio," *Ph.D thesis*, KTH Royal Institute of Technology, Stockholm, Sweden, 2000
- [2] H. Arslan, S. Yarkan, and Mustafa E. Sahin, "Cognitive Radio, Software Defined Radio and Adaptive Wireless Systems", Springer, ISBN 978-1-4020-5542-3, 2007
- [3] FCC Spectrum Policy Task Force, "Report of the spectrum efficiency working group," Nov 2002. [Online-May,2<sup>nd</sup> 2011] <http://www.fcc.gov/sptf/reports.html>
- [4] H. Kushwara, Y. Xing, R. Chandamouli, and H. Hefes, "Reliable Multimedia Transmission Over Cognitive Radio Networks Using Fountain Codes", Proceedings of the IEEE, Vol 96. No 1 January 2008, pp. 155-165.

- [5] H. Shiang and M. Schaar, 'Queueing-Based Dynamic Channel Selection for Heterogeneous Multimedia Applications Over Cognitive Radio Networks', *IEEE Transactions on Multimedia*, 2008, pp 896-909.
- [6] I. A. Akyildiz, T. Melodia, and K. R Chowdhury, "A Survey on Wireless Multimedia Sensor Networks", *Computer Networks* (Elsevier), Vol. 51, Issue 4, 2007, pp. 921-960.
- [7] M. Younis, M. Eltoweissy and A. Wadaa, "On Handling QoS Traffic in Wireless Sensor Network", Proceedings of the International Conference Hawaii International Conference on System & Sciences (HICSS-37), Big Island, Hawaii, January 2004.
- [8] F. Fu and M. van der Schaar, "A New Systematic Framework for Autonomous Cross-layer Optimization", *IEEE Transactions on Vehicular Technology*, Vol. 58, No. 4, May 2009, pp. 1887-1903
- [9] S. Ghassemzadeh and V. Tarokh, 'The Ultra-Wideband Indoor Path Loss Model', *IEEE P802.15-02/208r1-SG3a and IEEE P802.15-02/277r0-SG3a*
- [10] IEEE802.15.3 Wireless Medium Access Control (MAC) and Physical (PHY) Layer Specifications for High Rate Wireless Personal Area Networks, September 2003
- [11] N. M. Aripin, Rozeha A. Rashid, N. Faisal, and S.K.S Yusof, "Evaluation of Required Sensing Time for Multimedia Transmission Over Cognitive Ultra Wideband System", *IEEE International Conference on Ultra Modern Technology (ICUMT)*, October 2009, pp. 1-5.
- [12] N. M Aripin, Rozeha A. Rashid, N. Faisal, A.C.C Lo, S.H.S. Ariffin, and S.K.S. Yusof, "A Cross Layer Approach in Sensing and Resource Allocation for Multimedia Transmission over Cognitive UWB Networks", *EURASIP Journal on Wireless Communications and Networking*, Volume 2010, Article ID 467813, 10 pg
- [13] Norazizah Mohd Aripin et.al, "Joint Resource Allocation and Sensing Scheduling for Cognitive Ultra Wideband", The Australasian Telecommunication Networks and Applications Conference (ATNAC) 2010, pp. 66-71
- [14] Norazizah Mohd Aripin, Rozeha A. Rashid, and N. Faisal, "Issues in Resource Allocation & Sensing Scheduling for Multimedia Transmission Over Cognitive UWB Networks", *Wireless World Research Forum* 2010.
- [15] G. J Sullivan and T. Wiegand, "Rate Distortion Optimization for Video Compression", *IEEE Signal Processing Magazine*, November 1998, pp. 74-90.

# A Survey on Multi-Channel Based, Digital Media-Driven 802.11 Wireless Mesh Networks

Christian Köbel\*, Walter Baluja García\* and Joachim Habermann†

\**Department of Telecommunications, Polytechnic University ISPJAE José Antonio Echeverría  
Calle 114, No. 11901, Marianao, Havana, Cuba  
{Kobel, Walter}@Tesla.CUJAE.edu.cu*

†*Department for Information Technology, Electrical Engineering & Mechatronics  
TH Mittelhessen University of Applied Sciences  
Wilhelm-Leuschner-Str. 13, D-61169 Friedberg Germany  
Joachim.Habermann@IEM.TH-Mittelhessen.de*

**Abstract**—802.11 hardware has ultimately made its way into the electronic consumer market. But, especially in home networks, it clashes with high quality demands of rich multimedia streaming applications, that WirelessLAN can only partly cope with. This work describes basic features and fundamental flaws of 802.11, particularly with regard to multimedia streaming and Quality-of-Service (QoS) requirements. To overcome current problems of missing bandwidth resources in wireless home environments, the presented work points towards future multi-interface extensions of 802.11-based networks, in combination with decentralized Wireless Mesh Network (WMN) structures. The work provides an overview of all relevant 802.11 and WMN components with reference to important research works and indicates basic design aspects, crucial success factors and challenges of a future QoS-ready, multi-layer system that combines the advantages of decentralized wireless networks and of multi-channel MAC usage.

**Keywords**-802.11; Mesh; Multi-Interface; QoS

## I. INTRODUCTION

All things considered global consumer industry has made great experiences with the IEEE family of WirelessLAN 802.11 standards. But, the increased usage of wireless connections by consumer end devices has made wireless networks, especially home networks, become more heterogeneous. At the same time, new media applications raised the requirements on wireless networks and users expect the same quality of experience (QoE) from web based service platforms (e.g., YouTube) as it is well known from broadcast TV signals for example. So what has been achieved for wireless connectivity has to be achieved for *Quality of Service (QoS)* on future wireless links. The 802.11 spectrum is not fully exploited in most setups, although for example IEEE 802.11a would provide up to 12 orthogonal channels for simultaneous usage. Multi-channel / multi-interface WLAN networks offer a viable solution for these problems and some key aspects of their exploitation will therefore be presented in this work.

Furthermore, this work deals with 802.11 multi-channel systems in combination with *Wireless Mesh Networks (WMN)*. Recent hardware development favors the shortfall of access points in home setups, since nodes with multimedia capabilities provide more CPU resources and are able to compute complex routing decisions and forward packets in mesh networks on their own. To decrease the dependency on wired backbones and to enable a more scalable collective network, with enhanced all-wireless coverage and connectivity over a larger area, a WMN is the most suitable approach [1]. This advantage is in particular valuable in residential buildings, where often a single AP is not sufficient to connect all clients over several floors. Home networks with the purpose to essentially transport real-time media and less data traffic (e.g., a file download) are a key motivation scenario of this work.

The rest of the work is structured as followed: Section II provides essential knowledge about underlying 802.11 PHY and LLC techniques. Relevant mesh routing strategies are highlighted in Section III. Section IV identifies general QoS requirements. Finally, Section V outlines challenges and chances of possible multi-channel solutions for WMNs. In Section VI representative 802.11 measurements are presented and Section VII concludes the presented aspects. References to related works are included throughout the article.

## II. 802.11 - OVERVIEW AND RELEVANT COMPONENTS

The IEEE 802.11 standard covers the Medium Access Control (MAC) sublayer and the physical layer of the OSI reference model. But, an effective QoS-sensitive transmission chain requires the revision of QoS-related parameters (e.g., CPU power, buffer sizes, encryption, etc.) on almost every OSI layer. In this chain, 802.11 sub-protocols, especially in MAC layer, are at risk to become a bottleneck [2]. This section now covers the threat potential of known 802.11

issues in reference to QoS constraints, in order to efficiently apply mesh structures in home networks.

#### A. 802.11 Physical Layer Schemes

To conquer interference problems, 802.11 PHY layer natively provides frequency diversity: Depending on the region, 802.11b/g offers between 2 and 3, and 802.11a up to 12 non-overlapping channels. This generally favors the deployment of mesh nodes equipped with multiple wireless network interface cards (WNIC), but, current WLAN stations mostly deploy only a single interface, which can only communicate on one channel at a time. The first standardized attempt (in consumer hardware) to increase spectrum utilization within a single WNIC is described by 802.11n [2]. Besides MIMO antenna features, 802.11n allows to bond two 20 MHz channels and thus to increase the bandwidth on the PHY layer. Still, applying 802.11n physical bonding in 2.4 GHz band would consume up to two-thirds of the available spectrum and might hence increase interference levels. Those considerations about effective spectrum utilization may lead to a later addressed *dynamic bundling* approach, which envisions MAC coordinated channel bundling.

Due to physical limitations, PHY layer does not always match the requirement that “using  $n$  interfaces equals  $n$ -fold performance” [3] and so only offers a limited modularity for WNIC combinations. Per node performance is also limited by hardware capacities [4] and upper layer buffers; a less relevant effect for consumer hardware. Disadvantageous alignment of wireless cards and distances and angles between antennas might also jeopardize performance [5]. In reality negative radio effects like ACI might also occur on low SIR levels, despite the general assumption that two non-overlapping channels with physical proximity are supposed to be interference free [6]. Apart from that, capacity gain is always superior to all single-interface setups.

#### B. 802.11 MAC Layer Schemes

Besides physical interference, also the chance of a high amount of nodes sharing the same coverage area is responsible for the inherent unreliability of wireless systems. The more nodes are active, the higher is the packet congestion: Although 802.11 PHY layer is designed for multi-channel purposes, the medium access control (MAC) scheme implies a single channel shared by all stations, using CSMA/CA (Carrier Sense Multiple Access with Collision Avoidance). CSMA/CA was originally designed for solely best effort transmission, not considering distinction and prioritization of different packet classes. The lack of QoS support in 802.11 led to the first standard amendment 802.11e [7], that takes different traffic priorities into account. In 802.11e, standard 802.11 Distributed Control Function (DCF, with its strong best effort character) is replaced with the Hybrid Coordination Function (HCF), which increases the transmit probability for QoS-sensitive packets. 802.11e performance

strongly depends on packet inspection methods in the MAC layer in order to detect protection-worthy traffic, or on the corresponding QoS signaling in IP headers, performed by upper layers (e.g., DiffServ). This denoted dependency underlines the advances of a cross-layer QoS system [8].

Because of their poor scalability, random access protocols such as CSMA/CA are not an efficient solution for WMNs. Standard 802.11 MAC is not designed for multi-hop communication, which creates elementary problems in intermediate nodes, which forward packets between source and destination. These packets have to be replicated multiple times through multiple nodes, but still belong to the same end-to-end flow. The 802.11 standard does not consider this and forces every to-be-sent packet to compete for the medium anew, although it would be sensible to coordinate common time slots for coherent forwarding packets on all involved intermediate nodes on the multi-hop route. In affect, the available throughput for each node is not only limited by the raw channel capacity, but also by the forwarding loads imposed by other nodes, since only one node can access the medium at a time. This results in an inefficient store-and-forward process along the route. To reduce overhead time wasted for channel access negotiation, Bononi et. al. [8] described a basic and a fast-forward negotiation mode. The former mode equals standard DCF procedure, the latter mode extends RTS / CTS messages by a flow identifier and by a variable reservation period for each forwarded stream, facilitating multi-hop transmissions. These schemes can be further optimized by including a distributed multi-interface usage, for example when a set of interfaces A is responsible for the actual data transmission, while another set B coordinates the synchronized medium access along the route. Such enhanced functionality, when packet forwarding indicated by the network layer is supported by link layer, further argues for a cross-layer approach to decrease delay. IEEE standard committee recognized these needs and has formed the *Robust Audio/Video Streaming Task Group*, who defined general QoS requirements on WLAN in the 802.11aa standard amendment [9]. 802.11aa mainly depicts enhanced signaling between APs/stations and robust multi-cast, but also relies on packet inspection methods. A general overview of related 802.11 amendments is given in [10].

### III. WIRELESS MESH NETWORKS AND ROUTING PROTOCOLS

Especially in WMNs deployed in the private sector (e.g., home network) users frequently access external Internet resources and thus routes to the gateway node are more often penetrated than any other node-to-node connection, which provokes bottlenecks near the Internet gateway. If routing measures cannot compensate such limitations, alternative solutions have to be found in the underlying 802.11 access technology (see Section V).

To extend regular wireless mesh networks to multi-interface, QoS-ready WMNs, distance vector protocols like AODV are suitable in principle [12], because they already share a lot of neighbor information (included in signaling messages) locally, which could be extended with channel allocation information of the involved interfaces. On the other hand distance vector (DV) protocols are less scalable and produce high amounts of overhead, since comprehensive routing information is exchanged. In addition, DV protocols do not converge fast enough, due to the unsynchronized and unacknowledged way that distance vector information is exchanged. A fast convergence might be a critical factor for a home network, since users expect multimedia devices to be quickly integrated into the network. Link state protocols like OLSR [11] offer contrary and thus positive characteristics: Low network overhead (since exchanged information only contains single link characteristics), good scalability and fast convergence.

Research works show that cross-layer routing metrics may heavily increase performance in mesh routing [12]. Including medium-specific characteristics like e.g., the actual throughput, packet delay or packet delivery ratio values, directly obtained from driver API, allows a precise evaluation of the link quality, but, at the expense of compatibility (the system then always depends on a special operating system or WLAN driver). For a better multimedia performance in home networks, standard metrics should be replaced with *residual bandwidth capacity* and *round-trip-time*, which plays to the requirements of a video/audio stream. For a best effort traffic transmission high speed and low delay are not necessarily mandatory to carry out its duty, but rather a *high packet delivery rate*.

To enhance this multi-layer metric concept, in a next step routing metrics should adapt to changing network conditions [13]. Given the assumption that traffic is spread heterogeneously, the cumulative equation that describes the used routing metric in total could beneficially combine several sub-metrics and weight them with a factor that depends on dynamic zones in the mesh network. In zones with low QoS-related traffic activity and low packet collision/contention probability, it is better to traverse packets as fast as possible through this area, rather than observing the general bandwidth utilization in this zone for example. In this case *packet delay* and *hop count* dominate (represented by a higher weight factor). In zones with high network loads, that might threaten QoS performance on busy links, it is sensible to consider *the residual bandwidth capacity* instead. To realize a proactive indication whether a part of a route is about to carry real-time traffic or not, a source node could broadcast network-wide QoS-reservation messages for a planned end-to-end route, before starting the actual stream. Köhnen et al. [14] already implemented such functionalities in Ethernet networks. The reservations are included in the cumulative metric equation as weight factors.

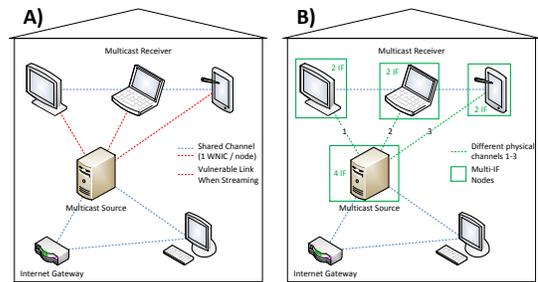


Figure 1. Single-Interface / Multi-Interface Home Topology

#### IV. TRANSPORT LAYER AND BEYOND

Introducing QoS in a home mesh reveals structural issues of standard WLAN setups: Advanced scenarios might require multicasting of local video/audio sources, which is contrary to common point-to-point data communication and may further provoke bottlenecks (see red links in Figure 1, A)). To implement QoS, the correct identification and classification of real-time packets is mandatory, but also hard to accomplish, due to the convergence of Media over IP traffic (e.g., YouTube, Hulu or Telekom Entertain), best effort traffic (e.g., HTTP protocol) and linear DVB broadcasts. Traffic can be identified in several ways and the identification process should never slow down the actual packet processing of the node. Most of the current setups require a deep packet inspection with a thorough payload analysis [14] to determine the real nature of a packet's content, which is generally not a problem with common home consumer devices. After a successful identification, transport protocols like the Resource Reservation Protocol (RSVP) can reserve bandwidth for certain flows. RSVP requires that every client supports the protocol and is therefore mainly applied in homogeneous cable backbones and less in wireless mesh networks. Besides bandwidth reservation and packet prioritization (e.g., 802.11e) the consideration of a multi-interface environment marks a third viable alternative to protect traffic. The benefits are regarded in the following sections.

#### V. CHALLENGES OF MULTI-INTERFACE ENVIRONMENTS

This section attends to initial problems of multi-interface / multi-channel (MIMC) networks. In most multi-interface solutions a single virtual MAC interface (which is presented as such to upper layers) combines and manages PHY interfaces. This architecture provides a good trade-off between portability and channel diversity exploitation, since the original 802.11 standard needs no further modifications and future PHY modulation schemes such as 802.11ac [10] can be easily adopted.

Xu et al. [13] examined differences between the aim to *exploit channel diversity* in MIMCs, which implies assigning as much different channels as possible to different neighbor-to-neighbor links, and sole *channel bonding*, which means

aggregating bandwidth capacities of the involved interfaces between two neighboring nodes. The group states that under high traffic load conditions, with a high degree of packet collisions that deteriorate the throughput and delay performances, relying on channel diversity greatly reduces collisions and may improve both throughput and delay since traffic can be equally distributed on more channels. Channel bonding only slightly decreases packet delays in high-traffic conditions. On the other hand, when traffic is low, the bonding of two interfaces achieves low delay values on a link and therefore favors QoS streams. Bonding has a better leverage effect when applied to concrete links that form parts of routes carrying QoS-packets between two nodes. MIMC is rather suitable to balance and absorb high packet quantities (both best effort and real-time) in the mesh network. The two concepts are regarded and shall be combined in the following subsections.

#### A. Channel Distribution in MIMC Environments

Channel distribution for multi-interface nodes must aim to at least equally distribute available channels in a hop-by-hop fashion, in order to decrease per channel congestion and mutual interference levels. Also, in this way the network topology might be actually shaped through (physical) route separations by frequencies; for example to separate three different streaming/multicast single hop links, as depicted in the in-house scenario B in Figure 1. First of all channel distribution has to be applied in a way that a *logical connectivity* (appearance in the routing table, despite of the used channel) between a node and every neighbor within its coverage range is guaranteed. If nodes transmit on different channels without further knowledge of the channels their neighbors are reachable on, they become invisible to each other. In the literature this problem is referred to as *deafness*, which further leads to common hidden terminal issues on multiple hops [15]. Therefore a dedicated *control channel* (CC) is needed [16]. The scope of the CC ranges from sole transmission of signaling packets generated by MAC (RTS, CTS, ACK) and network layer (e.g., OLSR HELLO and *Topology Control* overhead traffic) to hybrid solutions where overhead traffic is mixed with best effort traffic. CC must be always accessible by all stations and must carry information about a node's general access to a set of channels and which of them are currently assigned to / used by its interfaces. To somehow consider the limited coverage area, for the sake of simplicity a node distributes this information only to its 1-hop and 2-hop neighborhood, to not provoke excessive overhead by network-wide flooding of its channel signaling information. The information itself can be distributed through an arbitrary message type or, more economically, embedded in existing signaling messages.

Another deciding design attribute of a multi-channel / multi-interface network is the question whether channel allocation is performed pro- or reactively. Bononi et. al. [8]

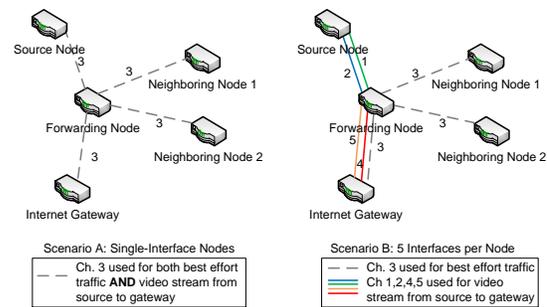


Figure 2. Separate Resources for Gateway Routes and QoS Traffic

deal with the question if channel assignment should follow after the actual route finding mechanism, or vice versa. Wu et. al. [16] implement channel allocation in a *reactive* manner by combining RTS/CTS management with the identification and distribution of unused channels within the 1-hop neighborhood, all based on a common CC. A proactive solution on the other hand might be more suitable, since a basic connectivity between nodes has to be established anyway. (Re-) Assigning channels in reaction to a transmit request by upper layers only consumes additional channel (re-)allocation time; a critical factor for delay-sensitive real-time traffic. Additionally, interference is prevented a priori when channels (orthogonal to each other) are assigned proactively, before upcoming flows.

The *dynamic adaption* to actual traffic loads and QoS-demands in a node, instead of simply assuring a uniform distribution of channels, make assignment mechanisms even more effective. To classify next-hop links and thus the majority of traffic they are bearing, we distinguish between best effort links (low priority), links that carry traffic towards or coming from Internet gateways and links that carry QoS traffic (both high priority). Identification of the type of traffic may be provided by upper layers (deep packet inspection methods). Gateway nodes are identified through the routing table. As mentioned before, gateway routes generally carry the majority of traffic in WMNs and should receive more resources to prevent bottlenecks in the mesh region near the actual gateway node. The distributable channel resources depend on the amount of available interfaces per node. In a first step, QoS- and gateway routes are favored with separate interfaces, while best effort links may communicate on a shared channel, applying CSMA/CA. Figure 2 depicts a related scenario. In this case the channel distribution is calculated and seen from the forwarding nodes's point of view, which deploys 5 WNICs. A video stream that takes a gateway route receives 4 different orthogonal channels. This significantly increases bandwidth, lowers delay and prevents negative interference effects for this stream. Based on this example, *resources in MIMC environments* are either represented by separate channels for each neighbor, or in the advanced case, by *multiple* separate channels for each

neighbor, which lead to the application of *channel bundles*.

### B. Channel Bundling in MIMC Environments

Bundles may be dynamically generated respectively degraded, depending on the variable amount of 1-hop neighbors (assuming that links suddenly break down or come up). In the particular case of fig. 2, each link of the QoS route aggregates capacities of two interfaces. Bundles are formed reactively (for example directly after the identification of a concrete QoS-stream), when extra capacity is needed and if the involved nodes can provide sufficient interface resources. Channel Bundling requires *load balancing*. To double bandwidth capacity, both channels can be loaded simultaneously with packets. Another concept sets the focus on stability: One channel might act as a fall-back option in case the main channel (that carries all packets) fails, or if the packet error rate on the main channel exceeds a threshold. Alternatively, redundancy (multiple transmission of the same packet) might increase stability as well. On the arrival side *packet reordering* has to be performed, depending on the chosen load balancing scheme. *Packet reordering* is a critical performance factor as well since wireless links are less predictable and packets may not always arrive in the correct order, due to different delay values of the interfaces.

In current research works, channel bundling options are still poorly considered, mostly because it is supposed that nodes may use all interfaces to serve different neighbors first. Therefore additional WNICs for bundling are rarely available. Still, channel bundling is the future concept to increase network capacity in 802.11. Tradeoffs between MIMC and bundling have to be defined here to enable efficient hybrid scenarios like in Figure 2, where all low-priority links can be served on a shared channel.

Channel bundling decreases packet delay times and for this reason it is especially interesting for supporting media streams. Since the overall transmission bandwidth of a link that contains bundled interfaces is enlarged, the transmission rate increases, which decreases packet transmission time, assuming a constant packet size.

## VI. MEASUREMENTS

The presented measurements were simulated with Network Simulator NS-2, in combination with OLSR as proactive link state routing protocol. They reflect the promising performance gain of multi-interface (IF) over single-IF 802.11 mesh networks. Comparative throughput (respectively TCP/FTP goodput), packet delay and packet drop rate observations are considered to evaluate the performance of each scenario.

Before addressing a home WMN, our first setup shall represent a large scale WMN and depicts a 150m x 150m area where 50 nodes were placed at random positions (cf. Figure 3). 4 extra nodes (yellow marked in fig. 3) are placed at fixed positions at the corners of the square

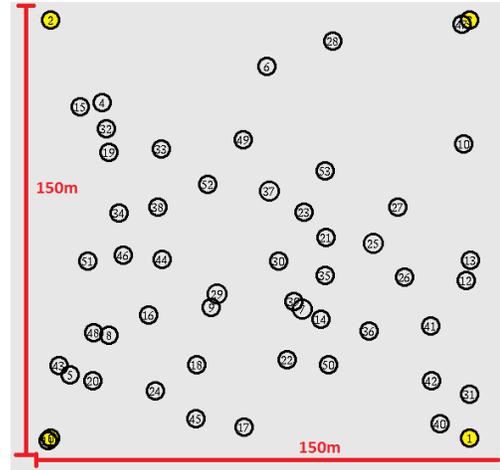


Figure 3. WMN with random node positions

area. Yellow node 1 initiates a TCP stream to node 2 in each of the 200 simulation runs, as well as node 0 to node 3. The two streams take different routes in each run due to their random character. A single-IF scenario, where all 54 nodes communicate on a shared 802.11a channel (54Mbits/s raw bitrate), is opposed to a multi-IF setup, where each of the 50 randomly positioned node contains two .11a interfaces. First interface is either tuned to channel 36 or 40, second IF to channel 44 or 48. Channels are selected randomly in each run. In this second scenario the 4 extra nodes apply 4 interfaces each to access all 4 used channels and to ensure a physical connection to all possible neighbors. Results are shown in Figure 4. Multi-IF setup clearly outperforms single-IF because limited single-channel resources are spatially disseminated on other frequencies, optimizing channel utilization. The “total” column of each graph refers to the overall value across the entire mesh.

Second measurement maps a typical indoor scenario in a 10m x 10m home (see Figure 5) where the left node on the top floor maintains a UDP stream for 200 seconds, which is interfered by a FTP stream after 100s. Again OLSR is used, but any other mesh protocol may be applied instead. Compared are both situations; undisturbed UDP/TCP stream and both UDP and TCP stream at the same time. Again, a single-IF setup (802.11a) is compared to a multi-IF environment (cp. legend in fig. 5). Results in Figures 6, 7 and 8 reveal that using a separate orthogonal channel for each link dramatically increases performance and makes the home network more suitable for QoS streams.

## VII. CONCLUSIONS AND FUTURE WORKS

This work focuses on 802.11-based wireless mesh networks, used to transport real-time streams. As shown by the so far discussed flaws of 802.11, the development of network solutions that provide abundant streaming capacities for future rich multimedia formats is critical for the ongoing

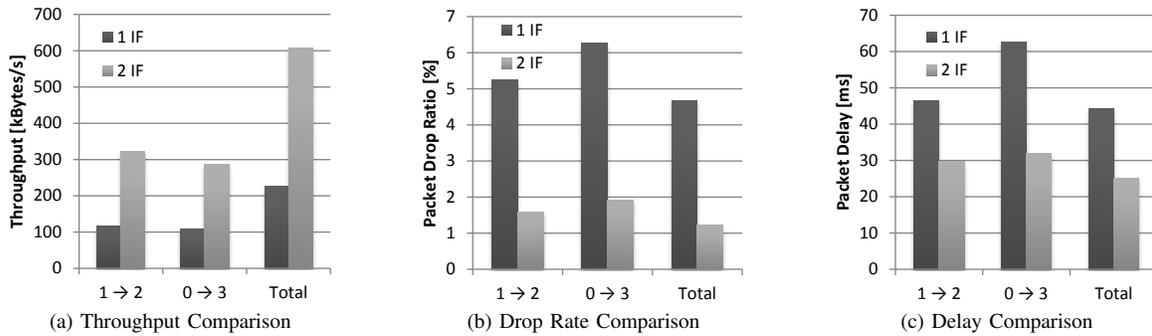


Figure 4. Single-IF vs. Multi-IF - Performance Comparison

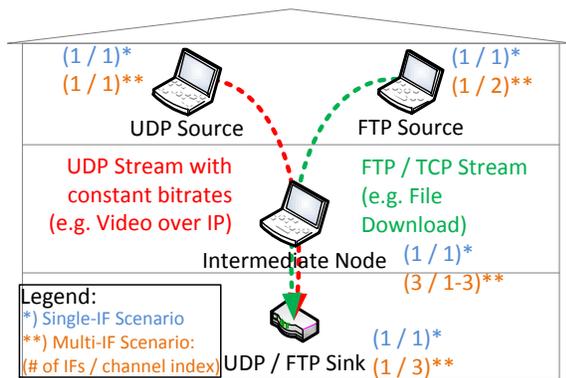


Figure 5. Home Mesh Scenario

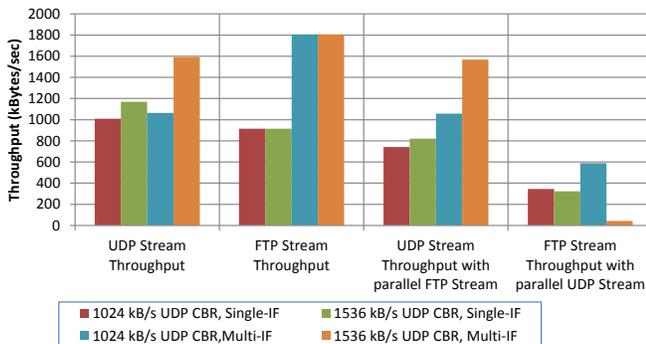


Figure 6. Throughput Results

adaption process of 802.11. Especially home networks based on standard single channel 802.11 links may not be able to provide sufficient bandwidth resources and high quality links to satisfy the constantly growing quality-of-service demands. This work provides a base for future multi-layer approaches. Multi-Layer Architecture allows routing metrics to get a more adequate view on the actual link states. Moreover, results of packet inspection methods can be used to facilitate traffic distribution over physically separated routes in the MAC layer, to decrease the general packet congestion per channel. A novel MAC layer instance to control the channel

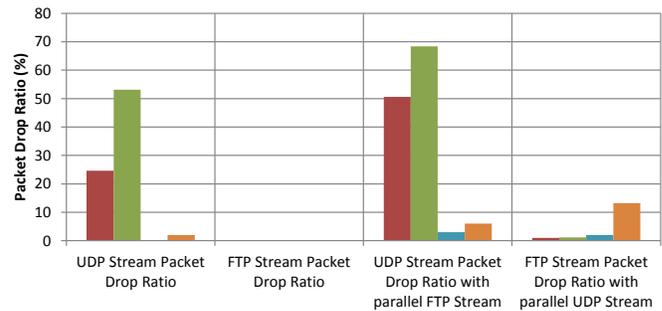


Figure 7. Packet Drop Ratio Results

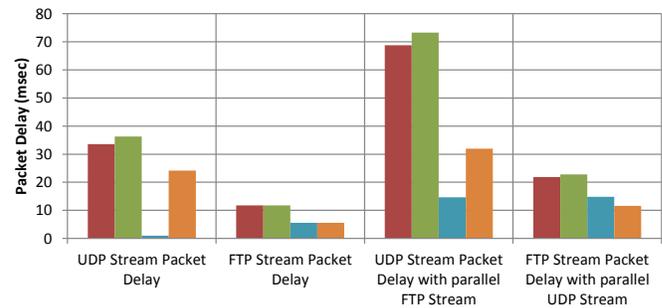


Figure 8. Packet Delay Results

assignment and channel bundling of connected interfaces forms part of the multi-layer strategy. To optimize the mixture of MIMC and bundling approaches is an important task for the future.

Now we are on the boarder to second and third generation mesh networks that offer quality features beyond simple connectivity of all nodes. 802.11s [18] is clear evidence that mesh structures are finally accepted in a broad range of wireless applications. To help the IEEE 802.11 standard to evolve with future requirements, its network structures have to become more dynamic and able to fully exploit the channel diversity offered by the PHY layer. Enhanced spectral efficiency and adaptive multi-channel utilization are the keywords for a significantly increased network capacity.

## REFERENCES

- [1] G. R. Hertz, S. Max, E. Wei, L. Berlemann, D. Denteneer, and S. Mangold, "Mesh Technology enabling Ubiquitous Wireless Networks", in Proc. of the 2nd Annual International Wireless Internet Conference (WICON 06), Boston, MA, USA: ACM, Invited Paper, pp. 11, August, 2006.
- [2] S. Maaroufi, W. Ajib, and H. Elbiaze, "Performance Evaluation of New MAC Mechanisms for IEEE 802.11n", in the Proc. of First International Global Information Infrastructure Symposium, 2007.
- [3] Chenzhe Zhang, Karol Kowalik, and Mark Davis, "An Experimental Study of the Impact of Using Multi-Radio in WLAN Mesh Networks", in the Proc. of IEEE WICOM '09, pp. 2587-2590, 2009.
- [4] R. Draves, J. Padhye, and B.Zill, "Routing in Multi-radio, Multi-hop Wireless Mesh Networks", in the Proc. of ACM Mobicom, Philadelphia, PA, pp. 114 - 128, 2004.
- [5] J. Robinson, K. Papagiannaki, C. Diot, X. Guo, and L.Krishnamurthy, "Experimenting With a Multi-Radio Mesh Networking Testbed", in the Proc. of WinMee, 2005.
- [6] Chen-Mou Cheng, Pai-Hsiang Hsiao, H. Kung, and D. Vlah, "Adjacent Channel Interference in Dual-radio 802.11a Nodes and Its Impact on Multi-hop Networking", in the Proc. of IEEE Global Telecommunications Conference GLOBECOM '06, pp. 1-6, 2006.
- [7] Nikos Passas, Dimitris Skyrianoglou, and Panagiotis Mouziouras, "Prioritized support of different traffic classes in IEEE 802.11e wireless LANs", ACM Computer Communications Volume 29, Issue 15, pp. 2867-2880, September 2006.
- [8] Luciano Bononi, Marco Di Felice, Antonella Molinaro, and Sara Pizzi, "A Cross-Layer Architecture for Efficient Multi-Hop Communication in Multi-Channel Multi-Radio Wireless Mesh Networks", in the Proc. of 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops SECON '09, pp. 1-6, 2009.
- [9] Ganesh Venkatesan, *Multimedia Streaming Over 802.11 Links*, IEEE Wireless Communications - Industry Perspectives, pp. 4-5, April 2010.
- [10] Guido R. Hiertz, Dee Denteneer, Lothar Stibor et. al., "The IEEE 802.11 Universe", IEEE Communications Magazine, pp. 62-70, January 2010.
- [11] A. Tonnesen, A. Hafslund, and Ø. Kure, "The Unik - OLSR Plugin Library", OLSR Interop and Workshop, 2004.
- [12] Sana Ghannay, Sonia Gammar, Fethi Filali, and Farouk Kamoun, "Multi-Radio Multi-Channel Routing Metrics in IEEE 802.11s-based Wireless Mesh Networks - And the Winner Is ...", in the Proc. of First International Conference on Communications and Networking, ComNet, 2009.
- [13] L. Xu, K. Yamamoto, and S. Yoshida, "Performance Comparison between Channel-Bonding and Multi-Channel CSMA", in the Proc. of WCNC '07, pp. 37-40, 2007.
- [14] C. Köhnen, C. Überall, F. Adamsky, V. Rakocevic, M. Rajarajan, and R. Jäger, "Enhancements to Statistical Protocol Identification (SPID) for Self-Organised QoS in LANs", IEEE ICCCN '10, pp.1-6, August 2010.
- [15] M. Di Felice, G. Zhu, and L. Bononi, "Future Channel Reservation Medium Access Control (FCR-MAC) Protocol for Multi-Radio Multi-Channel Wireless Mesh Networks", in the Proc. of IEEE PE-WASUN'08, pp. 71-79, 2008.
- [16] S.-L. Wu, C.-Y. Lin et al., "A New Multi-Channel MAC Protocol with On-Demand Channel Assignment for Multi-Hop Mobile Ad Hoc Networks", in the Proc. of IEEE ISPAN '00, Dallas, pp. 232-237, 2000.
- [17] Ali Hamidian, Claudio E. Palazzi, and Tin Y. Chong, "Deployment and Evaluation of a Wireless Mesh Network", in the Proc. of the Second International IEEE Conference on Advances in Mesh Networks, pp. 66-72, 2009.
- [18] S. Jeon, J. Lee, and J. Choi, "Performance Evaluation of Enhanced 802.11 in Wireless Mesh Networks", in Proc. of IEEE International Conference on Advanced Communication Technology, ICACT '08, pp. 839-842, 2008.

# A Bidirectional Semi-Passive DS-CDMA-RFID Transponder for the UHF Band

Andreas Loeffler, Ingo Altmann  
 Chair of Information Technologies  
 Friedrich-Alexander-University of Erlangen-Nuremberg  
 Erlangen, Germany  
 {loeffler, altmann}@like.eei.uni-erlangen.de

**Abstract**—This paper presents a first realization of a semi-passive UHF RFID transponder featuring a direct sequence code division multiple access channel access method for the RFID uplink, and, in addition, some limited downlink abilities, including carrier frequency and field strength measurements. The paper shows experimental results of the proposed transponder, which is already included within an UHF RFID system, containing an RFID reader being able to read out the transponders' data. The focus of this work is on the transponder's modulator, generating an ASK-like modulation scheme by impedance miss-matching techniques. This paper presents results obtained from real life measurements of a realized UHF RFID transponder qualified for the application in CDMA-based RFID systems.

**Keywords**-Radiofrequency identification; UHF; transponder; Backscatter.

## I. INTRODUCTION

Through an increasing number of RFID (Radio Frequency Identification [1], [2]) tags (transponders), particularly UHF tags [3]–[6], the need for a fast recognition of innumerable RFID tags puts great demand on future RFID readers [7]. If several tags are located within the range of a reader, signals from some of these tags collide [8]. This is the reason why anti-collision procedures [9] are widely used to prevent tags from broadcasting their information simultaneously. Existing RFID multiple access solutions are based on Time Division Multiple Access (TDMA [10]). Fig. 1a shows the TDMA method, pointing out that the tags in the reader's field transmit their data at different moments in time [1].

A simple TDMA protocol is the well known ALOHA protocol [11]. Basically, there are two modes the ALOHA protocol may be used with: pure (unslotted) and slotted ALOHA [12]. When using pure ALOHA [13], data are being sent at any times the channel is available. Because there is no control over the starting time, a collision is difficult to predict. Slotted ALOHA, on the other hand, is based on time slots which synchronize the start of transmission. Collisions can be predicted more easily and only happen, if at least two tags are requesting at the same time a free slot (e.g., slot 2, 3 and 5 in Fig. 1a).

The anti-collision procedure of the commonly used RFID standard EPC Class1 Gen2 [14] is based on slotted ALOHA. A so called  $Q$ -parameter controls the inventory process by determining the maximum number of possible time slots,

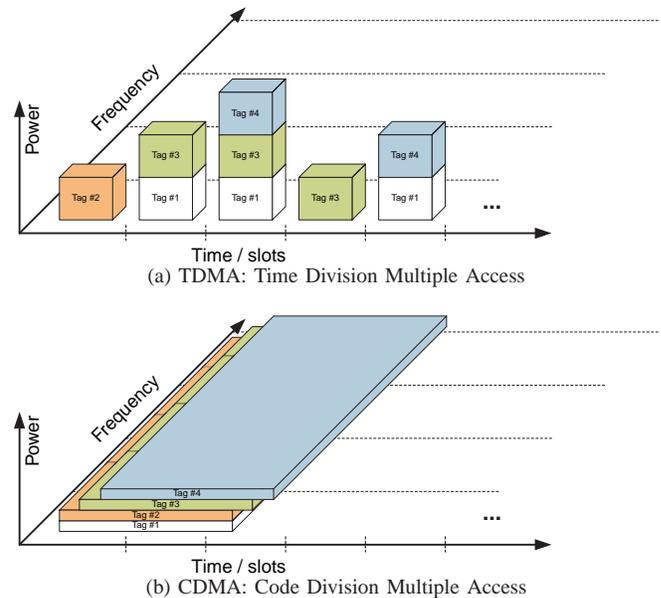


Figure 1. Comparison of TDMA and CDMA communication channel access techniques for RFID

which is set to  $2^Q$ . Each transponder in the reading range of an RFID reader fills its slot counter with a random number between 0 and  $2^Q - 1$ . Within each round, the slot counter is decreased by one. Once the slot counter of one or more tags reaches a zero state, the tag(s) send an acknowledgment command back to the reader. However, if tags chose equal random numbers, the tags will collide once their slot counters are zero. Hence, the choice of  $Q$  is a typical trade-off. Choosing a high  $Q$  will lead to a smaller number of collisions, at the expense of an increasing time needed for an inventory round. A smaller  $Q$  will lead to less acquisition time, but to more collisions, indeed. In addition, the usage of TDMA methods pushes the envelope of the system (time-wise) when a very high number of tags have to be scanned in very short time.

The introduction of Code Division Multiple Access (CDMA [15]) may find a remedy (Fig. 1b). The transponders, each equipped with a unique quasi-orthogonal spreading code, may use the radio channel whenever the transponders are ready to. The objective is the realization of a DS

(direct sequence)-CDMA-based RFID system using semi-passive UHF transponders [16], with the reader providing the recognition of multiple transponders simultaneously. This means that the transponders are transmitting data within the same time range and frequency band (Fig. 1b), in contrast to the existing systems based on TDMA.

The realized UHF transponders operate in semi-passive mode, meaning that the digital part of the transponder, i.e., the data generation, has an active power supply, whereas the high frequency (HF) part works in passive mode taking advantage of the backscatter principle. The attendant RFID reader, though, is separated into two parts. Part one, described as transmitting system, generates a carrier wave at around 867 MHz. Part two, the receiving system, mainly demodulates the incoming backscattered signals of the RFID tags.

The work described in this paper is based upon previous work, e.g., described in [17]. Also, prior work may be found in this reference. Although it seems there are no big differences between this work and previous work it can be stated that this is not a given fact. Several improvements (e.g., modulator with higher efficiency, new downlink capabilities of the transponder, higher range, etc.) have been carried out to push the system's performance. Also, a complete new receiving part has been realized in order to accelerate the inventory rounds.

The remainder of the paper is organized as follows. Section II depicts the system at a whole, Section III describes in more detail the structure of the UHF transponder, mainly focusing on the modulator. The section concludes with experimental measurements. Section IV outlines the results, particularly pointing out the received superimposed signals of multiple transponders. Finally, the paper concludes by showing references for future work in Section V.

## II. SYSTEM OVERVIEW

The whole RFID system is built upon three parts; a transmitting part and a receiving part, both defined as RFID reader, and a third part, involving one or more CDMA-based RFID UHF transponders. However, the RFID reader parts will not be specified in much detail in this section. Fig. 2 shows the system's setup.

The transmit part consists of a simple sine wave generator (TX source) and an antenna (TX antenna) creating the carrier wave on which the transponders respond using the backscatter method [1]. The carrier signal is created upon a tunable, PLL-based RF synthesizer board (within the setup as in Fig. 2, the PLL is substituted by two frequency synchronized signal generators). This PLL generates two sine waves (shifted  $\pi$  in phase), one for the transmitting part and one for the receiving part (i.e., the demodulator). The receiving part consists of an antenna (RX antenna), a LNA (low noise amplifier), an IQ-demodulator and a succeeding DSP-based platform for the baseband processing.

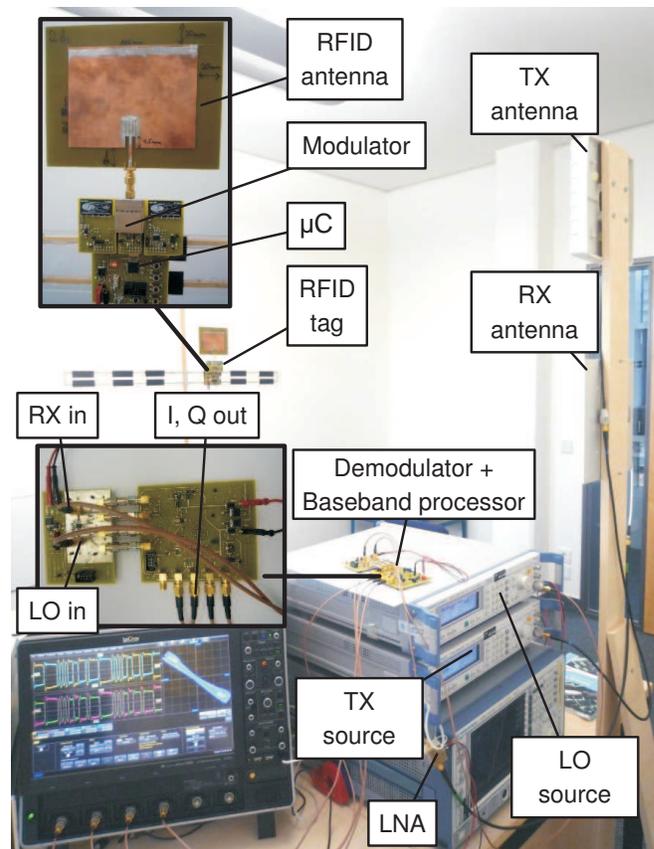


Figure 2. Setup of CDMA-based UHF-RFID system

The reflected signals from the transponders are low-noise amplified and directly mixed into baseband using a zero-IF demodulator. The demodulator's output consists of two I- and two Q-signals, each of them processed differentially. The following filtering and amplifying stages handle the signals differentially, too, in order to hold the current signal-to-noise ratio at a high level. As an example, the measured baseband signals at the demodulator's output, are shown in Fig. 2 (see oscilloscope) and Subsection III-B. Subsequently, these preprocessed signals are A/D-converted and pipelined to a DSP-based acquisition board, dealing with the mandatory baseband processing. The DSP is in charge of providing an appropriate despreading scheme to separate, in terms of data, various transponders from each other. This is mostly done through cross-correlation techniques. Some of these methods are evaluated in [17].

## III. DS-CDMA TRANSPONDER

The proposed transponders consist of four main parts. Part one, the tag antenna, is designed as patch antenna. Part two (see subsection III-A) is known as *backscatter* modulator. This part generates different impedance values seen by the tag antenna (see [17] for more details) in order to generate the wanted backscatter modulation scheme [1]

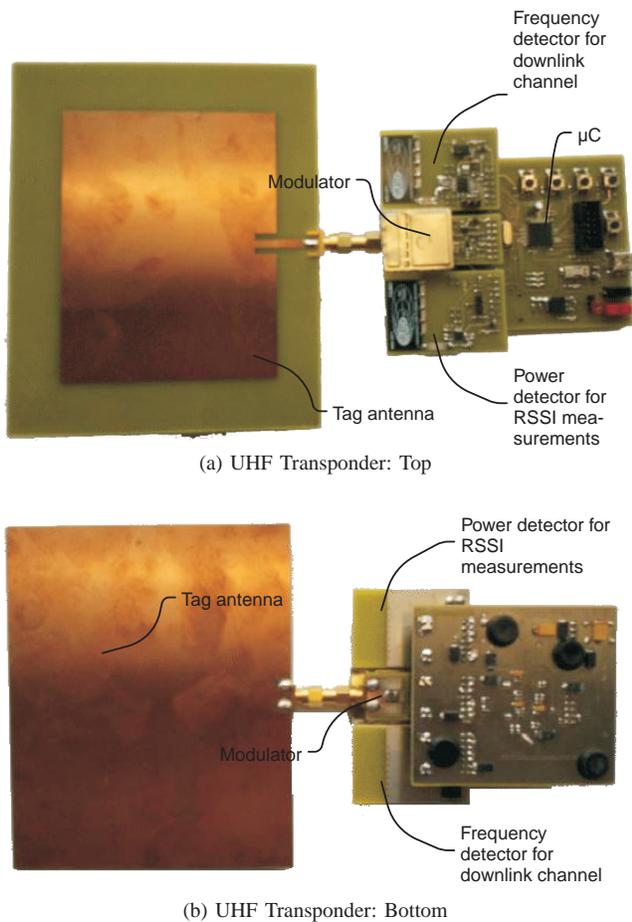


Figure 3. Implemented UHF transponder

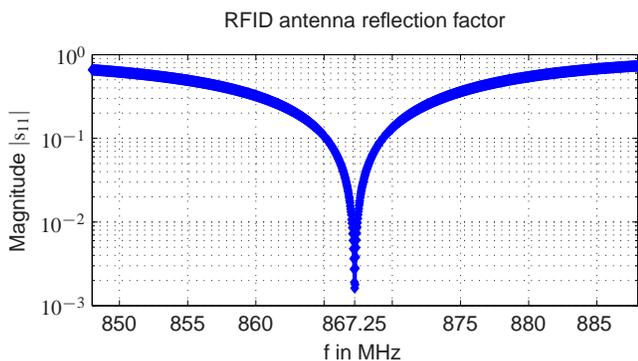


Figure 4. Measurement (vector analyzer) of reflection factor of RFID antenna within a range of 40 MHz

used along with non-inductive and non-capacitive working RFID systems. In most RFID systems, the realized backscatter modulation is a mixture between amplitude and phase modulation [2], [16].

Part three handles the baseband processing of the transponder, which includes mostly the generation of appropriate spreading codes, e.g., gold codes, and data spreading techniques for the CDMA scheme.

The last part describes a first realization of equipping the transponders with downlink (reader to tag) functionalities. This is realized using two elements; a frequency counter (referred to as frequency detector in Fig. 3a, 3b) and a detector of the incoming field strength (referred to as power detector in Fig. 3a, 3b). By implementing these two functionalities, the transponder may respond with different data to different incoming carrier frequencies. On the other hand, the transponder may use longer respectively shorter spreading codes according to the incoming field strength (lower respectively higher). This should mild the Near-Far effect [18], CDMA-based systems have to deal with. Anyway, the main part of the work focuses on the Antenna - Modulator effects, i.e., the uplink (tag to reader). A top and bottom view of the UHF transponder is given in Fig. 3a and Fig. 3b, respectively.

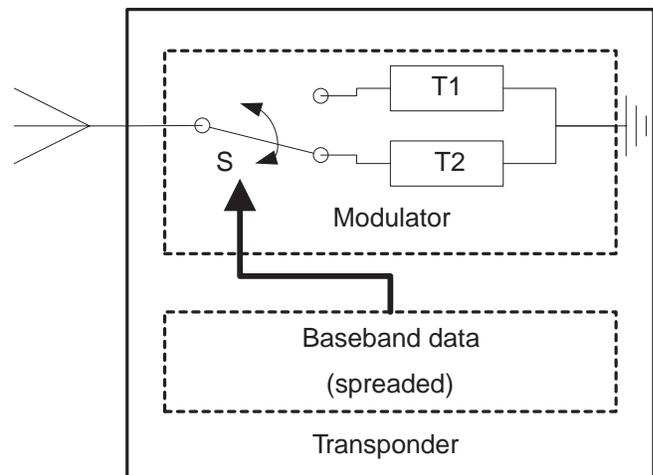


Figure 5. RFID transponder with HF switch as modulator

### A. Backscatter Modulator

The RFID tag antenna is designed for the usage with linear polarized waves at approximately 867 MHz at an impedance of  $50 \Omega$ . The antenna's gain has been calculated to  $G = 6.1$  dBi. The measurement of the antenna reflection factor is shown in Fig. 4. The current dimension of the antenna is still quite high (around  $3.5 \times$  ISO card size) but will be reduced to current tag sizes, certainly with the effect of a lower gain.

To be flexible in designing the modulator and in measuring various effects of the impedance, the choice fell upon an HF switch (single pole, double throw) for tuning respectively switching the two different impedances (binary modulation). Fig. 5 shows the principle of the modulator designed as HF switch (S in Fig. 5). Examining the resulting backscatter effect, one has to say, that the well-directed two miss-matchings between the HF switch terminations (T1, T2) and the RFID patch antenna generate different

amounts of energy scattered back to the receiving part of the RFID reader [2], [16], [19]–[22]. These reflected waves of the transponder differ in amplitude and/or phase. This difference, indeed, is used by the reader’s demodulator to recognize the binary states of the transponder’s data and demodulate them accordingly. The values of the impedances and therefore the amount of reflected energy can be tuned by varying the terminations (T1 and T2 in Fig. 5) of the HF switch. A good first approach is achieved by inserting a short circuit on termination T1 and an open circuit on T2. Feeding the HF switch with spreaded data from the transponder (i.e., binary ‘0’s and ‘1’s), leads to impedance alternations between short and open circuit state, thus generating backscattered waves with differing phase changes, leading to a BPSK-like modulation scheme. However, these signals are superimposed by various scattered waves of the carrier from the reader, finally forming an ASK-like modulation at the demodulator’s input. The following results of carried out experiments were received by using the setup as in Fig. 2.

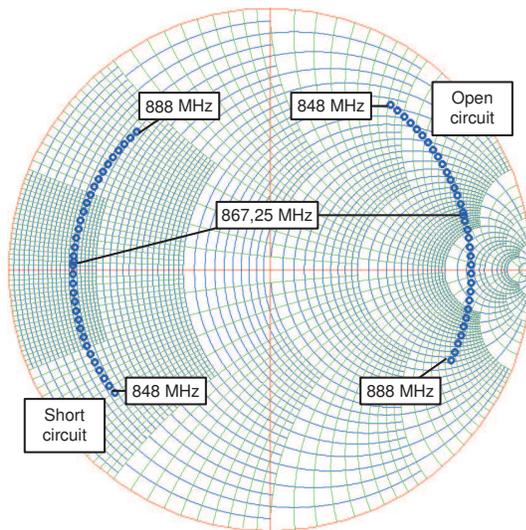


Figure 6. Smith Chart of the two different binary states of the modulator (short and open circuit)

### B. Measurements

To show the working principle of the proposed system, measurements were carried out. The first measurement includes the reflection factor (i.e.,  $s_{11}$ ) of the RFID transponder’s patch antenna. The second measurement involves the impedance measurement (i.e., reflection factor) of the modulator for both binary states, open and short circuit. The last measurements carried out within the setup as of in Fig. 2, show the demodulated baseband signals of the backscattered power from one or more tags within IQ constellation and timing diagrams.

The result of the first measurement is shown in Fig. 4,

indicating the absolute value of the measured reflection factor  $|s_{11}|$  of the tag antenna used. It can be shown that the proposed (simulated)  $s_{11}$  at a frequency of  $f_c = 867$  MHz could be approximately hit with an offset of only 250 kHz. The measurement was carried out in a laboratory environment.

The measurement results of the modulator are shown in Fig. 6. As expected, the results of the short circuit states are on the left side of the Smith chart, whereas the open circuit states are located at the right side. The frequency range was limited to 40 MHz with around 867 MHz at the center. It can be stated, that the measured values show a good approximation of a BPSK modulation scheme.

The measurement, showing the whole effect of the transponder’s properties and the principle of the demodulator board, is shown in Fig. 7. The resulting values of this measurement are the processed (i.e., filtered and amplified) baseband signals of the transponder’s backscatter response. The four underlying signals (i.e., two differential signals for the I- and two for the Q-channel, respectively) were sampled by an oscilloscope. After building the differences ( $I_+ - I_-$  and  $Q_+ - Q_-$ ) in order to receive the pure I- and Q-signals, the resulting signals are shown in an IQ constellation diagram. It is important to keep in mind, that the measured signals are sampled DC-free to show the effect of the BPSK modulation. Showing the full DC-afflicted signal, would result into an ASK modulation scheme (which, of course, is received at the receiving antenna). Anyway, Fig. 7 not only shows the baseband signals, it shows signal regions, sorted colored (white to black) according to their frequency of occurrence. The more often particular signal states are present within a given region, the darker the region is. That means, that the two darkest states at the upper left and lower right, actually represent the two states of the modulator, i.e., open circuit and short circuit. The gray states in between the darker states show the switching operation of the HF switch. A more detailed view on the measurement showed, that the switching process from one stable state to the other, is executed in two different ways (non-reciprocal), as the gray region in Fig. 7 is divided by a white region. The distance between RFID reader antenna and transponder is approximately 3 m. However, further experiments showed a verified distance of around 15 m.

The remaining measurements show the complex baseband signals after mixing and filtering by the demodulator. These measurements were carried out using up to three transponders in the reader’s field. Current baseband signals with 2 transponders in the field are shown in Fig. 8. The number of fixed signal states is 4, corresponding to 2 transponders. The reason for that is given by the fact, that the two transponders respond on the same “wave” emitted by the reader (coherent backscattering). Therefore, both transponders backscatter their very own signature on the coherent carrier of the reader. As each transponder inherits two stable states, the resulting

waveform consists of  $2^2$  states. This issue is exemplary shown in Fig. 9. Fig. 9a shows the complex baseband signal with 1 transponder, Fig. 9b with 2 transponders, and Fig. 9c with three transponders in the field. However, the results describing the currently used CDMA methods (spreading, despreading) are part of coming up work.

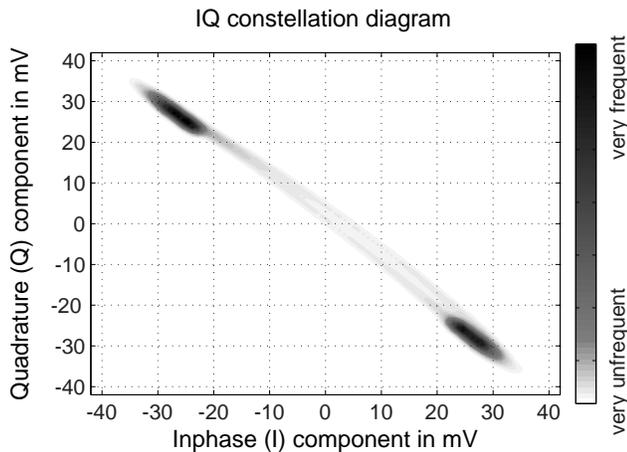


Figure 7. IQ constellation diagram showing baseband I- and Q-signals according to their frequency of occurrence, with black being very frequent and white being no occurrence at all

#### IV. RESULTS

The measurement results presented in Subsection III-B show the basic operation principles of the proposed CDMA-transponder. Using above realized impedance states, i.e., open and short circuit for the transponder's modulator, the RFID reader, particularly the receiving system, may receive ASK-like modulated and spreaded transponder data, although a BPSK modulation scheme is provided by the transponder. Having this in mind, the receiver may easily recognize a lot more transponders simultaneously, by

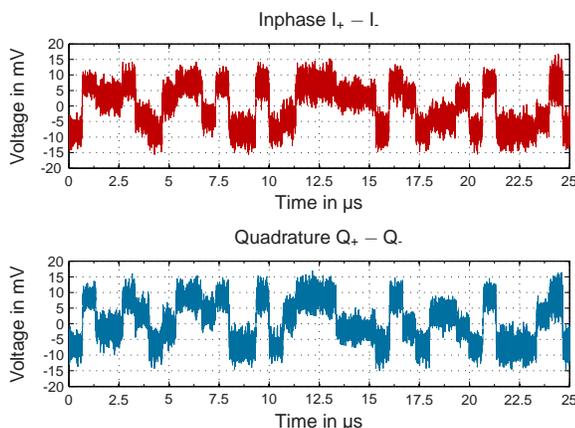


Figure 8. Differential signals of I- and Q-component at the output of the demodulator showing two transponders with superimposed CDMA signals

downconverting the spreaded and overlapped data of the individual transponders into baseband, and by evaluating the modulated phase and/or amplitude [17] through despreading techniques. The methods describing these despreading techniques in detail are part of different work.

#### V. CONCLUSION AND FUTURE WORK

The work shown in this paper proposes a new semi-passive DS-CDMA-based RFID transponder system for the UHF frequency band. The RFID reader uses a transmitting part to generate the carrier wave, again used by the transponders to reflect the incoming electromagnetic wave using different reflection states. These different states are achieved within the transponder's modulator, mainly consisting of an HF switch, creating impedance miss-matchings (open and short circuit) against the  $50\ \Omega$  RFID tag patch antenna, in order to generate two different reflected waves (in terms of energy). These backscattered waves are received by the receiving part of the RFID reader for further processing. The applied CDMA channel access method is used to recognize transponders responding at an arbitrary point in time.

Further work has to be done by verifying the downlink part for the transponders, and the data respectively signal despreading of various transponders in the reading range of the RFID reader.

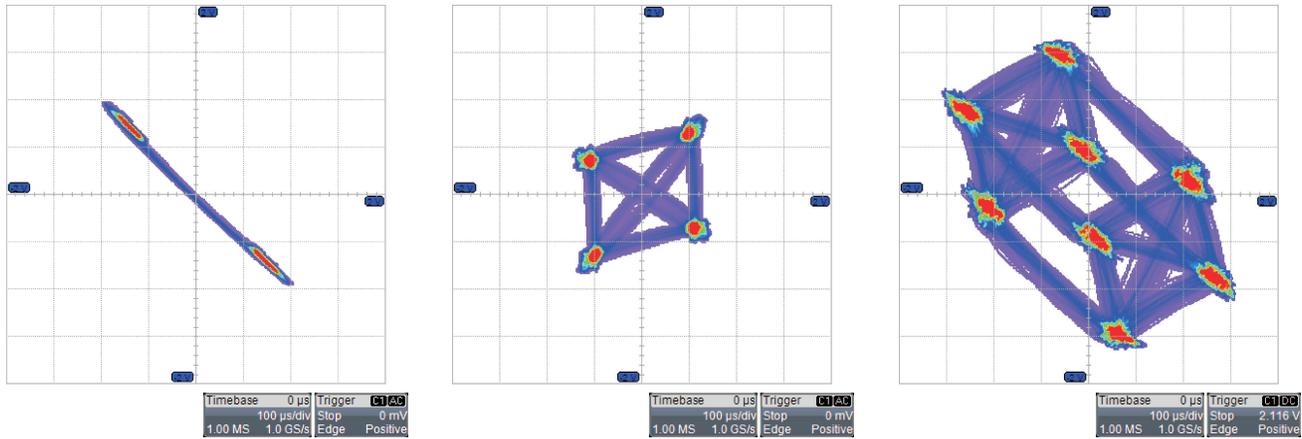
Tuning the impedance miss-matchings between modulator and antenna by applying other elements such as resistors, capacitors or inductors, or a mixture between them to the modulator, could provide better readability of the tags.

#### ACKNOWLEDGMENT

The authors would like to thank Fabian Schuh for his preceding work on this subject.

#### REFERENCES

- [1] K. Finkenzeller, *RFID handbook*. Wiley West Sussex, England, 2003.
- [2] K. Finkenzeller and D. Müller, *RFID Handbook: Fundamentals and Applications in Contactless Smart Cards, Radio Frequency Identification and Near-Field Communication*. Wiley, 2010.
- [3] RNCOS, "Global RFID Market Analysis till 2010," Aug. 2010.
- [4] Frost & Sullivan, "Global Market RFID," Dec. 2010.
- [5] Infiniti Research Limited, "Global RFID Market 2010-2014," Feb. 2011.
- [6] VDC Research Group, "RFID Transponders, ICs & Inlays: Track 1, Volume 1: The 2009 RFID Business Planning Service," 2009.
- [7] A. Loeffler, U. Wissendheit, H. Gerhaeuser, and D. Kuznetsova, "A multi-purpose RFID reader supporting indoor navigation systems," in *RFID-Technology and Applications (RFID-TA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 43–48.



(a) Measurement: IQ constellation diagram with 1 transponder in the field; signal is sampled at baseband (1 GSps);  $2^n = 2$  signal states correspond to  $n = 1$  transponder(s)

(b) Measurement: IQ constellation diagram with 2 transponders in the field; signal is sampled at baseband (1 GSps);  $2^n = 4$  signal states correspond to  $n = 2$  transponder(s)

(c) Measurement: IQ constellation diagram with 3 transponders in the field; signal is sampled at baseband (1 GSps);  $2^n = 8$  signal states correspond to  $n = 3$  transponder(s)

Figure 9. Baseband measurements (IQ constellation diagrams) with 1, 2 and 3 transponders in the reading range of the RFID reader; signals are sampled after baseband processing, i.e., IQ demodulated, filtered and amplified

- [8] D. Shih, P. Sun, D. Yen, and S. Huang, "Taxonomy and survey of RFID anti-collision protocols," *Computer communications*, vol. 29, no. 11, pp. 2150–2166, 2006.
- [9] P. Peris-Lopez, J. Hernandez-Castro, J. Estevez-Tapiador, and A. Ribagorda, "RFID systems: A survey on security threats and proposed solutions," in *Personal Wireless Communications*. Springer, 2006, pp. 159–170.
- [10] D. R. Campbell, "MESA: A Time Division Multiple Access System," *Int'l. Symp. Global Commun.*, p. 61, Jun. 1964.
- [11] N. Abramson, "THE ALOHA SYSTEM: another alternative for computer communications," in *Proceedings of the November 17-19, 1970, fall joint computer conference*, ser. AFIPS '70 (Fall). New York, NY, USA: ACM, 1970, pp. 281–285. [Online]. Available: <http://doi.acm.org/10.1145/1478462.1478502>
- [12] L. G. Roberts, "ALOHA packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, pp. 28–42, April 1975. [Online]. Available: <http://doi.acm.org/10.1145/1024916.1024920>
- [13] D. Bertsekas and R. Gallager, *Data networks (2nd ed.)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1992.
- [14] "Class 1 Generation 2 UHF Air Interface Protocol Standard," EPCglobal Inc., 2008.
- [15] A. Viterbi *et al.*, *CDMA: Principles of spread spectrum communication*. Addison-Wesley Norwell, MA, 1995, vol. 129.
- [16] D. Dobkin, *The RF in RFID: passive UHF RFID in practice*. Newnes, 2007.
- [17] A. Loeffler, F. Schuh, and H. Gerhaeuser, "Realization of a CDMA-Based RFID System Using a Semi-Active UHF Transponder," in *Wireless and Mobile Communications, 2010. ICWMC '10. Sixth International Conference on*, September 2010.
- [18] J. Andrews, "Interference cancellation for cellular systems: A contemporary overview," *IEEE Wireless Communications*, vol. 12, no. 2, pp. 19–29, 2005.
- [19] P. D.-I. B. Rembold, "Optimum modulation efficiency and sideband backscatter power response of rfid-tags," *Frequenz - Journal of RF-Engineering and Telecommunications*, vol. 63, no. 1-2, pp. 9 – 13, Januar 2009.
- [20] P. V. Nikitin and K. V. S. Rao, *Theory and Measurement of Backscattering from RFID Tags*. Intermec Technologies Corporation, 2003.
- [21] J. Andersen and R. Vaughan, "Transmitting, receiving, and scattering properties of antennas," *Antennas and Propagation Magazine, IEEE*, vol. 45, no. 4, pp. 93 – 98, August 2003.
- [22] W. Kahn and H. Kurss, "Minimum-scattering antennas," *IEEE Transactions on Antennas and Propagation*, vol. 13, no. 5, pp. 671 – 675, September 1965.

## Efficient Symbol Detector for MIMO Communication Systems

Soohyun Jang

School of Electronics, Telecomm., Computer Eng.  
Korea Aerospace University  
Goyang-si, Korea  
shjang@kau.ac.kr

Yunho Jung

School of Electronics, Telecomm., Computer Eng.  
Korea Aerospace University  
Goyang-si, Korea  
yjung@kau.ac.kr

**Abstract**—In this paper, an area-efficient symbol detector is proposed for multiple-input multiple-output (MIMO) communication systems with two transmit and two receive antennas. The proposed symbol detector can support both the spatial multiplexing mode and spatial diversity mode in a single hardware, and shows the optimal maximum likelihood (ML) performance. By applying the multi-stage pipeline structure and using the complex multiplier based on polar-coordinate, the complexity of the proposed architecture is dramatically decreased. The proposed symbol detector was designed in hardware description language (HDL) and synthesized to gate-level circuits using 0.13 $\mu$ m CMOS standard cell library. With the proposed architecture, the total logic gate count for the detector is 393K, which is reduced by 57% compared with the conventional architecture.

**Keywords**—MIMO; ML; multi-stage pipeline; spatial diversity; spatial multiplexing; symbol detector

### I. INTRODUCTION

As the demand for a high-rate and reliable wireless data transmission increases, MIMO techniques have attracted considerable attention in recent wireless communication systems such as IEEE 802.11n wireless local area network (WLAN), IEEE 802.16e mobile worldwide interoperability for microwave access (WiMAX), and 3GPP long-term evolution (LTE) [1].

The MIMO techniques can basically be classified into spatial diversity (SD) scheme [2] and spatial multiplexing (SM) scheme [1]. In SM scheme, since independent data streams are transmitted from the individual transmit antennas, the overall data rate is significantly increased as the number of transmit antenna increases. Meanwhile, since SD systems transmit multiple streams bearing the same information, the link reliability is considerably improved from the spatial diversity gain even though there is no increase in data rate.

In SD scheme, the optimal ML symbol detection can be easily accomplished by simple linear combination at receiver [2]. However, since the ML detection for SM scheme requires exhaustive search for all transmitted symbols from all transmit antennas, its complexity is proportional to  $C^{N_T}$  ( $C$  is the constellation size and  $N_T$  is the number of transmit antennas) and exponentially increases as

$C$  and  $N_T$  increase. Therefore, its real-time implementation is infeasible when a large number of antennas are used together with high constellation size, e.g., 64QAM.

In order to solve this complexity problem, there have been vigorous researches in recent decades [3]-[12]. Among them, modified ML (MML) detection algorithm, which can reduce the complexity by the ratio of  $1/C$ , was proposed in [10], and was applied to several implementations as in [11]-[12]. Since the recent communication systems mostly support two transmit and two receive antennas to be incorporated into mobile device, MML detection can be considered to be suitable for the symbol detector of those systems because its complexity is proportional to only  $C$ .

Although MML detection provides a lower complexity than the classical ML detection, its complexity is still too high to be implemented in real time especially when supporting 64QAM because 64 complex calculations for Euclidean distance should be performed in parallel. Also, since SD scheme such as space-time block coding (STBC) [2] and space-frequency block coding (SFBC) [13] should be supported together with SM scheme in most systems, the design of the efficient hardware architecture is really important for the MIMO symbol detector.

In this paper, we propose an area-efficient MIMO symbol detector supporting both SD and SM modes and present its design and implementation results. By fully sharing the common function blocks and applying the multi-stage pipelining, the proposed detector is implemented with very low-complexity.

This paper is organized as follows: In Section II, MIMO system model is presented, and ML and MML symbol detection algorithms are introduced in Section III. The hardware architecture for the proposed symbol detector is described in Section IV, and the implementation results are presented in Section V. Finally, Section VI concludes the paper.

### II. SYSTEM MODEL

Fig. 1 depicts the MIMO system model with  $N_T$  transmit and  $N_R$  receive antennas  $N_T \geq N_R$ . The receive signal vector is given by

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{X} + \mathbf{N} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_{N_T} \end{bmatrix} \cdot \mathbf{X} + \mathbf{N}$$

$$= \begin{bmatrix} h_{11} & h_{21} & \cdots & h_{N_T 1} \\ h_{12} & h_{22} & \cdots & h_{N_T 2} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1N_R} & h_{2N_R} & \cdots & h_{N_T N_R} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{N_T} \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{N_R} \end{bmatrix}, \quad (1)$$

where  $x_j$ , ( $j=1,2,\dots,N_T$ ) is the signal transmitted from the  $j$ -th transmit antenna,  $y_i$ , ( $i=1,2,\dots,N_R$ ) is the signal transmitted from the  $i$ -th transmit antenna; and  $h_{j,i}$  ( $j=1,2,\dots,N_T$ ,  $i=1,2,\dots,N_R$ ) is the fading channel between the  $j$ -th transmit antenna, and the  $i$ -th receive antenna. Also,  $n_i$ , ( $i=1,2,\dots,N_R$ ) is an independent and identically distributed (*i.i.d.*) complex zero-mean Gaussian noise with variance  $\sigma^2$  per dimension.

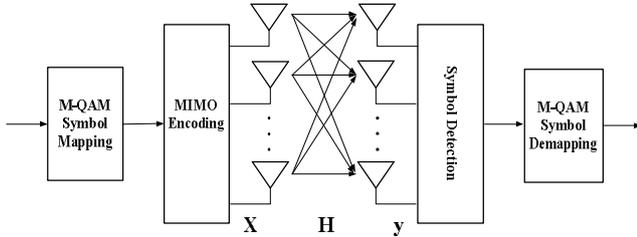


Figure 1. MIMO system model with  $N_T$  transmit and  $N_R$  receive antennas

### III. SYMBOL DETECTION ALGORITHMS

#### A. ML Symbol Detection Algorithm

The ML detection method, which achieves the best performance by searching for the transmitted signal vector that leads to the minimum Euclidian distance from the received signal vector, can be represented as

$$\mathbf{X}_{ML} = \arg \min_{\mathbf{X}} \|\mathbf{y} - \mathbf{H} \cdot \mathbf{X}\|. \quad (2)$$

Since its complexity exponentially increases as the number of transmit antennas and the constellation size increase, its real-time implementation is infeasible. For example, in case of  $N_T=2$  and 64QAM,  $64^2=4096$  searches is required for each received signal vector.

#### B. Prior Research in Symbol Detection Algorithm

As an alternative of MLD, the concept of sphere detection (SD) was introduced in [4] and has further discussed in various publications [5]–[6]. In order to avoid the exponential complexity of the MLD, the search for the closest lattice point is restricted to include only vector constellation points that fall within a certain search sphere. This approach allows for finding the ML solution with only polynomial complexity for sufficiently high signal-to-noise ratio (SNR) [4]. However, SD has a disadvantage that the computational complexity varies with different signals and channels. Hence, the detection throughput is non-fixed,

which is not desirable for the real-time hardware implementation. To resolve this problem, the MLD with QR decomposition and M-algorithm (QRM-MLD) was proposed in [7]. At each search layer in QRM-MLD, only the best  $M$  candidates are kept for the next level search and therefore, it has the fixed complexity and throughput which is suitable for pipeline hardware implementation. However, its complexity is still exponentially increasing with the number of transmit antennas, which may bring very large computational complexity when high spectral efficiencies are required to support higher communication rates [9].

#### C. MML Symbol Detection Algorithm

The MML technique that reduces dramatically the computational complexity of ML technique without degradation performance was introduced in [10]. Table I describes the symbol detection algorithm based on MML detection, where  $\mathbf{C}$  is the set consisting of the constellation points and  $Q(\cdot)$  represents a slicing (quantization) function. As depicted in Table I, with the MML symbol detection, the required computational number of MML metric for symbol detection is  $C^{N_T-1}$ , while that of ML metric is  $C^{N_T}$ . The MML algorithm reduces the computational complexity of the ML algorithm significantly, however, especially when supporting 64QAM, its complexity is still too high to be implemented in real-time. Therefore, efficient architecture design for the real-time implementation is required.

TABLE I. MML SYMBOL DETECTION ALGORITHM

Step	Description
(1)	$metric = \text{big number},$
(2)	$a = \mathbf{h}_2^H \mathbf{y},$
(3)	$b = \mathbf{h}_2^H \mathbf{h}_1,$
(4)	$norm = \ \mathbf{h}_2\ ^2,$
(5)	for $l = 1 :  C ,$
(6)	$x_1 = C(l),$
(7)	$\tilde{\mathbf{y}}_1 = \mathbf{y} - \mathbf{h}_1 x_1,$
(8)	$x_2 = Q(a - b x_1),$
(9)	$metric\_temp = \ \tilde{\mathbf{y}}_1 - \mathbf{h}_2 x_2\ ^2,$
(10)	if $metric\_temp < metric$
(11)	$x_{ML} = [x_1 \ x_2]^T,$
(12)	$metric = metric\_temp,$
(13)	end,
(14)	end.

### IV. DESIGN OF HARDWARE ARCHITECTURE FOR THE PROPOSED MIMO SYMBOL DETECTOR

The efficient hardware structure of the MIMO symbol detector to support all MIMO transmission modes is presented in this section. In order to achieve the more

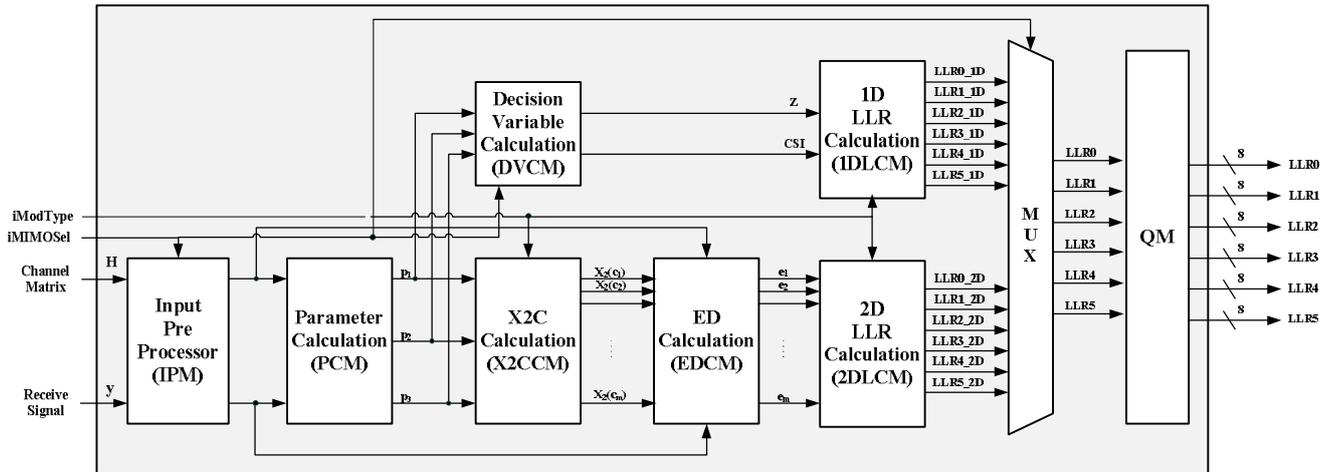


Figure 2. Block diagram of the proposed symbol detector for 2x2 MIMO systems.

reliable performance and higher-rate data transmission, the latest wireless communication systems specify to support SD mode such as single-input multiple-output (SIMO), multiple input single-output (MISO), STBC and SFBC as well as SM mode. If the symbol detector for each mode is designed independently, it is not efficient because the hardware complexity is a very burdensome to be implemented in real time. By sharing commonly used function block for all MIMO modes, the complexity of the proposed architecture is dramatically decreased.

Fig. 2 shows the proposed hardware structure of 2x2 MIMO symbol detector and Table II summarizes the procedure of proposed symbol detection to support all of transmission modes. The proposed structure of MIMO symbol detector is composed of input preprocessor module (IPM), parameter calculation module (PCM), decision variable calculation module (DVCM), X2C calculation module (X2CCM), Euclidean distance calculation module (EDCM), 1D LLR calculation module (1DLCM), 2D LLR

calculation module (2DLCM) and 8-bit quantization module (QM).

#### A. Input Preprocessor Module (IPM)

Fig. 3 shows the hardware block diagram of IPM, which set the input data of PCM for MIMO mode by reordering the estimated channel vectors and received signal vector.

Especially, the column-switching of the channel matrix,  $\mathbf{H}$ , is performed for multi-stage pipelining in case of SM mode. Since the vertical coding [13] for SM mode is generally specified in most recent wireless communication standards such as IEEE 802.16e mobile WiMAX and 3GPP LTE, LLR values from each transmit antenna need not to be generated simultaneously. Therefore, LLR values are generated sequentially by column-switching in IPM and the hardware blocks are fully shared to reduce the complexity in the proposed architecture.

TABLE II. ALGORITHMIC STEPS FOR THE PROPOSED 2X2 MIMO SYMBOL DETECTOR

STEP	BLOCK	OPERATION		
1	INPUT	$\mathbf{H} = \begin{pmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$		
2	IPM	Input of PCM, $\mathbf{a}$ , $\mathbf{b}$ , $\mathbf{c}$ , $\mathbf{d}$ , and $\mathbf{e}$ is set as depicted in Table III.		
3	PCM	$p_1 = \mathbf{a}^H \mathbf{b}$ , $p_2 = \mathbf{c}^H \mathbf{d}$ , $p_3 = \ \mathbf{e}\ ^2$		
4	X2CCM	Spatial Multiplexing (SM) Mode	Spatial Diversity (SD) Mode	
		$x_2(c_m) = \mathbf{Q}(p_1 - p_2 c_m, p_3)$ , ( $m = 1, 2, \dots, C$ )	The DVCM calculates the CSI and decision variables by using $p_1$ , $p_2$ and $p_3$ .	
5	EDCM	$e_m = \ \mathbf{y} - \mathbf{h}_1 c_m - \mathbf{h}_2 x_2(c_m)\ ^2$		DVCM
6	2DLCM	$\text{LLR} = \min(\text{bit 0 group}) - \min(\text{bit 1 group})$	1DLCM	Simplified demapping scheme in [14] is applied.
7	QM	The QM quantizes the LLR values in 8 bit.		

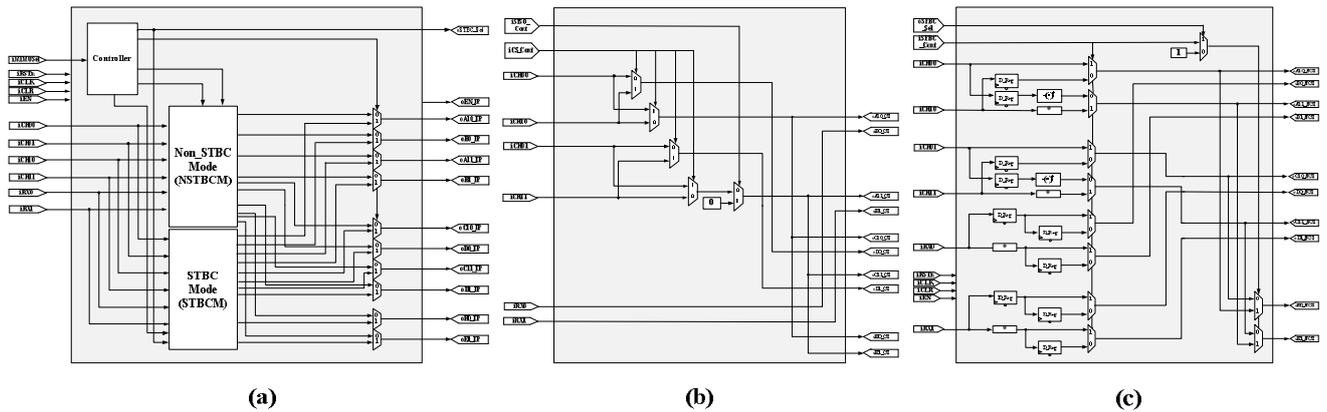


Figure 3. Block diagram of (a) IPM, (b) NSTBCM, and (c) STBCM

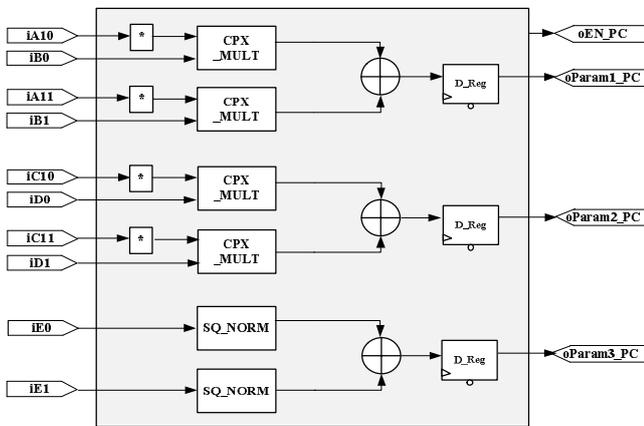


Figure 4. Block diagram of PCM

### B. Parameter Calculation Module (PCM)

As shown in Fig. 4, PCM calculates the parameters,  $p_1$ ,  $p_2$  and  $p_3$ , which are the commonly required operations for both SD and SM modes. In case of SD mode,  $p_1$  and  $p_2$  are

used for calculating the decision variables in DVCM and  $p_3$  is utilized as the channel state information (CSI). In case of SM mode, all the parameters are mapped to the input data of X2CCM.

### C. X2C Calculation Module (X2CCM)

As shown in Fig. 5, X2CCM consists of the polar-coordinate based multiplier (PBM) and slicer module (SCM). SCM makes the output,  $x_2(c_m)$ ,  $m=1, 2, \dots, C$ , and is implemented without division operations through the scaled-constellation as in (5).

$$x_2(c_m) = \mathbf{Q} \left( \frac{\mathbf{h}_2^H}{\|\mathbf{h}_2\|^2} [\mathbf{y} - \mathbf{h}_1 c_m] \right) = \mathbf{Q}(p_1 - p_2 c_m, p_3). \quad (5)$$

In order to calculate  $p_2 c_m$  in (5),  $C$  number of complex multiplications should be performed in parallel, which makes it very difficult to design of X2CCM. For example, in case of 64QAM, 64 complex multiplications are required.

TABLE III. INPUT DATA MAPPING SCHEME OF PCM FOR MIMO MODE

MODE	$p_1 = \mathbf{a}^H \mathbf{b}$		$p_2 = \mathbf{c}^H \mathbf{d}$		$p_3 = \ \mathbf{e}\ ^2$	
SISO / SIMO	$\mathbf{a} = \begin{pmatrix} h_{11} \\ 0 \end{pmatrix} / \mathbf{a} = \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} y_{11} \\ y_{12} \end{pmatrix}$		-		$\mathbf{e} = \mathbf{a}$	
MISO	1st clk.	2nd clk.	-		1st clk.	2nd clk.
	$\mathbf{a}_1 = \begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix} \quad \mathbf{b}_1 = \begin{pmatrix} y_{11} \\ y_{12}^* \end{pmatrix}$	$\mathbf{a}_2 = \begin{pmatrix} h_{21} \\ -h_{11} \end{pmatrix} \quad \mathbf{b}_2 = \begin{pmatrix} y_{11} \\ y_{12}^* \end{pmatrix}$			$\mathbf{e}_1 = \mathbf{a}_1$	$\mathbf{e}_2 = \mathbf{a}_2$
STBC & SFBC	1st clk.	2nd clk.	1st clk.	2nd clk.	1st clk.	2nd clk.
	$\mathbf{a}_1 = \begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix} \quad \mathbf{b}_1 = \begin{pmatrix} y_{11} \\ y_{12}^* \end{pmatrix}$	$\mathbf{a}_2 = \begin{pmatrix} h_{21} \\ -h_{11} \end{pmatrix} \quad \mathbf{b}_2 = \begin{pmatrix} y_{11} \\ y_{12}^* \end{pmatrix}$	$\mathbf{c}_1 = \begin{pmatrix} h_{12} \\ h_{22} \end{pmatrix} \quad \mathbf{d}_1 = \begin{pmatrix} y_{21} \\ y_{22}^* \end{pmatrix}$	$\mathbf{c}_2 = \begin{pmatrix} h_{22} \\ -h_{12} \end{pmatrix} \quad \mathbf{d}_2 = \begin{pmatrix} y_{21} \\ y_{22}^* \end{pmatrix}$	$\mathbf{e}_1 = \mathbf{a}_1$	$\mathbf{e}_2 = \mathbf{a}_2$
SM	1~4 clk. (LLR1)	5~8 clk. (LLR2)	1~4 clk. (LLR1)	5~8 clk. (LLR2)	1~4 clk.	5~8 clk.
	$\mathbf{a}_1 = \begin{pmatrix} h_{21} \\ h_{22} \end{pmatrix} \quad \mathbf{b}_1 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$	$\mathbf{a}_2 = \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} \quad \mathbf{b}_2 = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$	$\mathbf{c}_1 = \begin{pmatrix} h_{21} \\ h_{22} \end{pmatrix} \quad \mathbf{d}_1 = \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix}$	$\mathbf{c}_2 = \begin{pmatrix} h_{11} \\ h_{12} \end{pmatrix} \quad \mathbf{d}_2 = \begin{pmatrix} h_{21} \\ h_{22} \end{pmatrix}$	$\mathbf{e}_1 = \mathbf{a}_1$	$\mathbf{e}_2 = \mathbf{a}_2$

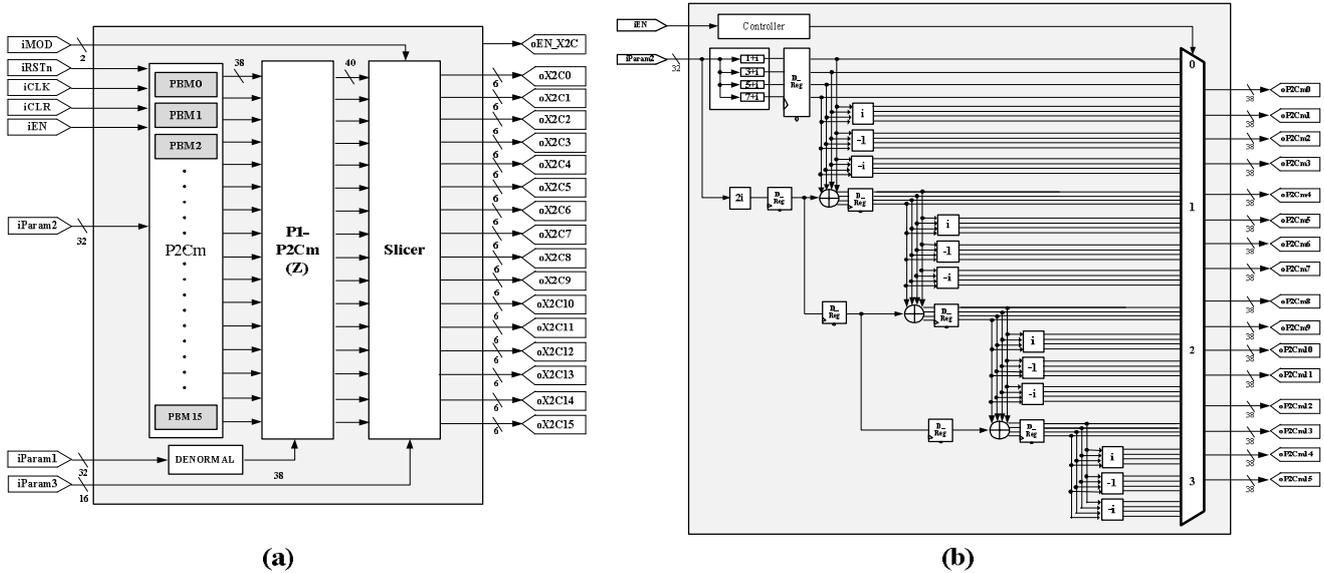


Figure 5. Block diagram of (a) X2CCM and (b) PBM

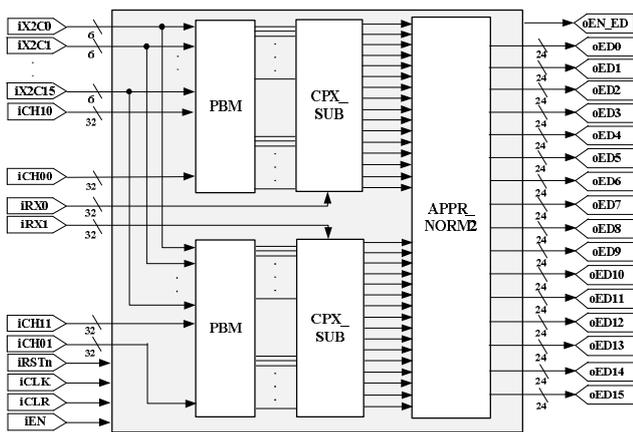


Figure 6. Block diagram of EDCM

In the proposed architecture, the complex multiplication is replaced by the PBM which can be simply implemented with sign-inverter, shifters and adders as in Fig. 5(b) because  $c_m$  in constellation is constant and symmetric. Especially, PBM is designed with 4-stage pipeline architecture to reduce the computational complexity by sharing the hardware resources. Although the throughput performance may be degraded, it is practically negligible because the throughput bottleneck of the baseband modem is mostly in the forward error correction (FEC) module such as turbo decoder. For example, when the proposed detector is applied to the mobile WiMAX baseband processor including turbo decoder with 6 iterations, it is verified from the timing analysis that 4-stage pipelining of PBM does not make any degradation of throughput performance.

D. ED Calculation Module (EDCM)

EDCM calculates the Euclidean distance,  $e_m$ , which is given by

$$e_m = \|y - h_1 c_m - h_2 x_2(c_m)\|^2 \tag{6}$$

As shown in Fig. 6,  $h_1 c_m$  and  $h_2 x_2(c_m)$  is computed by PBM, and norm calculation is approximated as in [15], which makes the negligible performance degradation as shown in the performance evaluation results of Fig. 7 for a Rayleigh fading channel with additive white Gaussian noise (AWGN). After calculating ED, log-likelihood ratio (LLR) values are computed from 1DLCM or 2DLCM, and the final quantized LLR value is generated by 8-bit QM for FEC module such as turbo decoder. The quantization is performed through the analysis for the distribution of the calculated LLR value from 1DLCM or 2DLCM.

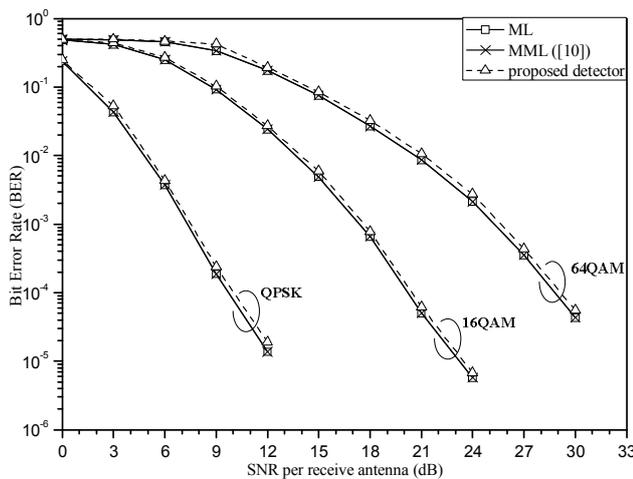


Figure 7. Performance evaluation results of the proposed symbol detector

## V. IMPLEMENTATION RESULTS

The MIMO symbol detector supporting all MIMO modes with the proposed architecture was designed in HDL and synthesized to gate-level circuits using 0.13 $\mu$ m CMOS standard cell library. Table IV depicts the logic synthesis results for 80MHz operating clock frequency of the proposed MIMO symbol detector, which shows that EDCM is the most complex block in the proposed detector.

Table V shows the comparison results of our design and the existing detectors. Even though the proposed detector can support all MIMO modes such as SD and SM, its complexity is rather reduced by 57% compared with that of [10]. Compared with the results in [16], the proposed detector has the similar complexity. However, it can support all MIMO modes and the optimal ML performance.

TABLE IV. LOGIC SYNTHESIS RESULTS OF THE PROPOSED MIMO SYMBOL DETECTOR

	Gate Count (K)	Prop. (%)
IPM	11.4	2.9
PCM	32.2	8.2
X2CCM	31.0	7.7
EDCM	255.8	64.5
1DLCCM	4.2	1.2
2DLCCM	28.0	7.1
QM	1.0	0.2
<i>Etc.</i>	<i>29.4</i>	<i>8.2</i>
<b>Total</b>	<b>393 K</b>	<b>100 %</b>

TABLE V. COMPARISON RESULTS OF THE PROPOSED DESIGN AND EXISTING DETECTORS

	[10]	[16]	Proposed
MIMO Configuration	SM Only	SM Only	SM & SD
Detection Algorithm	MML	LORD*	MML
Gate count	921 K	408 K	393 K

\* LORD denotes Layered Orthogonal Lattice Detector, which supports near-ML performance.

## VI. CONCLUSION

In this paper, we proposed an area-efficient hardware architecture for the MIMO symbol detector which can support all MIMO modes such as SD and SM. With the multi-stage pipelining, simplified multiplication based on polar-coordinate, and the approximation of norm operation, the complexity of the proposed detector is reduced by 57% compared with that of the conventional architecture. Since the recent wireless systems specify to support both SD and SM modes and need to be implemented with low-complexity and low-power consumption, the proposed MIMO symbol detector can be considered to be suitable for those systems.

## ACKNOWLEDGMENT

This work was supported by the Technology Innovation Program, 10035389, funded by the Ministry of Knowledge Economy (MKE, Korea) and CAD tools were supported by IDEC.

## REFERENCES

- [1] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bölcskei, "An overview of MIMO communications—A key to gigabit wireless," *IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.
- [2] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal design," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, July 1999.
- [3] D. Wübben, R. Böhnke, V. Kühn, and K. D. Kammeyer, "MMSE extension of V-BLAST based on sorted QR decomposition," *IEEE VTC*, vol. 1, pp. 508–512, Oct. 2003.
- [4] B. Hassibe and H. Vikalo, "On sphere decoding algorithm I. Expected complexity," *IEEE Trans. Signal Processing*, vol. 53, no. 8, pp. 2806–2818, Aug. 2005.
- [5] J. Jalden and B. Ottersten, "An exponential lower bound on the expected complexity of sphere decoding," *IEEE ICASSP*, pp. 393–396, 2004.
- [6] H. Kawai, K. Higuchi, N. Maeda, and M. Sawahashi, "Adaptive control of surviving symbol replica candidates in QRD-MLD for OFDM MIMO multiplexing," *IEEE J. Sel. Areas Commun.*, pp. 1130–1140, 2006.
- [7] C. Studer, A. Burg, and H. Bölcskei, "Soft-output sphere decoding: Algorithms and VLSI implementation," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 2, pp. 290–300, Feb. 2008.
- [8] S. Noh, Y. Jung, S. Lee, and J. Kim, "Low complexity symbol detector for MIMO-OFDM-based wireless LANs," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 12, pp. 1403–1407, Dec. 2006.
- [9] S. Mondal, A. Eltawil, S. Chung-An, and K.N. Salama, "Design and Implementation of a Sort-Free K-Best Sphere Decoder," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 18, no. 10, pp. 1497–1501, Oct. 2010.
- [10] S. Yu, T. Im, C. Park, J. Kim, and Y. Cho, "An FPGA Implementation of MML-DFE for Spatially Multiplexed MIMO Systems," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 55, no. 7, pp. 705–709, Jul. 2008.
- [11] C. Huang, C. Yu, and H. Ma, "A Power-Efficient Configurable Low-Complexity MIMO Detector," *IEEE Trans. Circuits Syst. I*, vol. 56, no. 2, pp. 485–496, Feb. 2009.
- [12] D. Wu, E. G. Larsson, and D. Liu, "Implementation aspects of fixed-complexity soft-output MIMO detection," *IEEE Vehicular Technology Conf. (VTC)*, April 2009.
- [13] B. Vucetic and J. Yuan, "Space-Time Coding." Wiley, 2003.
- [14] F. Tosato, P. Bisaglia, "Simplified soft-output demapper for binary interleaved COFDM with application to HIPERLAN/2," *IEEE ICC*, vol.2, pp.664–668, May 2002.
- [15] A. Adjoudani, E.C. Beck, A.P. Burg, G.M. Djuknic, T.G. Gvoth, D. Haessig, S. Manji, M.A. Milbrodt, M. Rupp, D. Samardzija, A.B. Siegel, T. Sizer, C. Tran, S. Walker, S.A. Wilkus, and P.W. Wolniansky, "Prototype experience for MIMO BLAST over third-generation wireless system," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, pp. 440–451, Apr. 2003.
- [16] T. Cupaiuolo, M. Siti, and A. Tomasoni, "Low-complexity high throughput VLSI architecture of soft-output ML MIMO detector," *Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1396–1401, Mar. 2010.

# Software Defined Radio as a DS-CDMA Receiver

Ryszard J. Katulski

Gdansk University of Technology  
Department of Radiocommunications Systems and  
Networks  
Gdansk, Poland  
[rjkat@eti.pg.gda.pl](mailto:rjkat@eti.pg.gda.pl)

Wojciech Siwicki

Gdansk University of Technology  
Department of Radiocommunications Systems and  
Networks  
Gdansk, Poland  
[wojciech.siwicki@eti.pg.gda.pl](mailto:wojciech.siwicki@eti.pg.gda.pl)

Jacek Stefański

Gdansk University of Technology  
Department of Radiocommunications Systems and Networks  
Gdansk, Poland  
[jstef@eti.pg.gda.pl](mailto:jstef@eti.pg.gda.pl)

**Abstract** — Programmable radio is one of the latest trends in the construction of multi-standard receivers. The technology, called Software Defined Radio (SDR), is also an ideal test platform that allows trying out different algorithms of signal receiving. This particular feature led to choose this platform to implement a DS-CDMA receiver (Direct Sequence Code Division Multiple Access). This receiver is a part of the radiolocation system called AEGIR. The main task of the receiver is to calculate geographical position on basis of received signals from ground stations. That is why Software Radio has been chosen. It allows a gradual upgrade of data processing algorithms. The main goal was to create a wireless receiver that allows calculating the correct position on the basis of the received signals and the processing time will be minimized as much as possible. This paper mainly focuses on algorithms in terms of correctness of the received signal and, what is equally important, its processing time.

**Keywords** – Programmable radio; Software Defined Radio; SDR; phase correction; AEGIR

## I. INTRODUCTION

The variety of standards for radio systems makes it necessary to construct a mobile terminal that has the technical possibilities of cooperation with different radio standards. This simple concept created an idea of a programmable radio called Software Defined Radio (SDR), based on a universal hardware layer, with only a layer of software determining its functionality [4-8].

One of the Software Defined Radio features is great flexibility; that is why this platform has been chosen to implement a localization receiver for AEGIR system [1-3]. The use of SDR technology in this approach allowed a flexible shaping of functionality of the receiver. This technique is also an ideal testing platform, because it allows implementing and testing different algorithms.

It consists of a wideband receiver and a personal computer (PC) equipped with an acquisition card. PC with appropriate

software is used to control both the receiver and the acquisition card; also it serves for signal processing.

In this article, a basic concept of SDR will be presented. Next, the physical implementation of DS-CDMA signal receiver and its structure will be shown. Selected parameters of received signal and encountered problems with data processing (e.g., initial phase correction, processing time) will be presented as well.

## II. SOFTWARE DEFINED RADIO

The concept behind Software Defined Radio is to implement – to the greatest possible extent – signal processing blocks of a radio transceiver in software rather than in dedicated hardware. The differences between the classic "analogue" version of a receiver and the programmable one are illustrated, respectively, in Figures 1 and 2 [4], [5].

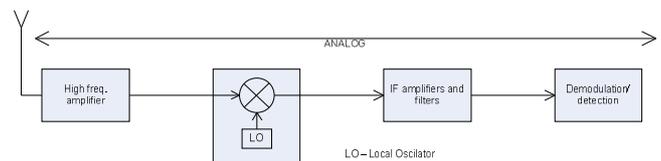


Figure 1. Block diagram of analogue receiver

Receiver shown in Figure 2 can be divided into two different parts of a system:

- hardware (analogue radio) in the form of a set of classic radio components,
- software (digital), whose main element is fast signal processor DSP (Digital Signal Processor).

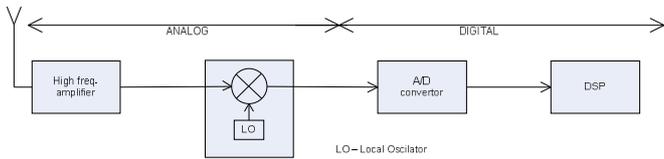


Figure 2. Block diagram of Software Defined Radio

Conceptually speaking, a SDR should have the following properties [7]:

- reconfigurable RX/TX architecture, controlled by software,
- most part of the radio functionality performed by software,
- system specification (bandwidth, bit rate, demodulation) can be updated whenever needed to do so.

The task of the analogue radio part is to strengthen appropriately and convert the received radio signal from the high-band radio frequency to intermediate frequency band. Then, in this band with a fast A/D converter (Analogue to Digital Converter) a received analogue signal is converted into its digital form. Processing is performed in a properly programmed digital signal processor.

III. STATE OF ART

One of the first steps was to check the literature of already implemented receivers in SDR technology. Most of the articles concerned implementations as computer simulations [9-10]. There were also articles on the hardware implementation based on FPGA (Field Programmable Gate Array) platform [11-12]. The concept of building a receiver on a FPGA platform has been rejected for the reasons explained later in this article.

Due to the fact that the whole radiolocation system was designed from scratch, the receiver has been created from scratch as well. The aim was to develop a simple and fast algorithm that would allow receiving signals from several sources.

IV. HARDWARE IMPLEMENTATION

Figure 3 shows the diagram of the receiver made in software defined radio technology.

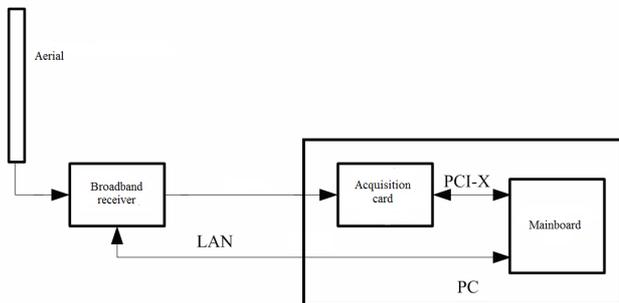


Figure 3. Block diagram of programmable receiver

It consists of a broadband receiver and a PC fitted with a data acquisition card (Figure 4), whereby the analogue radio signal is converted into a sequence of discrete data samples. Computer task is to control the operation of the receiver and to process data received from data acquisition card.

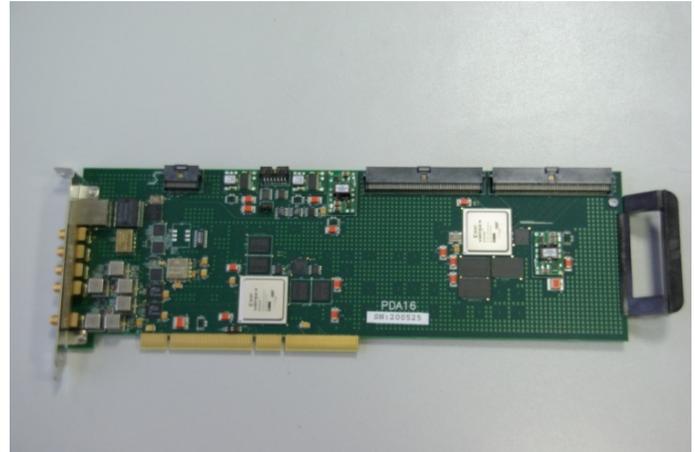


Figure 4. Acquisition card

Parameters of a PC are:

- Processor: i7 980EE (6 cores/12 threads)
- RAM: 6GB
- HDD: 1TB
- Mainboard with PCI-X slot for acquisition card.

Frame structure of receiving data is presented on Figure 5.

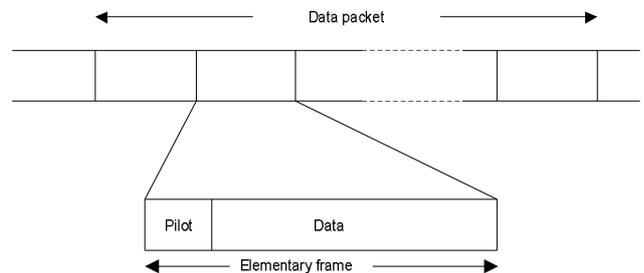


Figure 5. Frame structure

It consists of known sequence of pilot bits and sequence of transmitted data. Sequence of pilot bits are used to improve time of demodulation and are used for initial phase correction. Both mentioned algorithms will be presented later in this article.

Actual received signal parameters are as follows:

- The signal carrier frequency: 450MHz,
- Data rate: 1kHz,
- Bandwidth after spreading: 1MHz,
- Modulation: QPSK,

- In-phase and quadrature-phase components carry independent data.

As already mentioned, this receiver is part of the radiolocation system AEGIR. For this reason it was decided to build a programmable radio, based on a PC computer (instead of FPGA), which allows to record data obtained from an acquisition card to hard disk and perform test with upgraded algorithms in a post processing.

V. ALGORITHMS

A. Main algorithm

The main algorithm is shown on Figure 6. The program can be divided into two segments. First: configuration and data acquisition, second – digital signal processing.

Program begins with launching and configuring the data acquisition card and setting the parameters of the receiver. Then, the process of acquisition (writing to a binary file data samples) begins. After a certain time, acquisition is completed and acquisition card goes into standby mode. The next step is to process the collected samples. They are multiplied by carrier (both the cosine and sine) and by the spreading sequences.

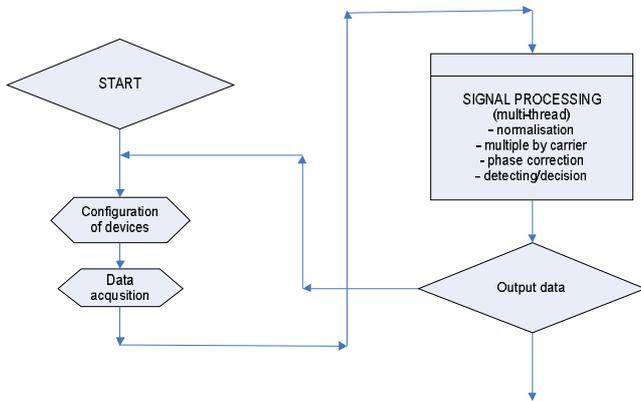


Figure 6. Main algorithm

B. Phase correction algorithm

During the processing of data, we should consider errors resulting from inaccuracies in the internal clocks of the receiver and data acquisition card. Correction is being determined after receiving a few bits; next, it is added in the following stage of detection. This frequency drift and phase error can be illustrated as a constellation of the received signal. Due to the independent component in in-phase and quadrature-phase, received signal can be detected as two BPSK modulation. In this case when *Re value* (read from real axis) is greater then 0 it means that we received logical '1' if is lower then 0 – logical '0'.

Bearing in mind the above, received signal constellation may take the form shown on Figure 7. Bellow the following sequence of bits has been given for illustration:

no	1	2	3	4	5	6	7	8	...	16	17	18	19
bit	1	1	1	1	1	0	0	1	...	0	1	1	0

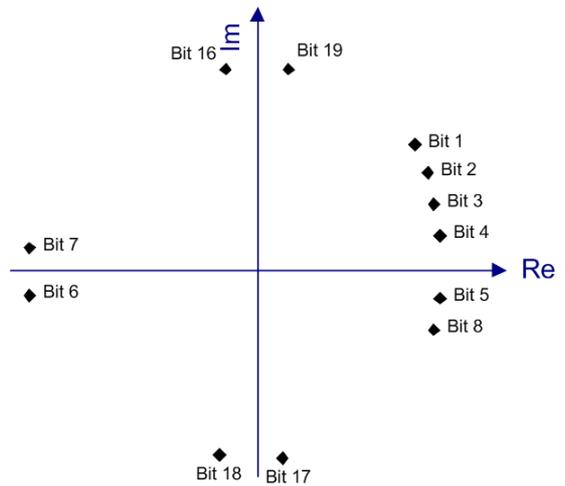


Figure 7. BPSK constellation of a received signal

Analysing the presented sequence with constellation shown in Figure 7 it can be seen that bits 18 and 19 without phase correction have opposite values. Phase correction is based on known sequence of pilot bits. Because sequence is known, correction can be calculated (in order to 'twist' received point on a constellation) and applied on next received bit.

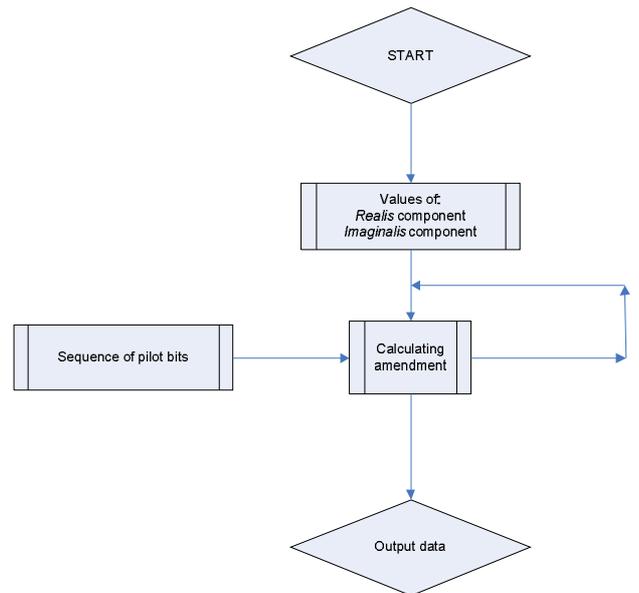


Figure 8. Algorithm of phase correction

After applying the correction phase algorithm (Figure 8), both points are on the side of their actual representation (Figure 9).

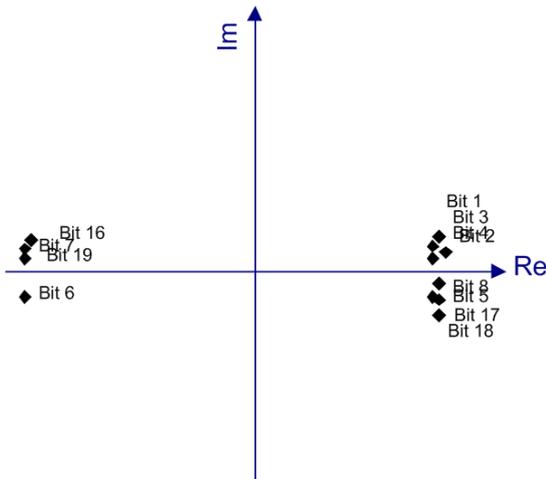


Figure 9. BPSK constellation after phase correction

C. Processing time

Time of signal processing is one of the most important parameters of a programmable radio. It is recommended to obtain the strongest possible CPU for a PC. If possible, convert each calculating subprogram to the application, using several processor cores. Only the development of multithreaded applications can effectively cope with the complexity of computing with which we meet in implementing a programmable radio. The hardware layer affects the processing speed; however, it is equally important to optimize the code. Also an algorithm of decoding/detection is very important as well.

Figure 10 shows first algorithm of data decoding. It is based on a classic correlation technic. Received data are correct, but time of computing is very long.

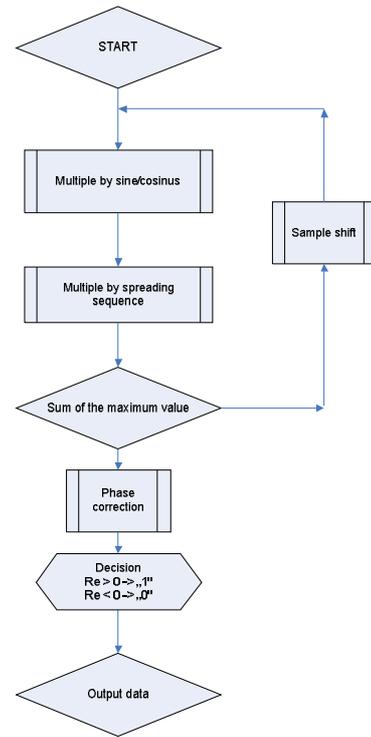


Figure 10. Algorithm of decoding

Because of long time of computing another algorithm was developed. Figure 11 presents this algorithm. It uses Fast Fourier Transformation (FFT) to calculate where the pilot bits start. Based on these results, data are multiplied by spreading sequence already at those locations and get *Re* and *Im* components.

Using the first algorithm, attempts to receive transmitted signal took over an hour of processing time (for 40 seconds of acquisition). Second algorithm reduced this time to approximately 40-60 seconds. The software was rewritten to work in a multithread environment. This procedure allows receiving signal and decoding information from many sources without significant time consumption.

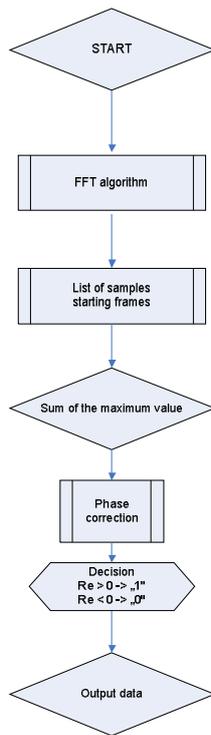


Figure 11. Algorithm of decoding

Given the fact that it is a locational receiver, it was necessary to make such changes in the algorithm that calculated position was generated every few seconds. To this end, the algorithm has undergone another change. Two types of length of acquisition were chosen. Long one with a minimum sampling frequency is used to store long sequences respectively, which allows reading of data contained in the radio signal and a short one at the higher sampling rate, which serves only to determine the differences of time needed to calculate the correct geographical position. Short acquisition time was reduced to about 2 seconds. Increased sampling frequency resulted in much more data to process. That is why FFT algorithm was internally divided to work in a few threads. Through this approach, a short acquisition (two seconds) gave expected time of computation (i.e., generating geographic position) of 4-5 seconds.

## VI. CONCLUSION AND FUTURE WORK

Application of the DS-CDMA signal receiver in programmable radio technology allows a flexible approach to updating and verification of the implemented software receiver. It also allows implementing several methods of reception in order to compare their efficiency, correctness and processing time. Studies conducted in the lab and in real conditions confirm the versatility of the platform.

Mentioned AEGIR system was tested twice in real conditions at the Bay of Gdansk. Last test were carried out in October 2010. System is still being updated. Next test will be performed in June 2011 and another bandwidth of spreaded signal will be tested (respectively 2MHz and 4MHz).

## ACKNOWLEDGMENT

This research, hereby described, is funded by the Polish Ministry of Science and Higher Education as a part of research and development project No O R00 0049 06. The authors express their sincere thanks for allocated funds for this purpose.

## REFERENCES

- [1] Ambroziak S.J., Katulski R.J., Sadowski J., Siwicki W. and Stefański J., "Autonomous ground-based and self-organized radiolocation systems", Naval Armament & Technology Conference NATCon'2010, Gdynia 20-21.X.2010.
- [2] Patent application no. P393181, "Asynchronous system and method for determining position of persons and/or objects" (in polish), 2010-12-08.
- [3] European patent application no. EP11460023, *Asynchronous system and method for estimating position of persons and/or objects*, May 2011.
- [4] Mitola III J.; *Software Radio Architecture*, A Wiley & Sons 2000.
- [5] Buracchini E.; *The Software Radio Concept*, IEEE Commun. Mag., vol 38, No. 9, Sep. 2000, pp. 138-143.
- [6] Chamberlain, M.W.; *A software defined HF radio*, Military Communications Conference, 2005. MILCOM 2005. IEEE, 17-20 Oct. 2005, vol. 4, pp. 2448 – 2453.
- [7] Duan L.; *The Key Issues to Design Software Radio*, Radio Science Conference, 2004. Asia-Pacific, 24-27 Aug. 2004, pp. 119 – 122.
- [8] Katulski R., Marczak. A., and Stefański J.; "Software Radio Technology" (in polish), Telecommunication review and telecommunication news No. 10/2004, pp. 402-406.
- [9] Fulghum T.L., Cairns D.A., Cozzo C., Wang Y.-P. E. and Bottomley G.E.; "Adaptive Generalized Rake Reception in DS-CDMA Systems", IEEE Transactions on wireless communications, vol. XX, no. XX, 2009, pp. 1-11
- [10] Abuelmaatti A., Thayne I. and Beaumont S.; "A new approach to QPSK: Mechanism and Implementation", IEEE Wireless Communications and Networking Conference 2007, pp. 2393-2398
- [11] Correal N.S., Buehrer R.M. and Woerner B.D.; "A DSP-Based DS-CDMA Multiuser Receiver Employing Partial Parallel Interference Cancellation", IEEE Journal on selected areas in communications, vol. 17, no. 4, April 1999, pp. 613-630
- [12] Zhuo-zhi D., Hao-xian W., Ying-hong S. and Xing-peng M.; "The FPGA implementation of high-speed DS-CDMA receiver", 2010 First International Conference on Pervasive Computing, Signal Processing and Applications, pp. 751-754.

## Performance Improvements of Mobile SCTP during handover period

Sinh Chung Nguyen, Thi Mai Trang Nguyen and Guy Pujolle

*University of Pierre and Marie Curie (UMPC - Paris 6)*

*Laboratoire d'Informatique de Paris 6 (LIP6)*

*4 Place Jussieu, 7005 Paris, France*

*Emails: {Sinh-Chung.Nguyen, Thi-Mai-Trang.Nguyen, Guy.Pujolle}@lip6.fr*

**Abstract**—Handling mobility at the transport layer is an interesting approach to attain seamless handover in the context of heterogeneous access networks. To support terminal mobility, an extension of SCTP called SCTP-DAR (SCTP Dynamic Address Reconfiguration) (a.k.a. mobile SCTP) has been defined allowing SCTP endpoints to update IP address and inform the remote peer about primary address change. However, mobile SCTP suffers from performance degradation during handover. In this paper, we investigate two handover-related issues: reduction of data rate immediately after handover and packet reordering. The first issue is caused by the congestion control mechanism, which forces the ongoing traffic to enter the slow-start phase when moving into the new path after handover. The second issue is caused by the asymmetric characteristics of access networks such as the different bandwidths and delays in heterogeneous access networks. In order to solve these issues, we propose to use a combination of two mechanisms, congestion control parameter update and buffer retransmission.

**Keywords**—SCTP; mSCTP handover; multi-homing; congestion control; vertical handover.

### I. INTRODUCTION

Stream Control Transmission Protocol [1] has been designed by the IETF SIGTRAN (Signaling Transport) working group as a transport protocol for the delivery of packet-based PSTN (Public Switched Telephone Network) signaling over IP networks. Similar to TCP (Transmission Control Protocol), SCTP offers reliable data transfer and has mechanisms for congestion control, data duplication detection and data retransmission. In addition, it supports two new features, which are multi-streaming and multi-homing. Even though SCTP was explicitly designed for signaling transport, the multi-homing feature of this protocol can be used for broader applications and becomes an interesting solution to support multi-homing in multi-interface mobile terminals [2]. Each SCTP endpoint owns a list of IP addresses. During the initial phase of the association, SCTP endpoints have to exchange their lists of addresses. Among these IP addresses, the endpoint chooses one address as primary address. The others are secondary ones. Based on this principle, an SCTP association has a primary path and several secondary paths. If a failure occurs along the primary path, SCTP switches to an alternative path to maintain the communication.

To support the mobility of a mobile terminal, an extension of SCTP called mobile SCTP (mSCTP) [3] using SCTP

Dynamic Address Reconfiguration [4], allows endpoints to update the IP address lists in an SCTP association. Each endpoint can update its address list if a new address is received or a current address becomes obsolete. This update procedure typically happens during the handover of a mobile terminal. For example, an endpoint moving from one access point to another may receive a new IP address from the new access point and the old IP address is no longer reachable. Thus, the endpoint needs to add a new IP address and deletes the old one from the address list of the association [5]. Mobile SCTP also allows an endpoint to inform the remote peer about its primary address change. For updating address lists and changing the primary address, mSCTP defines two new chunks: Address CONFIguration Change Chunk (ASCONF) and Address CONFIguration ACKnowledgment Chunk (ASCONFACK) [4].

However, in the context of heterogeneous networks with the existing congestion control mechanism and multi-homing feature, mSCTP has to suffer from two problems: (i) entrance into slow-start phase of new primary path just after handover, and (ii) packet reordering problem occurring in handover between heterogeneous access networks. These problems seriously cause mSCTP performance reduction after handover.

Some existing studies have already tried to improve handover performance of the mSCTP handover in the context of heterogeneous wireless networks. To provide a better throughput in handover procedure, [6] proposes to duplicate data on both existing primary path and new added one. This proposition is only useful in high loss rate environment, but it will create redundancy in stable context. In [7], the authors present a SCTP Efficient Flow Control (SCTP EFC) mechanism in order to minimize the change of traffic rate during handover. SCTP EFC updates new primary path with the same information of the old one to bypass slow-start phase. However, in heterogeneous networks environment, access network features such as bandwidth and delay may be different, so using the same information for new primary path update is not efficient. Another interesting approach named Buffer retransmission, which is introduced in [8], aims to ameliorate mSCTP handover performance but this solution can resolve only the packet reordering problem. In this paper, we propose to use a combination of two

mechanisms: congestion control parameter update and buffer retransmission. The main contribution is a new congestion control parameter update mechanism based on bandwidth estimation to avoid the slow-start phase in the new primary path after handover. Buffer retransmission proposed in [8] is used to overcome the packet reordering problem.

The remainder of this paper is organized as follows. Section II studies the two problems encountered in SCTP handover procedure. In Section III, we present the details of our proposition on congestion control parameter update and the use of this mechanism in combination with buffer retransmission. Section IV describes the simulation scenario and results. Finally, we conclude the paper in Section V.

## II. PROBLEM DESCRIPTION

Two main problems that the SCTP protocol can have in a handover between heterogeneous wireless networks are performance degradation due to the slow-start phase and packet reordering. These problems cause SCTP performance reduction after handover period. This section explains the reasons and the conditions, in which these problems can occur.

### A. Problem related to the slow-start phase

The entrance in slow-start phase can reduce throughput after handover. Like TCP, SCTP also supports flow control and congestion control mechanisms such as slow-start, congestion avoidance, fast recovery, etc. In terms of control variables, endpoint maintains only one receiver window (rwnd) for the whole association. Other control variables such as congestion window (cwnd), slow-start threshold (ssthresh) and RTT (Round-Trip Time), are kept on a per destination address basis. Because SCTP uses only one address at a time for transmission, SCTP handover can be understood as a procedure, in which an endpoint replaces the primary address with another one and redirects data flows from the old path to the new one. Since the ssthresh variable is controlled on a per destination address basis, the ongoing traffic must enter the slow-start phase on the new primary path after handover. This procedure leads to an immediate throughput reduction.

Figure 1 illustrates the effect of slow-start phase to mSCTP handover performance. We use Cumulative Transmission Sequence Number Acknowledgement (TSN Ack), which represents the last data sequence number received successfully, as a performance metric. As seen in Figure 1, handover execution takes place at 41.9 seconds. After that, ongoing traffic enters slow-start phase on the new primary path and its congestion control parameters start with initial values. Therefore, right after handover procedure, cumulative TSN Ack increases slowly, which proves a throughput reduction.

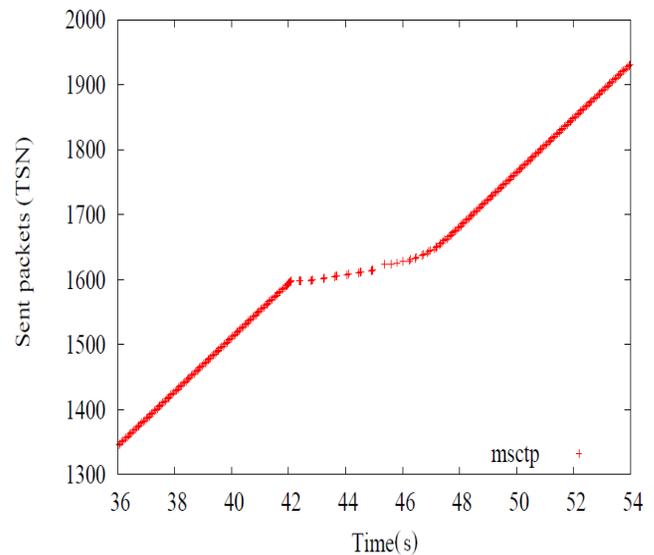


Figure 1. Throughput reduction after handover

### B. Problem related to packet reordering

In the SCTP primary path switching operations, the packet reordering problem may occur due to the asymmetric characteristics of access networks such as bandwidth and delay [8][9]. When a Mobile node (MN) moves between heterogeneous wireless networks (e.g from 3G to WLAN), the mSCTP endpoint should change its primary address to a newly received IP address. It is obvious that characteristics of heterogeneous access networks are not symmetric. If SCTP endpoints switch the primary path into a new path with higher bandwidth (or smaller delay), data chunks sent over the new path can arrive at peer endpoint earlier than those sent to the old path with lower bandwidth (or bigger delay). Therefore, packet may arrive at peer endpoint out-of-order, which is referred to the packet reordering problem. Consequently, endpoint can evoke spurious fast retransmission and unnecessarily reduce congestion window, which causes degradation of throughput in handover period.

Figure 2 shows an example of packet reordering problem due to the asymmetric characteristics between paths in one association. Each endpoint has two interfaces, interface 1 and interface 2. At the beginning, they use interface 1 (path 1) for their communication. When handover happens, the traffic is switched to interface 2 (path 2). We assume that the bandwidth of path 2 via interface 2 is higher than the one of path 1. As shown in Figure 2, endpoint S sends 4 data packets (TSNs 1-4) to endpoint R via path 1 just before the handover operation. After the handover, endpoint S uses the new path, path 2, to send new data packets. Because of the different bandwidths, data packets sent via path 2 (TSNs 5-6) arrive at endpoint R earlier than ones sent via

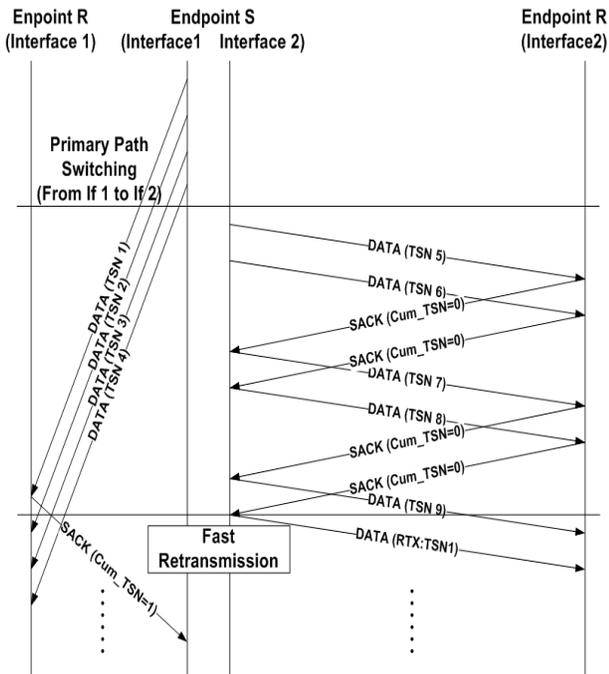


Figure 2. Packet reordering problem in mSCTP handover [8]

path 1. Therefore, a notification about packet reordering is sent to endpoint S via SACK chunk. When the accumulation of duplicate SACKs of the same gap received at endpoint S equals four, endpoint S triggers spurious fast retransmission immediately.

To solve this problem, the authors of [8] propose a buffer retransmission mechanism, which retransmits all outstanding data over the new primary path. All outstanding data are packets that have been transmitted over the old primary path but have not been acknowledged. It means that, after primary path change, endpoint will not send new data until all outstanding data are transmitted and acknowledged. Consequently, spurious fast retransmission and unnecessary reduction of congestion window may be avoided.

### III. PROPOSED SCHEMES

In order to deal with these problems of performance reduction, we present a solution combining two mechanisms: congestion control parameter update and buffer retransmission proposed in [8]. Although each mechanism helps to solve only one problem described above, we will show in the following of this section that the combining of both has better results.

#### A. Proposal of congestion control parameter update

In order to bypass slow-start phase after the change of primary path in handover, we propose a mechanism named congestion control parameter update. The mechanism principle is forcing the new primary path to update its congestion

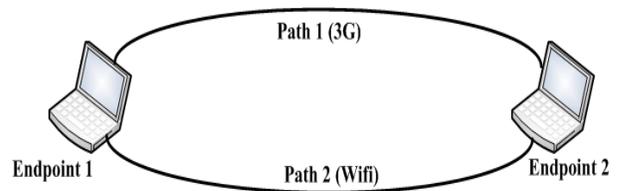


Figure 3. Network topology

control parameters using the old ones information. Just after the primary path switching operation, the congestion window (cwnd) and threshold (ssthresh) parameters of new primary path are calculated by the following formulas:

$$cwnd_{new} = \alpha \times cwnd_{old} \quad (1)$$

$$ssthresh_{new} = \alpha \times ssthresh_{old} \quad (2)$$

$$\alpha = \frac{bandwidth_{new}}{bandwidth_{old}} \quad (3)$$

in which  $\alpha$  is the bandwidth ratio of the two networks concerning handover.

With this mechanism, the new primary path will conserve the old ones status. For example, if the old primary path is in congestion avoidance status before the handover, the new primary path will be also in congestion avoidance status after handover. Therefore, mSCTP may avoid entering slow-start phase and its performance can be improved.

For bandwidth estimation purpose, in the old primary path, we use available bandwidth, which is equal to  $\frac{cwnd}{RTT}$ . This estimation reflects real bandwidth that the old path can support for data transfer. In the new primary path after handover, bandwidth is estimated thanks to Packet Pair algorithm [10]. It means that, sender sends two HeartBeat packets back-to-back via the new primary path. The size of probe HeartBeat packet is set to the Maximum Transfer Unit (MTU) [9]. Therefore, the available bandwidth estimation (ABE) of new primary path is calculated by

$$ABE = \frac{S_{HB}}{\Delta_T} \quad (4)$$

in which  $S_{HB}$  is the size of probe HeartBeat packet, and  $\Delta_T$  is the disperse time between the probe HeartBeat acks.

### IV. SIMULATION AND RESULTS

In this section, we present the performance evaluation of the proposed solution. In order to analyze mSCTP handover performances, we use the network simulator ns-2.33 [11].

#### A. Simulation scenarios

In this simulation, each endpoint has two interfaces: one representing 3G interface, which has a bandwidth of 384Kbps and a delay time of 150 ms, and one representing Wi-Fi interface, which has a bandwidth of 11 Mbps and a delay time of 4 ms. Although we use wired interfaces

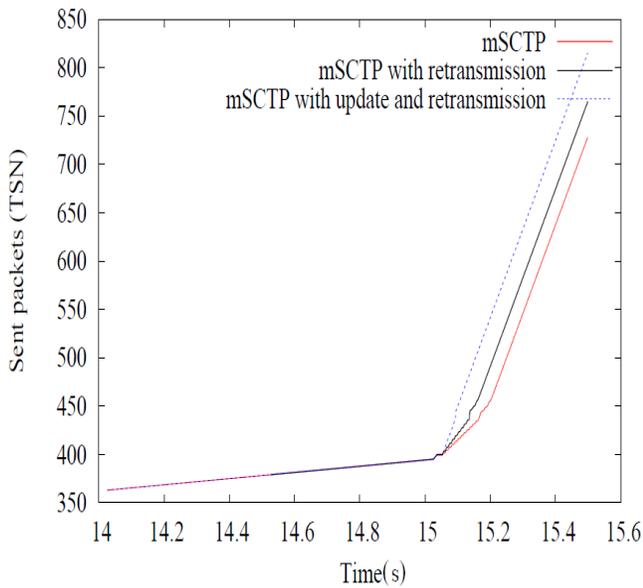


Figure 4. Handover from low to high bandwidth access network

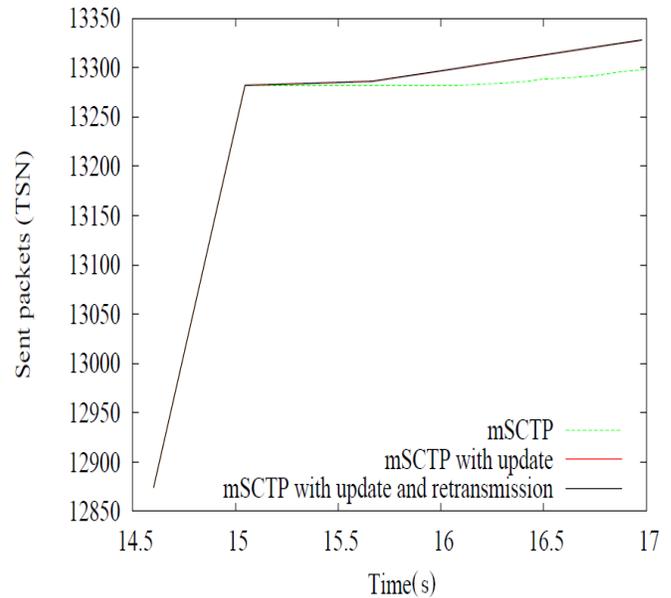


Figure 5. Handover from high to low bandwidth access network

in simulation, our proposition can be applicable for general interfaces with any technology type. In order to investigate our proposed scheme efficiencies, we execute two different scenarios. In the first one, endpoint executes handover from the 3G interface to the Wi-Fi interface, in which packet reordering problem happens. In the second one, handover is carried out, in inverse direction, from the Wi-Fi interface to the 3G interface. In the both scenarios, we have implemented bandwidth estimation mechanisms, which are presented above, to calculate the bandwidth ratio of the two networks  $\alpha$ .

**B. Simulation results**

1) *Scenario 1-Handover from low to high bandwidth access network (from 3G to Wi-Fi interface):* In this scenario, Sctp primary switching operation occurs at 15.1 seconds. Figure 4 illustrates three curves corresponding to the handover performances of three schemes: the red one for original mSCTP, the black one for mSCTP with buffer retransmission mechanism and the blue one for mSCTP with the combination of both congestion control parameter update and buffer retransmission. Cumulative TSN Ack value, which represents the last Transmission Sequence Number (TSN) of Data chunk successfully received in sequence, is used to measure the mSCTP handover performance.

As shown in Figure 4, mSCTP including our proposed schemes has the best handover performance as the accumulated TSN of the blue line increases most quickly. The handover performance of the original mSCTP is the worst among these three handover schemes, proved by the

fact that the accumulated TSN of the red line increases most slowly. The performance of mSCTP integrated two mechanisms is also better than the one of mSCTP only including buffer retransmission mechanism. Even though mSCTP with buffer retransmission can overcome the packet reordering problem, ongoing traffic has to still enter into slow-start phase in the new primary path after the primary path switching operation, causing performance degradation. The performance of mSCTP including buffer retransmission mechanism is better than the one of the original mSCTP because at least packet reordering problem is solved.

2) *Scenario 2- Handover from high to low bandwidth access network (from Wi-Fi to 3G interface):* In Figure 5, the handover performances of three schemes are presented by three curves: the green one for original mSCTP, the red one for mSCTP with congestion control parameter update and the black one for mSCTP with the combination of both mechanisms, congestion control parameter update and buffer retransmission. The handover occurs at 15.1 seconds. We also use the Cumulative TSN Ack value to represent the mSCTP handover performance. As shown in Figure 5, mSCTP integrating two proposed mechanisms has the best handover performance. The accumulated TSN of the black line increases most quickly. The original mSCTP gets the worst handover performance among these three handover schemes. The accumulated TSN of the green line increases most slowly. The red curve and the black one are almost the same. It means that, performance of mSCTP integrated two mechanisms is the same as the one of mSCTP with only update parameter congestion control mechanism. The fact is

that mSCTP does not suffer from packet reordering problem in this handover scenario (from high to low bandwidth access network). As described above, packet reordering problem happens when endpoint executes handover to a higher bandwidth (or smaller delay) network.

## V. CONCLUSION

In this paper, we have presented problems, which reduce mSCTP performance in handover procedure: slow-start phase and packet reordering problem. In order to solve these problems, we have proposed to combine congestion control parameter update and buffer retransmission mechanisms. The first one helps to bypass slow-start phase for a generic handover between two interfaces with mSCTP. Its principle is to adapt proportionally congestion control parameters considering the available bandwidth ratio of the two paths. The second one is used to solve packet reordering problem for handover in heterogeneous access networks environment. The simulation results show that the proposed solution has a better performance in comparison with the original mSCTP in terms of throughput and achievement of seamless handover.

In future works, we extend our simulation with the wireless interfaces such as 3G and Wi-Fi in order to verify our algorithm in a more realistic context. An implementation of our algorithm in a real test-bed with an Openmoko terminal [12] as well as with laptops equipped with two interfaces, Wifi and cellular (GPRS or 3G), is in progress and will be subject to future publications.

## ACKNOWLEDGMENT

Parts of this work are carried out in the framework of the 3MING (Mobilite Multitechnologie Multi-homING) project funded by Agence National de la Recherche (ANR-05-RNRT-07 TCOM 001) and the FP7 Integrated project 4WARD, which is partially funded by the European Union Commission.

## REFERENCES

- [1] R. Stewart, *Stream Control Transmission Protocol*. IETF RFC 4960, September 2007.
- [2] L. Ong and J. Yoakum, *An Introduction to the Stream Control Transmission Protocol (SCTP)*. IETF RFC 3286, May 2002.
- [3] M. Riegel, Q. Xie, and M. Tuexen, *Mobile SCTP*. draft-riegeltuexen-mobile-sctp-09.txt, Internet-Draft, November 2007.
- [4] R. Stewart, Q. Xie, M. Tuexen, S. Maruyama, and M. Kozuka, *Stream Control Transmission Protocol (SCTP) Dynamic Address Reconfiguration*. IETF RFC 5061, September 2007.
- [5] S.J. Koh, M.J. Chang, and M. Lee, *mSCTP for Soft Handover in Transport Layer*. IEEE communications letters, Vol. 8, No. 3, pp. 189-191, March 2004.
- [6] S.H. Kwon, T.W. Um, and W. Ryu, *Mobile SCTP with Bicast-ing for Vertical Handover*. Third International Conference on Convergence and Hybrid Information Technology (ICCIT), Busan, Korea, November 2008.
- [7] K.J. Lee, S. S. Nam, and B. I. Mun, *SCTP Efficient Flow Control During Handover*. Wireless Communications and Networking Conference (WCNC), Las Vegas, United States of America, April 2006.
- [8] D.P. Kim, S. J. Koh, and V. Leung, *On the Packet Reordering of mSCTP for Vertical Handover in Heterogeneous Wireless Networks*. Vehicular Technology Conference (VTC), Calgary, Canada, September 2008.
- [9] R. Fracchia, C. Casetti, C.-F. Chiasserini, and M. Meo, *WiSE: Best-path selection in wireless multihoming environments*. IEEE Transactions on Mobile Computing, Vol. 6, No. 10, pp. 1130-1141, October 2007.
- [10] K. Lai and M. Baker, *Measuring bandwidth*. Proceedings of the Conference on Computer Communications (IEEE Infocom), Vol.1, pp. 235-245, New York, United States of America, March 1999.
- [11] ns-2 network simulator (ver. 2), URL: <http://www.isi.edu/nsnam/ns> 20.04.2011.
- [12] Openmoko terminal, URL: [http://wiki.openmoko.org/wiki/Main\\_Page](http://wiki.openmoko.org/wiki/Main_Page) 20.04.2011.

# Sink-Connected Barrier Coverage Optimization for Wireless Sensor Networks

Yung-Liang Lai and Jehn-Ruey Jiang

National Central University  
Jhongli, 32001, Taiwan

945402011@cc.ncu.edu.tw jrjiang@csie.ncu.edu.tw

**Abstract**—This paper addresses the sink-connected barrier coverage optimization problem, which is concerned with how to select randomly deployed sensor nodes of a wireless sensor network (WSN) to reach two optimization goals: (1) to maximize the degree of barrier coverage by the minimum number of detecting nodes, and (2) to make the detecting nodes sink-connected by the minimum number of forwarding nodes. The detecting nodes are those for detecting intruders crossing the boundary of a monitored region. On detecting intruders, they send intruding event notifications to one of the sink nodes with the help of the forwarding nodes relaying. An algorithm, called optimal node selection algorithm (ONSA), is proposed for solving the problem on the basis of the maximum flow minimum cost algorithm. We perform simulations for ONSA and compare the results with those of a related algorithm, the global determination algorithm (GDA). The simulation results show that ONSA is better than GDA in terms of the number of nodes required in constructing sink-connected barrier coverage.

**Keywords**—Wireless sensor networks; Barrier coverage; Maximum flow minimum cost algorithm; Sink connectivity

## I. INTRODUCTION

A *wireless sensor network (WSN)* consists of a large number of sensor nodes with the capabilities of sensing, computing, storing, and communicating data. Each sensor can sense physical phenomena, such as light, temperature, sound, vibration, or electromagnetic field strength, and can transmit sensed data to one or more *sink nodes* through a multiple-hop transmission link. WSNs are self-organizing in the sense that they can be formed without human intervention, adapt to sensor failure and degradation, and react to task changes. They have wide applications like battlefield surveillance, environment monitoring, and so on. Some recent research uses WSNs to establish a *virtual barrier* of sensor nodes for detecting intruders crossing a protected area boundary, such as coastlines, national borders [9], and battlefield boundaries [11].

The barrier coverage problem is concerned with how to deploy WSN sensor nodes to form sensor barrier coverage for detecting intruders crossing a boundary. Several studies [2, 3, 4, 9, 10, 11] address the problem. They try to measure the quality of barrier coverage and/or to design schemes to achieve high-quality barrier-coverage in WSNs. In general, the quality of barrier coverage is measured by the *degree*. A

WSN is said to form *k-degree barrier coverage* (or *k-barrier coverage*, for short) if any intruder crossing the barrier is to be detected by at least  $k$  sensors. To take the WSN in Fig. 1 as an example, it forms 2-barrier coverage and its degree of barrier coverage is 2. This is because any intruder will be detected by at least two different sensor nodes when the intruder crosses the WSN from one border side to the opposite side.

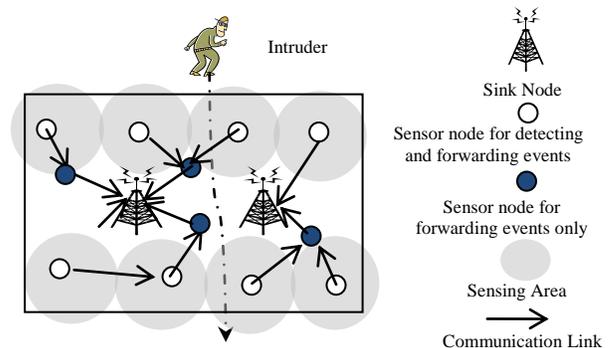


Figure 1. Sink-connected 2-barrier coverage

To the best of our knowledge, no earlier research addresses the barrier coverage problem with both the considerations that the sensor nodes should be connected to the sink node and that the number of the sensor nodes is minimized. In this paper, we take both considerations into account and propose an algorithm to solve the *sink-connected barrier coverage optimization problem*, which is concerned with how to select nodes from sensor nodes of a randomly deployed WSN to reach following two goals:

*Goal 1: Maximizing the degree of barrier coverage using the minimum number of detecting nodes*

*Goal 2: Minimizing the number of forwarding nodes to make detecting nodes sink-connected*

Randomly deployed nodes can be selected to be *detecting nodes* or *forwarding nodes*. The former is selected to be active for detecting intruders and sending *intruding event notifications* towards the sink nodes, and the latter, for forwarding the notifications. It is noted that unselected

nodes can remain inactive to save energy and detecting nodes can also help forward event notifications of other detecting nodes. The first goal is to maximize the degree of WSN barrier coverage while minimizing the number of the detecting nodes. The second goal is to make detecting nodes *sink-connected* (i.e., to make sure that every detecting node can find a path to send intruding event notification to a sink node) by adding a minimum number of forwarding nodes. When the number of forwarding nodes decreases, the collision probability goes down, and the energy consumption in transmissions is thus reduced.

The remainder of this paper is organized as follows. In Sections 2, we introduce some related work. In Section 3, we present the network model and problem definitions. The proposed algorithm is described in Section 4 and simulation results are reported in Section 5. Finally, conclusion is drawn in Section 6.

## II. RELATED WORK

The notion of barrier coverage was first introduced by Gage in [6] aiming at sensor-based surveillance of the boundary barrier to minimize the probability of undetected enemy penetration through the boundary barrier. In [10], Liu and Towsley defined *detectability* to be the probability that no path exists for an object to penetrate a barrier. They also characterize the detectability and showed that if the sensor node density is below a critical density, an intruder can almost surely find a path to cross the barrier without being detected. Wang and Cao in [12] also studied how to construct barrier coverage to monitor moving objects in camera sensor networks.

Kumar et al. [9] defined the notion of  $k$ -barrier coverage for precisely representing a WSN's ability of intruder detection. A WSN is said to have the  $k$ -barrier coverage property if any intruder crossing the barrier is detected successful by at least  $k$  sensor nodes. The authors developed theorems and proposed a centralized scheme using the maximum flow algorithm to determine whether a belt boundary region is  $k$ -barrier covered or not. Besides, they showed that the individual sensors cannot locally decide whether a network can form barrier coverage due to the lack of the global information. Unlike the algorithm in [9] that returns either true or false (0 or 1) for measuring the quality of barrier coverage, the method proposed by Chen et al. in [4] returns a non-binary value for the measurement. They also proposed a method to identify local regions whose qualities do not reach the desired level of quality.

Chen et al. [3] proposed a localized algorithm that guarantees the detection of intruders whose trajectory is confined to a slice of a belt boundary region. Saipulla et al. in [11] studied the barrier coverage of WSNs with line-based deployment, in which sensors are deployed along a line (e.g., sensors are dropped from an aircraft along a given path). Balister et al. [2] estimated the reliable node density

that achieves barrier coverage with  $s$ - $t$  connectivity in a thin strip with finite length, where  $s$ - $t$  connectivity means that a connected path exists between the two far ends of the thin strip.

## III. NETWORK MODEL AND PROBLEM FORMULATION

In this section, we first describe the network model and then formulate the sink-connected barrier coverage optimization problem to be solved in this paper.

### A. Network Model

Consider a WSN consisting of many sensor nodes and few sink nodes, in which sensor nodes are to form a virtual sensor barrier for monitoring a belt region to detect and send intruding events to one of the sink nodes. The sensor nodes are assumed to be randomly deployed; for example, they can be dropped from an aircraft as described in [11]. Each sensor node is equipped with a sensing module with a fixed sensing range to sense intruders and a communication module with a fixed communication range to communicate with other sensor nodes or sink nodes. Initially, a sensor node performs a bootstrapping task to pin point its location, discover its neighboring nodes, and report its information, such as the identification and the location, to one of the sink nodes. The sink nodes are more powerful than sensor nodes. They have more energy, memory, computing power and communication capacity. They can communicate with each other and with sensor nodes; they can also communicate with the backend system, which is assumed to have unlimited power supply and enormous computing power to gather all WSN nodes' information and perform the optimization computation.

Let  $V_s$  and  $V_k$  denote the set of sensor nodes and the set of sink nodes, respectively. Below, we define a *coverage graph*  $G_c$  to represent the sensing area coverage relationships of nodes. Moreover, we define a *transmission graph*  $G_t$  to represent the nodes' wireless transmission reachability relationships.

#### (1) Coverage Graph

A coverage graph  $G_c(V_s \cup \{S, T\}, E_c)$  is an undirected graph, in which  $V_s$  is the sensor node set,  $E_c$  is the edge set, and  $\{S, T\}$  are two virtual nodes. The edge set  $E_c$  represents the sensing area coverage overlap relationships. For two nodes  $N_i$  and  $N_j$  in  $V_s$ , there exists an edge  $(N_i, N_j)$  in  $E_c$  if  $N_i$ 's coverage and  $N_j$ 's coverage have overlap. As shown in Fig. 2, the monitored belt region has the outer side, inner side and lateral sides. Intruders are supposed to cross the belt region from outer side to inner side. The virtual nodes  $S$  and  $T$  are associated the lateral sides; an edge  $(N_i, S)$  or  $(N_i,$

T) exists in  $E_c$  if  $N_i$ 's sensing area overlap either lateral side. Fig. 2 shows the coverage graph  $G_c$  of the WSN with 8 sensor nodes  $N_1, \dots, N_8$ , which are represented by solid circles. Note that the gray shades around the solid circles represent the sensing areas of sensor nodes.

Now, we can define the traversable paths in  $G_c$ . A traversal path of a coverage graph  $G_c(V_s \cup \{S, T\}, E_c)$  is defined to be a path starting from  $S$ , going along edges in  $E_c$  through nodes in  $V_s$ , and stopping at  $T$ . Note that a coverage graph is similar to a flow network [1] and a traversable path is similar to a *flow* in the network. In the flowing context, the terms "traversable path" and "flow" will be used alternatively. The coverage graph and its traversal paths are very useful for measuring the degree of barrier coverage. By the theorems developed in [9], a WSN forms  $k$ -barrier coverage if and only if there exist  $k$  node-disjoint traversable paths in the coverage graph of the WSN. In the WSN of Fig. 2, there are two node-disjoint traversable paths  $S-N_1-N_2-N_3-N_4-T$  and  $S-N_5-N_6-N_7-N_8-T$  in the WSN coverage graph, so the WSN forms 2-barrier coverage.

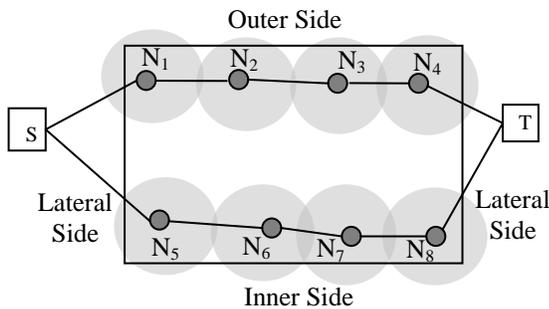


Figure 2. A WSN coverage graph and its 2 node-disjoint traversable paths

(2) Transmission Graph

A transmission graph  $G_t(V_s \cup V_k, E_t)$  is a directed graph, where  $V_s$  is the sensor node set,  $V_k$  is the sink node set, and  $E_t$  is an arc (or directed edge) set to represent transmission relationships. Note that we may use "edge" to stand for "arc" in the following context. That is, the two terms are used interchangeably when there is no ambiguity. For two nodes  $N_i$  and  $N_j$  in  $V_s$ , it exists an arc  $\langle N_i, N_j \rangle$  in  $E_t$  if the node  $N_i$  can successfully transmit data (or events) to node  $N_j$  over a direct wireless link. Based on the transmission graph  $G_t$  of a WSN, we can define the sink-connected property for a set of sensor nodes as follows. For the WSN with the transmission graph  $G_t(V_s \cup V_k, E_t)$ , a set  $S$  ( $S \subseteq V_s$ ) of sensor nodes is *sink-connected* if there exists a path for each node in  $S$  going through only nodes in  $S$  to reach a node in  $V_k$ . For example, for the WSN in Fig. 3 consisting of 14 sensor nodes  $N_1, \dots, N_{14}$  and 2 sink nodes  $K_1$  and  $K_2$ , the node sets  $\{N_4\}$ ,  $\{N_{11}\}$ ,  $\{N_1, N_9\}$ ,  $\{N_2, N_3, N_{11}\}$ ,  $\{N_4, N_7, N_8, N_{13}\}$ ,  $\{N_9, \dots, N_{13}\}$  and  $\{N_1, \dots, N_{13}\}$  all satisfy the sink-connected

property. However, the node sets  $\{N_1\}$ ,  $\{N_6, N_{11}\}$  and  $\{N_1, \dots, N_8\}$  do not satisfy the sink-connected property.

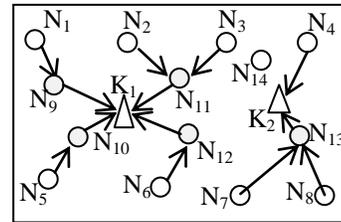


Figure 3. A WSN transmission graph with partial arcs for illustrating the sink-connected property

B. Sink-Connected Barrier Coverage Optimization Problem

The objective of the sink-connected barrier coverage optimization problem is to maximize the degree of barrier coverage by selecting the minimum number of nodes, while keeping the selected nodes sink-connected. Below, we formally define the problem.

Given a WSN with the coverage graph  $G_c(V_s \cup \{S, T\}, E_c)$  and transmission graph  $G_t(V_s \cup V_k, E_t)$ , the sink-connected barrier coverage optimization problem is to achieve the following two goals.

**Objective 1:** To find a minimum sensor node set  $V_r$  such that the number of node-disjoint traversable paths of  $V_r$  is maximized

**Objective 2:** To find a minimum forwarding node set  $V_t$  such that  $(V_r \cap V_t = \emptyset)$  and  $(V_r \cup V_t)$  satisfies the sink-connected property.

According to the above definition, a solution to the sink-connected barrier coverage optimization problem will return two node sets  $V_r$  and  $V_t$ . We can assume the nodes in  $V_r$  as detecting nodes to detect intruding events, and assume the nodes in  $V_t$  as forwarding nodes to forward events to one of the sink node. Certainly, since the detecting nodes remain active, they can also forward the intruding events sent by other detecting nodes. In reality, the detecting nodes can form barrier coverage with the highest degree. The solution is optimal in the sense that the degree of barrier coverage is maximized, while the number of detecting nodes and the number of forwarding nodes are both minimized. The solution is also practical in the sense that the detecting nodes are sink-connected with the help of forwarding nodes.

#### IV. OPTIMAL NODE SELECTION ALGORITHM (ONSA)

In this section, we propose an algorithm, called *optimal node selection algorithm (ONSA)*, to solve the sink-connected barrier coverage optimization problem. Given the sensor nodes  $V_s$ , sink nodes  $V_k$ , coverage relationship  $E_c$ , and transmission relationship  $E_t$ , ONSA can find the detecting node set  $V_r$  and the forwarding node set  $V_t$ . ONSA has three main tasks. The first task is to construct the coverage graph  $G_c$  and then perform the node-disjoint transformation to generate the graph  $G_c^*$  such that  $G_c^*$  is a flow network [5]. The second task is to find the minimum cost maximum flow in  $G_c^*$ . The third task is to construct the transmission graph  $G_t$  based on  $G_c^*$  and to find a flow plan by executing the maximum flow minimum cost algorithm. The nodes selected in the flow plan will be activated for constructing sink-connected barrier coverage. The details of ONSA are described below.

---

##### Optimal Node Selection Algorithm (ONSA)

---

Input:  $V_s, V_k, E_c, E_t$

Output:  $V_r$  and  $V_t$

Step 1: Construct a coverage graph  $G_c(V_s \cup \{S, T\}, E_c)$ , where  $S$  and  $T$  are virtual nodes. Each edge in  $E_c$  is associated with one capacity and zero cost.

Step 2: Execute *node-disjoint transformation* to transfer  $G_c$  into the new graph  $G_c^*$ .

Step 3: Execute the maximum flow minimum cost algorithm on  $G_c^*$  to decide the *minimum cost flow plan* ( $FP_{MinCt}$ ), and let node set  $V_r$  contain the selected nodes in  $FP_{MinCt}$ .

Step 4: Construct a transmission graph  $G_t(V_s \cup V_k, E_t)$ . Add a virtual source node  $S$  and a virtual target node  $T$  into  $G_t$ .

Step 5: For each node in  $V_r$  on graph  $G_t$ , add an edge from the virtual source node  $S$  to it. For each sink node, add an edge from it to the virtual target node.

Step 6: Execute *node-edge transformation* to convert  $G_t$  into  $G_t^*$ .

Step 7: Execute the maximum flow minimum cost algorithm to find the *minimum cost flow* ( $FP_T$ ) on  $G_t^*$ . Let  $V_b$  be the set of the nodes selected in  $FP_T$ .

Step 8: Set  $V_t = V_b - V_r$  and return  $V_r$  and  $V_t$

---

In step 1, ONSA constructs a coverage graph  $G_c$  with a virtual node  $S$  and a virtual node  $T$ . In step 2, ONSA executes the *node-disjoint transformation* to convert each

node with multiple inbound flows and multiple outbound flows into another form. The purpose of the transformation is to make the generated graphs node-disjoint. In the transformation, a node  $X$  will be transformed into a pair of nodes  $X'$  and  $X''$  with one capacity and one unity cost (refer to Fig. 4 (a)).

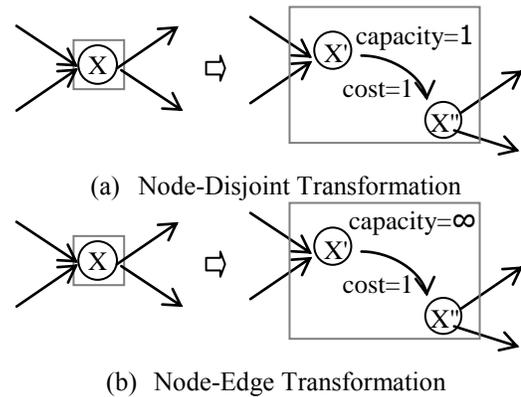


Figure 4. Two transformations of ONSA

In step 3, ONSA executes the well-known maximum flow minimum cost algorithm, which has two procedures. The first procedure is to find a maximum flow by executing the maximum flow algorithm. The second procedure is to execute the minimum cost flow algorithm. The readers are referred to [5] and [8] for the procedure details. After executing the two procedures on  $G_c^*$ , a set  $V_r$  of detecting nodes is obtained. Since the flow is maximized, the number of node-disjoint traversable paths is also maximized. Moreover, since the cost is minimized, the number of nodes in  $V_r$  is also minimized.

In step 4, ONSA constructs a transmission graph  $G_t(V_s \cup V_k, E_t)$  and adds a virtual source node  $S$  and a virtual target node  $T$  into  $G_t$ . In step 5, ONSA adds arcs for connecting the virtual source node  $S$  and the nodes in  $V_r$ . Each newly added arc is associated with  $Cost=0$  and  $Capacity=\infty$ .

In step 6, ONSA converts each node (excluding  $S$  and  $T$ ) into two nodes with one arc of  $Cost=1$  and  $Capacity=\infty$  by the *node-edge transformation*. Please refer to the Fig. 4(b).

In step 7, ONSA executes the maximum flow minimum cost algorithm on  $G_t^*$  to decide the minimum cost flow ( $FP_T$ ). In this step, the nodes containing in  $FP_T$  will be added into the node set  $V_b$ . Since  $FP_T$  has the minimum cost, the number of nodes in  $V_b$  will also be minimized.

In step 8, ONSA returns  $V_r$  as the set of detecting nodes and returns  $V_t = V_b - V_r$  as the set of forwarding nodes.

Below, we take the WSN in Fig. 3 as an example to illustrate the execution of ONSA. In step 1, a coverage graph  $G_c$  will be constructed. After step 2, the nodes with multiple inbound flows and multiple outbound flows are transformed by the node-disjoint transformation. The transformation results are shown in Fig. 5.

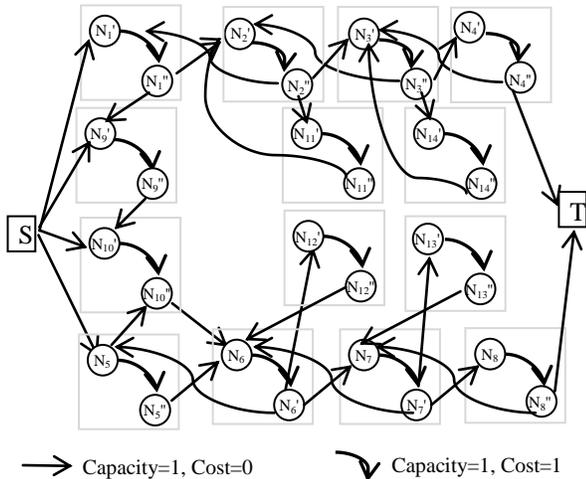


Figure 5. An example of the node-disjoint transformation

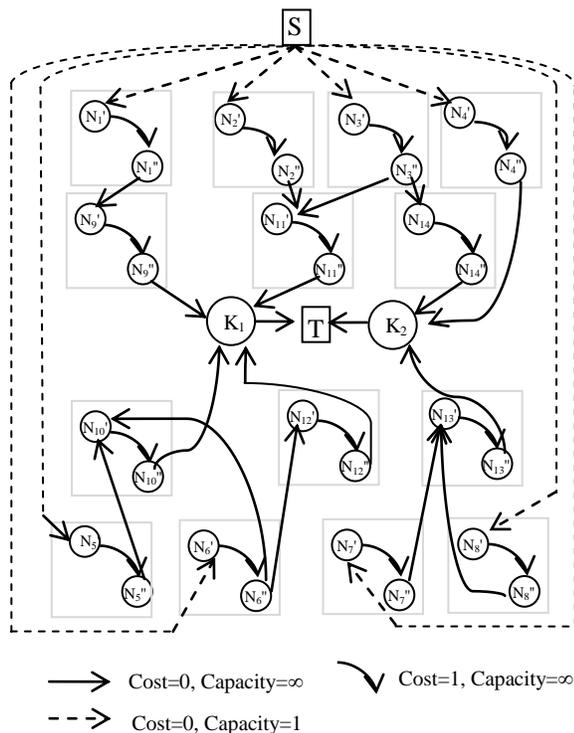


Figure 6. An example of the node-edge transformation

In step 3, a flow plan is decided by the maximum flow minimum cost algorithm. In this example, the maximum number of flows is two. In step 4, the graph transmission graph  $G_t$  is constructed and virtual nodes  $S$  and  $T$  are added into  $G_t$ . In step 5, a new arc is added between the virtual source node  $S$  and every node selected in step 3 (i.e., every node in  $V_r$ ). Moreover, a new arc is added between every sink node and the virtual target node.

In step 6, node-edge transformation is performed to generate  $G_{t^*}$ , as shown in Fig. 6. In step 7, the maximum flow minimum cost algorithm is executed to obtain  $V_b$ . In this example,  $V_b$  is  $\{N_1, N_2, \dots, N_{13}\}$ , which is a set containing the nodes selected in  $FP_T$ . In step 8,  $V_r = \{N_1, \dots, N_8\}$  and  $V_t = \{N_9, \dots, N_{13}\}$  are returned by ONSA, where  $V_r$  contains detecting nodes to form 2-barrier coverage for detecting intruding events and  $V_t$  contains forwarding nodes to forward events sent by detecting nodes to one of the sink nodes (i.e., either  $K_1$  or  $K_2$ ). Fig. 7 shows the execution results returned by ONSA.

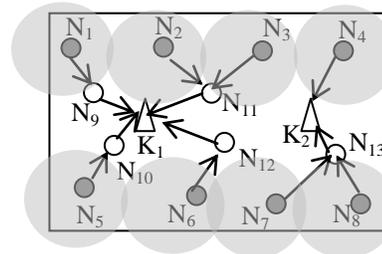


Figure 7. The execution result of ONSA, where  $N_1, \dots, N_8$  are selected as detecting nodes and  $N_9, \dots, N_{13}$  are selected as forwarding nodes

The time complexity of ONSA is dominated by Step 3 and Step 7, which execute the maximum flow minimum cost algorithm on  $G_{c^*}$  and  $G_{t^*}$ , respectively. The maximum flow minimum cost algorithm is actually the combination of the Edmonds-Karp algorithm [5], which is of  $O(V \cdot E^2)$  time complexity for a graph of vertex set  $V$  and edge set  $E$ , and the minimum cost flow algorithm (MinCostFlow) [8], which is of  $O(V \cdot E^2 \cdot \log(V))$  time complexity. The time complexity of ONSA is thus  $O(V_{c^*} \cdot E_{c^*}^2 \cdot \log(V_{c^*}) + V_{t^*} \cdot E_{t^*}^2 \cdot \log(V_{t^*}))$ , where  $V_{c^*}$  (resp.,  $V_{t^*}$ ) is the size of the vertex set in  $G_{c^*}$  (resp.,  $G_{t^*}$ ) and  $E_{c^*}$  (resp.,  $E_{t^*}$ ) is the size of the edge set in  $G_{c^*}$  (resp.,  $G_{t^*}$ ). To execute the optimization computation of ONSA will consume some computation power and memory storage. Fortunately, as we have mentioned earlier, ONSA is performed by the backend system, which is assumed to have unlimited power supply and enormous computing power. All the sensor nodes in the WSN only need to collaborate to deliver/forward their local information required by ONSA to the sink nodes, which in turn forward the information to the backend system. In other words, ONSA does not impose much computation and memory consumption on normal sensor nodes.

## V. SIMULATION

To demonstrate the advantages of ONSA, our proposed algorithm, we conduct simulation experiments. We also compare the simulation results with those of the *global determination algorithm (GDA)*, which is proposed in [9], for determining the degree of barrier coverage by using the maximum flow algorithm. Since GDA does not consider the

sink-connected property, we again use the maximum flow algorithm for GDA to select extra sensor nodes to serve as forwarding nodes to make GDA satisfy the property. In this way, GDA and ONSA can both achieve the highest-degree property and sink-connectivity property of barrier coverage.

We develop a simulator based on the MATLAB software [13] to solve the optimization problem. The simulations are conducted in the following settings. The sensing area coverage radius is 10m. The wireless transmission radius is equal to the coverage radius. All the sensors are randomly deployed by different number of sensors in a rectangle-shaped area of 100m x 10m. The number of sink nodes is 2 and their locations are respectively at  $(\frac{100}{3}\text{m}, 5\text{m})$  and  $(\frac{2 \times 100}{3}\text{m}, 5\text{m})$  relative to the left-most and lowest position of the rectangle.

We compare ONSA and GDA in terms of the number of nodes selected to achieve the highest-degreed and sink-connected barrier coverage. As shown in Fig. 8, ONSA selects fewer nodes than GDA for all cases. It implies that ONSA needs fewer nodes than GDA to achieve the highest-degreed and sink-connected barrier coverage.

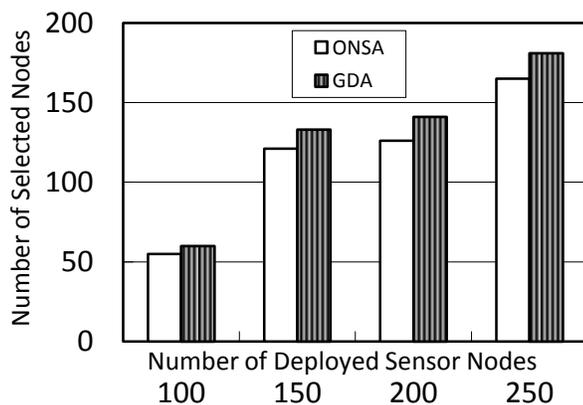


Figure 8. Comparisons of ONSA and GDA in terms of the number of selected nodes

## VI. CONCLUSION

In this paper, we studied the sink-connected barrier coverage problem to achieve two goals: (1) to maximize the degree of barrier coverage using the minimum number of detecting nodes and (2) to minimize the number of forwarding nodes to hold the sink-connected property. To the best of our knowledge, this is the first paper to consider the sink-connected property and the barrier coverage quality optimization at the same time. An optimal network flow planning algorithm, called optimal node selection algorithm (ONSA), is proposed to solve the problem. ONSA is based on the well-known maximum flow minimum cost algorithm. We also perform simulation experiments for ONSA and a related algorithm called global determination algorithm

(GDA), which uses the maximum flow algorithm to find out the maximum degree of barrier coverage and does not consider the sink-connected property. For the sake of comparison, the maximum flow algorithm is again used to make GDA satisfy the sink-connected property. The simulation results show that ONSA is better than GDA in terms of number of selected nodes.

In the future, we plan to study the optimization of barrier coverage with sink-node connectivity under the lifetime constraint. The lifetime is usually defined as the time span from the time of network deployment to the time when a certain fraction of sensor nodes run out of their energy. If the lifetime of a wireless sensor network is too short, it is likely that the network will soon be partitioned and fail to deliver the sensed data to sink nodes. The lifetime constraint is thus an important factor to be addressed.

## REFERENCES

- [1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall, Upper Saddle River, New Jersey, 1993.
- [2] P. Balister, B. Bollobas, A. Sarkar, and S. Kumar, "Reliable density estimates for coverage and connectivity in thin strips of finite length," *Proceedings of ACM Mobicom*, pp. 75-86, 2007.
- [3] A. Chen, S. Kumar, and T.-H. Lai, "Designing localized algorithms for barrier coverage," *Proceedings of ACM Mobicom*, pp. 63-74, 2007.
- [4] A. Chen, T.-H. Lai, and D. Xuan, "Measuring and guaranteeing quality of barrier-coverage in wireless sensor networks," *Proceedings of The ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pp. 421-430, 2008.
- [5] T.-H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms (Second Edition)*, MIT Press and McGraw-Hill, 2001.
- [6] D. Gage. "Command control for many-robot systems," *Proceedings of the Nineteenth Annual AUVS Technical Symposium (AUVS-92)*, pp. 28-34, 1992.
- [7] M. R. Garey and D. S. Johnson, *Computer and Intractability*, W. H. Freeman and Co., 1979.
- [8] A. V. Goldberg, "An efficient implementation of a scaling minimum-cost flow algorithm," *Journal of Algorithm*, vol. 22, pp. 1-29, 1997.
- [9] S. Kumar, T.-H. Lai, and A. Arora, "Barrier coverage with wireless sensors," *Wireless Networks*, vol. 13, pp. 817-834, 2007.
- [10] B. Liu and D. Towsley, "A study on the coverage of large-scale sensor networks," in *The 1st IEEE International Conference on Mobile Ad-hoc and Sensor Systems*, pp. 475-483, 2004.
- [11] A. Saipulla, C. Westphal, B. Liu and J. Wang, "Barrier Coverage of Line-Based Deployed Wireless Sensor Networks," *IEEE INFOCOM 2009*, pp. 127-135, Brazil, 2009.
- [12] Y. Wang, G. Cao, "Barrier Coverage in Camera Sensor Networks," to be presented at *ACM MobiHoc'11*, Paris, France, 2011.
- [13] MathWorks Inc, Matlab Software, URL: <http://www.mathworks.com>. (Accessed at January 8, 2011).

## The Joint Probability Density Function of the SSC Combiner Output Signal at Two Time Instants in the Presence of Hoyt Fading

Dragana Krstić

Department of Telecommunications,  
Faculty of Electronic Engineering,  
University of Niš  
Aleksandra Medvedeva 14  
Niš, Serbia  
dragana.krstic@elfak.ni.ac.rs

Petar Nikolić

Tigartyres, Pirot, Serbia  
nikpetar@gmail.com

Goran Stamenović

Tigar, Pirot, Serbia  
goran.stamenovic@tigar.com

Mihajlo Stefanović

Faculty of Electronic Engineering  
University of Niš  
Aleksandra Medvedeva 14  
Niš, Serbia  
mihajlo.stefanovic@elfak.ni.ac.rs

**Abstract** — In this paper, the probability density function of the Switch and Stay Combiner (SSC) output signal to noise ratio (SNR) at one time instant and the joint probability density function of the SSC combiner output signal to noise ratio at two time instants, in the presence of Hoyt fading, are determined in the closed form expressions. The results are shown graphically in several figures and the evaluation of the various parameters influence, such as distribution parameters and decision threshold values, is given.

**Keywords** - Probability Density Function; Joint Probability Density Function; Hoyt Fading; Diversity Reception; SSC Combining.

### I. INTRODUCTION

The radio wave propagation through wireless communications channels has received a great deal of research interest [1], [2]. The rapid and random fluctuations of the signal envelope and phase in a radio channel are caused with two propagation phenomena: multipath scattering (fast fading) and shadowing (slow fading). In wireless communications the multipath fading is modeled by several distributions such as: Rayleigh, Rice, Nakagami- $m$ , Weibull and so on.

Another distribution, which has recently received increased attention in modeling fading channels, is the Hoyt (Nakagami- $q$ ) distribution. The Hoyt fading model provides a very accurate fit to experimental channel measurements in a various communication applications, like mobile satellite propagation channels [3], and spans the range of the fading figure from the one-sided Gaussian to the Rayleigh distribution [4]. Similarly, the Hoyt distribution can be considered as an accurate fading model for satellite links with strong ionospheric scintillation [5]. Recently, in [6], an ergodic capacity analysis is presented, and in [7] the

information outage probability of OSTBC over Hoyt fading channels has been studied. Also in [8] this model has been used in outage analysis of cellular mobile radio systems, while in [9] a capacity analysis of Hoyt fading is provided.

In wireless communication systems, various techniques for reducing fading effect and influence of shadow effect are used. Such techniques are diversity reception, dynamic channel allocation and power control. Upgrading transmission reliability and increasing channel capacity without increasing transmission power and bandwidth is the main goal of diversity techniques.

Diversity reception, based on using multiple antennas at the receiver, space diversity, with two or more branches, is a very efficient method used for improving system's quality of service, so it provides efficient solution for reduction of signal level fluctuations in fading channels. Multiple received copies of signal could be combined on various ways. Among the most popular diversity techniques are: maximal ratio combining (MRC), equal gain combining (EGC), and generalized selection combining (GSC) [1], but their complexity of implementation is relatively high since they require a dedicated communication receiver for each diversity branch. On the other hand, among the simpler diversity combining schemes, the two most popular are selection combining (SC) and switch and stay combining (SSC). Selection combining (SC) and switch and stay combining (SSC) types of diversity systems process only one of the diversity branches, so they are less complicated.

Switch and stay combining (SSC) is an attempt at simplifying the complexity of the system but with loss in performance. In this case, the receiver selects a particular antenna until its quality drops below a predetermined threshold. When this happens, the receiver switches to another antenna and stays with it for the next time slot, regardless of whether or not the channel quality of that antenna is above or below the predetermined threshold. The consideration of SSC systems in the literature has been

restricted to low-complexity mobile units where the number of diversity antennas is typically limited to two [10-11]. Furthermore, in all these publications, only predetection SSC has thus far been considered wherein the switching of the receiver between the two receiving antennas is based on a comparison of the instantaneous SNR of the connected antenna with a predetermined threshold. This results in a reduction in complexity relative to SC in that the simultaneous and continuous monitoring of both branches SNRs is no longer necessary.

The probability density function (PDF) of the SSC combiner output signal at one time instant and the joint probability density function of the SSC combiner output signal at two time instants in the presence of Rayleigh, Nakagami-m, Weibull and log-normal fading are determined in [12-15], respectively.

In this paper the probability density function of the SSC combiner output signal to noise ratio at one time instant and the joint probability density function of the SSC combiner output signal to noise ratio at two time instants in the presence of Hoyt fading will be determined. The joint probability density function of the SSC combiner output signal to noise ratio at two time instants is important when the decision is based on multiple samples.

The remainder of the document is organized in the following way: Section II introduces the model of the SSC combiner is given and the probability density function of the SSC combiner output signal to noise ratio at one time instant is determined. Subsequently, in Section III, the joint probability density function of the SSC combiner output signal to noise ratio at two time instants is calculated. In fourth section the numerical results are presented.

## II. SYSTEM MODEL

The use of SSC combiner with great number of branches can minimize the bit error rate (BER) [16]. We determine SSC combiner with two inputs because the gain is the greatest when we use the SSC combiner with two inputs instead of one-channel system. When we enlarge the number of branches the improvement becomes less [16]. The ratio price/complexity is the best for a system with two branches. Because of that it is more economic using SSC combiner with two inputs.

The model of this system is shown in Fig. 1. The signal to noise ratios at the combiner inputs are  $\gamma_1$  and  $\gamma_2$ , and  $\gamma$  is the combiner output signal to noise ratio.

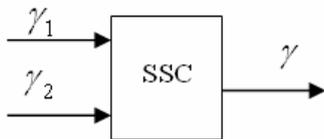


Figure 1. Model of the SSC combiner with two inputs

Let see how the SSC combiner with two inputs works. The probability of the event that the combiner first examines the signal at the first input is  $P_1$ , and for the second input is

$P_2$ . If the combiner examines first the signal at the first input and if the value of the signal to noise ratio at the first input is above the threshold,  $\gamma_T$ , SSC combiner forwards this signal to the circuit for the decision. If the value of the signal to noise ratio at the first input is below the threshold  $\gamma_T$ , SSC combiner forwards the signal from the other input to the circuit for the decision, regardless if it is above or below the predetermined threshold. If the SSC combiner first examines the signal from the second combiner input it works in the similar way.

The expression for the probability density of the combiner output signal to noise ratio will be determined first for the case  $\gamma < \gamma_T$ . Based on the work algorithm of the SSC combiner in this case, the probability density is equal, for  $\gamma < \gamma_T$ :

$$p_\gamma(\gamma) = P_1 \cdot F_{\gamma_1}(\gamma_T) \cdot p_{\gamma_2}(\gamma) + P_2 \cdot F_{\gamma_2}(\gamma_T) \cdot p_{\gamma_1}(\gamma) \quad (1)$$

In the case  $\gamma \geq \gamma_T$  the expression for the probability density of the signal to noise ratio at the combiner output is :

$$p_\gamma(\gamma) = P_1 \cdot p_{\gamma_1}(\gamma) + P_1 \cdot F_{\gamma_1}(\gamma_T) \cdot p_{\gamma_2}(\gamma) + P_2 \cdot p_{\gamma_2}(\gamma) + P_2 \cdot F_{\gamma_2}(\gamma_T) \cdot p_{\gamma_1}(\gamma) \quad (2)$$

where  $\gamma_T$  is the threshold of the decision, and the cumulative probability densities (CDFs) are given by [17]:

$$F_{\gamma_i}(\gamma_T) = \int_0^{\gamma_T} p_{\gamma_i}(x) dx, \quad i=1,2 \quad (3)$$

The probabilities  $P_1$  and  $P_2$  are [17]:

$$P_1 = \frac{F_{\gamma_2}(\gamma_T)}{F_{\gamma_1}(\gamma_T) + F_{\gamma_2}(\gamma_T)} \quad (4)$$

$$P_2 = \frac{F_{\gamma_1}(\gamma_T)}{F_{\gamma_1}(\gamma_T) + F_{\gamma_2}(\gamma_T)} \quad (5)$$

The probability densities of the SNRs at the combiner input,  $\gamma_1$  and  $\gamma_2$ , in the presence of Hoyt fading, are [17]:

$$p_{\gamma_1}(\gamma_1) = \frac{(1 + q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1 + q_1^2)^2 \gamma_1}{4q_1^2 \bar{\gamma}_1}\right) I_0\left(\frac{(1 - q_1^4) \gamma_1}{4q_1^2 \bar{\gamma}_1}\right) \quad \gamma_1 \geq 0 \quad (6)$$

$$p_{\gamma_2}(\gamma_2) = \frac{(1 + q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1 + q_2^2)^2 \gamma_2}{4q_2^2 \bar{\gamma}_2}\right) I_0\left(\frac{(1 - q_2^4) \gamma_2}{4q_2^2 \bar{\gamma}_2}\right) \quad \gamma_2 \geq 0 \quad (7)$$

where  $q_i$  are Nakagami-q fading parameters, which range from 0 to 1 and  $\bar{\gamma}_i$  are average SNRs for input channels

After putting of the expressions (4)-(7), (10) and (11) into (1), the probability density of the signal to noise ratio at the combiner output  $\gamma$ , is, for  $\gamma < \gamma_T$ :

$$\begin{aligned}
 p_\gamma(\gamma) &= P_1 \frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) \\
 &= \frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) \quad (10) \\
 &\cdot \frac{(1+q_2^2)}{2q_2 \bar{\gamma}_2} \exp \left( -\frac{(1+q_2^2)^2 \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) I_0 \left( \frac{(1-q_2^4) \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) + \\
 &+ P_2 \frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right) \\
 &\cdot \frac{(1+q_1^2)}{2q_1 \bar{\gamma}_1} \exp \left( -\frac{(1+q_1^2)^2 \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) I_0 \left( \frac{(1-q_1^4) \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) \quad (8)
 \end{aligned}$$

After putting of the expressions (4)-(7), (10) and (11) into (2), the probability density of the signal to noise ratio at the combiner output  $\gamma$ , is, for  $\gamma \geq \gamma_T$ :

$$\begin{aligned}
 p_\gamma(\gamma) &= P_1 \frac{(1+q_1^2)}{2q_1 \bar{\gamma}_1} \exp \left( -\frac{(1+q_1^2)^2 \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) I_0 \left( \frac{(1-q_1^4) \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) + \\
 &+ P_1 \frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) \\
 &\cdot \frac{(1+q_2^2)}{2q_2 \bar{\gamma}_2} \exp \left( -\frac{(1+q_2^2)^2 \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) I_0 \left( \frac{(1-q_2^4) \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) + \\
 &+ P_2 \frac{(1+q_2^2)}{2q_2 \bar{\gamma}_2} \exp \left( -\frac{(1+q_2^2)^2 \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) I_0 \left( \frac{(1-q_2^4) \gamma_2}{4q_2^2 \bar{\gamma}_2} \right) + \\
 &+ P_2 \frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right) \\
 &\cdot \frac{(1+q_1^2)}{2q_1 \bar{\gamma}_1} \exp \left( -\frac{(1+q_1^2)^2 \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) I_0 \left( \frac{(1-q_1^4) \gamma_1}{4q_1^2 \bar{\gamma}_1} \right) \quad (9)
 \end{aligned}$$

The cumulative probability densities (CDFs) of the SNRs at the combiner input in the presence of Hoyt fading, after putting of the expressions (6), (7), into (3), are given by:

$$F_{\gamma_1}(\gamma_T) = \int_0^{\gamma_T} \frac{(1+q_1^2)}{2q_1 \bar{\gamma}_1} \exp \left( -\frac{(1+q_1^2)^2 x}{4q_1^2 \bar{\gamma}_1} \right) I_0 \left( \frac{(1-q_1^4) x}{4q_1^2 \bar{\gamma}_1} \right) dx =$$

$$\begin{aligned}
 &= \frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) \quad (10) \\
 F_{\gamma_2}(\gamma_T) &= \int_0^{\gamma_T} \frac{(1+q_2^2)}{2q_2 \bar{\gamma}_2} \exp \left( -\frac{(1+q_2^2)^2 x}{4q_2^2 \bar{\gamma}_2} \right) I_0 \left( \frac{(1-q_2^4) x}{4q_2^2 \bar{\gamma}_2} \right) dx = \\
 &= \frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right) \quad (11)
 \end{aligned}$$

where  $I_e(k, x)$  is Rice's  $I_e$  function [18].

After putting of the expressions (10) and (11) into (4) and (5), the probabilities  $P_1$  and  $P_2$  are:

$$\begin{aligned}
 P_1 &= \frac{\frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right)}{\frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) + \frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right)} \quad (12) \\
 P_2 &= \frac{\frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right)}{\frac{2q_1}{1+q_1^2} I_e \left( \frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2 \gamma_T}{4q_1^2 \bar{\gamma}_1} \right) + \frac{2q_2}{1+q_2^2} I_e \left( \frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2 \gamma_T}{4q_2^2 \bar{\gamma}_2} \right)} \quad (13)
 \end{aligned}$$

The obtained expressions for the probability density function (PDF) of the output signal to noise ratio after diversity combining can be used to study the moments, the amount of fading, the outage probability and the average bit error rate of proposed system.

### III. SYSTEM PERFORMANCES AT TWO TIME INSTANTS

The model of the SSC combiner with two inputs at two time instants considering in this section is shown in Fig. 2. The signal to noise ratios at the inputs are  $\gamma_{11}$  and  $\gamma_{21}$  at the first time moment and they are  $\gamma_{12}$  and  $\gamma_{22}$  at the second time moment.

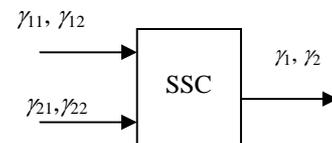


Figure 2. Model of the SSC combiner with two inputs at two time instants

The output signal to noise ratios are  $\gamma_1$  and  $\gamma_2$ . The indexes for the input signal to noise ratios are: first index is the number of the branch and the other signs time instant observed. For the output signal to noise ratios, the index represents the time instant observed.

The joint probability density function of uncorrelated signals at the input, with Hoyt distribution and same parameters, is [17]:

$$p_{\gamma_1\gamma_2}(\gamma_1, \gamma_2) = \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \quad (14)$$

Modified Bessel function of the first kind is defined by [19]:

$$I_m(x) = \left(\frac{x}{2}\right)^m \sum_{k=0}^{\infty} \frac{\left(\frac{x^2}{4}\right)^k}{k! \Gamma(m+k+1)} \quad (15)$$

Now we have four different cases. The first case is:  $\gamma_1 < \gamma_T$  and  $\gamma_2 < \gamma_T$ . In this case all signal to noise ratios at the input are below  $\gamma_T$ , i.e.:  $\gamma_{11} < \gamma_T$ ,  $\gamma_{12} < \gamma_T$ ,  $\gamma_{21} < \gamma_T$ , and  $\gamma_{22} < \gamma_T$ .

Let the combiner first examines the signal  $r_{11}$ . Because  $\gamma_{11} < \gamma_T$ , it follows that  $\gamma_1 = \gamma_{21}$ , and since  $\gamma_{22} < \gamma_T$  it is  $\gamma_2 = \gamma_{12}$ . The probability of this event is  $P_1$ .

When SSC combiner first examines the signal  $r_{21}$ , then  $\gamma_1 = \gamma_{11}$ , because  $\gamma_{21} < \gamma_T$ . Since  $\gamma_{12} < \gamma_T$ , then it is  $\gamma_2 = \gamma_{22}$ . The probability of this event is  $P_2$ . After previous, the joint probability density of the combiner output signal to noise ratios at two time instants,  $\gamma_1$  and  $\gamma_2$ , is, by using expression (14), for  $\gamma_1 < \gamma_T$  and  $\gamma_2 < \gamma_T$ :

$$p_{\gamma_1\gamma_2}(\gamma_1, \gamma_2) = P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}\gamma_{12}}(\gamma_{11}, \gamma_{12}) d\gamma_{11} \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}\gamma_{22}}(\gamma_{21}, \gamma_{22}) d\gamma_{21} \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} = P_1 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) \quad (16)$$

In the similar way we can derive the other joint probability density functions. The joint PDF is, for  $\gamma_1 \geq \gamma_T$  and  $\gamma_2 < \gamma_T$ :

$$p_{\gamma_1\gamma_2}(\gamma_1, \gamma_2) = P_1 \cdot p_{\gamma_{22}}(\gamma_2) \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} + P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}\gamma_{12}}(\gamma_{11}, \gamma_{12}) d\gamma_{11} \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} +$$

$$+ P_2 \cdot p_{\gamma_{12}}(\gamma_2) \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}\gamma_{22}}(\gamma_{21}, \gamma_{22}) d\gamma_{21} \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} = P_1 \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \cdot \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + P_1 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) \quad (17)$$

for  $\gamma_1 < \gamma_T$  and  $\gamma_2 \geq \gamma_T$ :

$$p_{\gamma_1\gamma_2}(\gamma_1, \gamma_2) = P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}}(\gamma_{11}) d\gamma_{11} \cdot p_{\gamma_{21}\gamma_{22}}(\gamma_1, \gamma_2) + P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}\gamma_{12}}(\gamma_{11}, \gamma_{12}) d\gamma_{11} \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}}(\gamma_{21}) d\gamma_{21} \cdot p_{\gamma_{11}\gamma_{12}}(\gamma_1, \gamma_2) + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}\gamma_{22}}(\gamma_{21}, \gamma_{22}) d\gamma_{21} \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} = P_1 \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) + P_2 \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) \cdot \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) + P_1 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) \quad (18)$$

$$\begin{aligned}
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + \\
 & + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \\
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right)
 \end{aligned} \quad (18)$$

for  $\gamma_1 \geq \gamma_T$  and  $\gamma_2 \geq \gamma_T$ :

$$\begin{aligned}
 p_{\gamma_1\gamma_2}(\gamma_1, \gamma_2) &= P_1 \cdot p_{\gamma_{11}\gamma_{12}}(\gamma_1, \gamma_2) + P_1 \cdot p_{\gamma_{22}}(\gamma_2) \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} + \\
 & + P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}}(\gamma_{11}) d\gamma_{11} \cdot p_{\gamma_{21}\gamma_{22}}(\gamma_1, \gamma_2) + P_1 \cdot \int_0^{\gamma_T} p_{\gamma_{11}\gamma_{12}}(\gamma_{11}, \gamma_{12}) d\gamma_{11} \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} + \\
 & + P_2 \cdot p_{\gamma_{21}\gamma_{22}}(\gamma_1, \gamma_2) + P_2 \cdot p_{\gamma_{12}}(\gamma_2) \int_0^{\gamma_T} p_{\gamma_{22}\gamma_{21}}(\gamma_{22}, \gamma_{21}) d\gamma_{22} + \\
 & + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}}(\gamma_{21}) d\gamma_{21} \cdot p_{\gamma_{11}\gamma_{12}}(\gamma_1, \gamma_2) + P_2 \cdot \int_0^{\gamma_T} p_{\gamma_{21}\gamma_{22}}(\gamma_{21}, \gamma_{22}) d\gamma_{21} \int_0^{\gamma_T} p_{\gamma_{12}\gamma_{11}}(\gamma_{12}, \gamma_{11}) d\gamma_{12} = \\
 & = P_1 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \\
 & \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) + \\
 & + P_1 \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \cdot \\
 & \cdot \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) + \\
 & + P_1 \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \\
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \\
 & \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) + \\
 & + P_1 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \\
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + \\
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right)
 \end{aligned}$$

$$\begin{aligned}
 & + P_2 \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \\
 & \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) + \\
 & + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) \cdot
 \end{aligned}$$

$$\begin{aligned}
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_1}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_1}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) + \\
 & + P_2 \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right) \cdot \\
 & \cdot \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \\
 & \exp\left(-\frac{(1+q_1^2)^2\gamma_2}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_2}{4q_1^2\bar{\gamma}_1}\right) + \\
 & + P_2 \frac{(1+q_1^2)}{2q_1\bar{\gamma}_1} \exp\left(-\frac{(1+q_1^2)^2\gamma_1}{4q_1^2\bar{\gamma}_1}\right) I_0\left(\frac{(1-q_1^4)\gamma_1}{4q_1^2\bar{\gamma}_1}\right) \frac{2q_1}{1+q_1^2} I_e\left(\frac{1-q_1^2}{1+q_1^2}, \frac{(1+q_1^2)^2\gamma_T}{4q_1^2\bar{\gamma}_1}\right) \cdot \\
 & \cdot \frac{(1+q_2^2)}{2q_2\bar{\gamma}_2} \exp\left(-\frac{(1+q_2^2)^2\gamma_2}{4q_2^2\bar{\gamma}_2}\right) I_0\left(\frac{(1-q_2^4)\gamma_2}{4q_2^2\bar{\gamma}_2}\right) \frac{2q_2}{1+q_2^2} I_e\left(\frac{1-q_2^2}{1+q_2^2}, \frac{(1+q_2^2)^2\gamma_T}{4q_2^2\bar{\gamma}_2}\right)
 \end{aligned} \quad (19)$$

#### IV. NUMERICAL RESULTS

It is simple to present these expressions grafically using mathematical software, for example “MatLab”. Because of simplicity we supposed that the variances of both signals at the combiner input are equal.

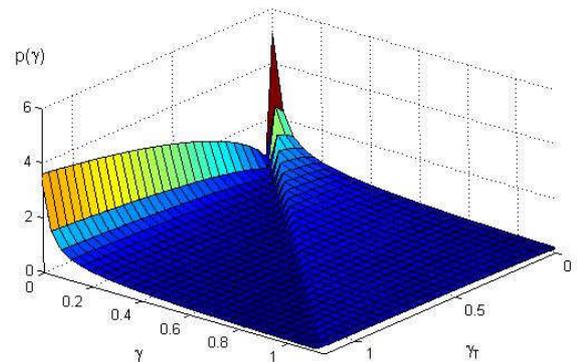


Figure 3. Probability density function of the combiner output signal at one time instant for  $\bar{\gamma}_1 = \bar{\gamma}_2 = 1$ ,  $q_1 = q_2 = 0.5$

In the case we observe one time instant, Fig. 3, the probability density function of the combiner output signal to noise ratio is determined as function of input signal to noise ratio  $\gamma$  and the threshold  $\gamma_T$ , for three different variance values for the same distribution parameters in branches of the receiver.

When we observe two time instants, Figs. 4-6, the PDF is given versus input signals at two time instants,  $\gamma_1$  and  $\gamma_2$ , for different values of the distribution parameters and the threshold  $\gamma_T$ .

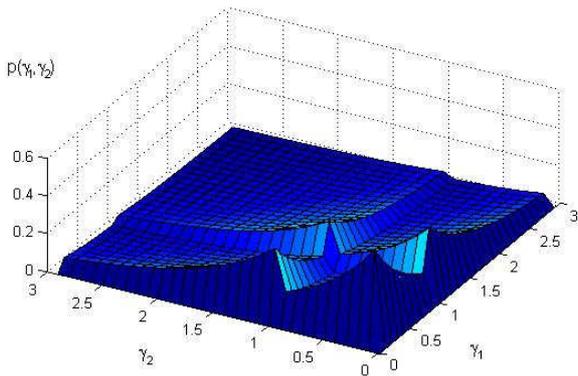


Figure 4. The probability density function of the combiner output signal at two time instants for  $\bar{\gamma}_1 = \bar{\gamma}_2 = 1, \gamma_T = 1, q_1=q_2=0.5$

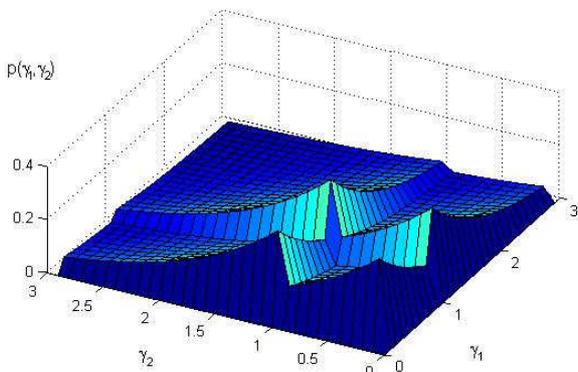


Figure 5. The probability density function of the combiner output signal at two time instants for  $\bar{\gamma}_1 = \bar{\gamma}_2 = 1, \gamma_T = 1, q_1=q_2=0.9$

The bit error probability of digital telecommunication systems in the presence of Hoyt fading can be calculated by the probability density function obtained here. The outage probability also can be calculated using PDF.

The performances of the Switch and Stay Combining/ Selection Combining (SSC/SC) combiner output signal at two time instants in the presence of different types of fading, are determined in our other papers where the results are shown graphically to highlight better performances of

the SSC/SC combiner compared to classical SSC and SC combiners at one time instant.

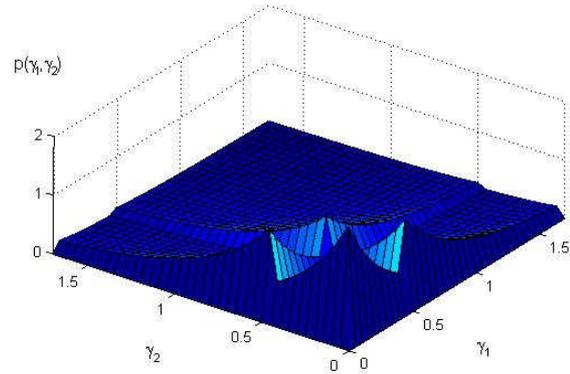


Figure 6. The probability density function of the combiner output signal at two time instants for  $\bar{\gamma}_1 = \bar{\gamma}_2 = 0.5, \gamma_T = 0.5, q_1=q_2=0.5$

## V. CONCLUSION

The probability density function of the dual branches SSC combiner output signal at one time instant and the joint probability density function of the SSC combiner output signal at two time instants are determined in closed form. The obtained results are shown graphically for different variance values and decision threshold values.

The bit error probability of digital telecommunication systems in the presence of Hoyt fading can be calculated by the probability density function. The system performances can be significantly improved using the sampling at two time instants. The authors showed in an other work, based on the results obtained in this paper, that the error probability is significantly reduced if the decision making is performed in two time instants. This fact shows that the results obtained in this study are very significant for further research and application in the designing of diversity receivers.

## ACKNOWLEDGMENT

This work has been funded by the Serbian Ministry for Science under the projects TR33035 and III 044006.

## REFERENCES

- [1] W.C.Y. Lee, Mobile communications engineering, Mc-Graw-Hill, New York, 0-7803-7005-8/01, 1992.
- [2] D. Drajić, The introduction into the statistical theory of telecommunications, Akademska misao, Beograd, 2003. (in serbian)
- [3] C-X. Wang, N. Youssef, and M. Patzold, "Level-crossing rate and average duration of fades of deterministic simulation models for Nakagami-Hoyt fading channels", 5th International Symposium on Wireless Personal Multimedia Communications, vol.1, pp. 272-276, Oct. 2002.
- [4] G. Fraidenraich, J. Filho, and M. Yacoub, "Second-order statistics of Maximal-Ratio and Equal-Gain Combining in Hoyt fading", IEEE Commun. Lett, vol. 9, No 1, pp. 19-21, Jan. 2005.

- [5] B. Chytil, "The distribution of amplitude scintillation and the conversion of scintillation indices," *J. Atmos. Terr. Phys.*, vol. 29, pp. 1175-1177, Sept. 1967.
- [6] S. Khatalin and J. P. Fonseka, "On the channel capacity in Rician and Hoyt fading environments with MRC diversity," *IEEE Trans. Veh. Technol.*, vol. 55, no. 1, pp. 137-141, Jan. 2006.
- [7] G. A. Ropokis, A. A. Rontogiannis, and P. T. Mathiopoulos, "Information outage probability of orthogonal space-time block codes over Hoyt fading channels," in *International Workshop on Satellite and Space Communications*, vol. 2, pp. 135-139, Sept. 2006.
- [8] A. Anamalai, C. Tellambura, and V. K. Bhargava, "Simple and accurate methods for the outage analysis in cellular mobile radio systems-A unified approach," *IEEE Trans. Commun.*, vol. 49, no. 2, pp. 303-316, Feb. 2001.
- [9] J. Cheng and T. Berger, "Capacity of Nakagami-q (Hoyt) fading channels with channel side information," in *Proc. IEEE International Conference on Communication Technology*, vol. 55, pp. 1915-1918, Apr. 2003.
- [10] M. S. Alouini, and M. K. Simon, "Postdetection Switched Combining- A simple Diversity Scheme with Improved BER Performance", *IEEE Trans. on Commun.*, vol. 51, No 9, pp.1591-1602, September 2003.
- [11] Y. C. Ko, M. S. Alouini, and M. K. Simon, "Analysis and optimization of switched diversity systems", *IEEE Trans. Veh. Technol.*, vol. 49, Issue 5, pp.1813-1831, Sep 2000.
- [12] M. Č. Stefanović., D. S. Krstić, M. Petrović, Dj. Bandjur, "The Joint Probability Density Function of the SSS Combiner Output Signal at Two Time Instants in the Presence of Rayleigh Fading", *The Second European Conference on Antennas and Propagation, EuCAP 2007*, Edinburgh, UK, 11-16 Nov. 2007.
- [13] D. Krstić, P. Nikolić., M. Matović, A. Matović, and M. Stefanović, "The Joint Probability Density Function of the SSC Combiner Output Signal in the Presence of Nakagami-m fading" *The Fourth International Conference on Wireless and Mobile Communications, ICWMC 2008*, ISBN: 978-0-7695-3274-5, Athens/Vouliagmeni, Greece, pp. 409-416, July 27-August 1, 2008.
- [14] P. B. Nikolic, M. C. Stefanović, D. S. Krstić, and G. Lj. Stamenović, "The Joint Probability Density Function of the SSS Combiner Output in the Presence of Weibull Fading", *XV International Symposium on Theoretical Electrical Engineering, ISTET'09*, Lübeck, Germany, pp. 20-24, 22 - 24 June 2009.
- [15] D. S. Krstić, P. B. Nikolić, F. Destović, and M. C. Stefanović, "The Joint Probability Density Function of the SSC Combiner Output Signal in the Presence of Log-Normal Fading in two Time Instants", *Electronics and Electrical Engineering*, Kaunas: Technologija ISSN 1392-1215, No. 3 (109), pp. 11-16, 2011.
- [16] B. Nikolic and G. Djordjevic, "Performance of SC and SSC Receivers in Hoyt Channel in the Presence of Imperfect Reference Signal Extraction", *9th International Conference on Applied Electromagnetics*, Niš, Serbia, Aug. 31 - Sept. 02, 2009.
- [17] Simon M. K., Alouini M. S., *Digital Communication over Fading Channels*, Second Edition, Wiley-Interscience, A John Wiley&Sons, Inc., Publications, New Jersey, 2005.
- [18] B. T. Tan, T. T. Tjhung, C. H. Teo, and P. Y. Leong, "Series representations for Rice's I<sub>0</sub> function," *IEEE Trans. on Communications*, vol. 32, no. 12, pp. 1342-1344, December 1996.
- [19] I. S. Gradshteyn and I. M. Ryzhik, "Table of Integrals, Series, and Products", 5th ed. San Diego, CA: Academic Press, 1994.

# Hybrid Log-MAP Algorithm for Turbo Decoding Over AWGN Channel

Li Li Lim  
 Faculty of Engineering  
 The University of Nottingham  
 Selangor, Malaysia  
 Email: keyx7lll@nottingham.edu.my

David Wee Gin Lim  
 Faculty of Engineering  
 The University of Nottingham  
 Selangor, Malaysia  
 Email: lim.wee-gin@nottingham.edu.my

**Abstract**—In Log-MAP turbo decoding, the complicated log exponential sum is often simplified with the Jacobian logarithm which consists of the max operation along with an exponential correction function. Although the Max-Log-MAP reduces the complexity of the Jacobian logarithm implementation by omitting the correction function, its performance is inferior to the exact Log-MAP algorithm. Hence, a simple approximation to the correction function is needed to complement the Max-Log-MAP algorithm. In this paper, a new suboptimal hybrid Log-MAP algorithm for decoding turbo code is proposed. It approximates the exact Log-MAP accurately and is simple for hardware implementation. The performance of the hybrid Log-MAP algorithm is shown to have the closest performance to the exact Log-MAP solution especially at low SNR. The result of other Log-MAP based algorithm are also presented.

**Index Terms**—Log-MAP, turbo codes, correction function, Jacobian logarithm.

## I. INTRODUCTION

Turbo codes were first introduced by Berrou et al [1] and is among the most powerful error correcting codes. The MAP decoder is often operated in the log domain in order to reduce computational complexity [2]. However, the computation for state metrics and log likelihood ratio (LLR) is still burdened by the log exponential sum calculation. The Jacobian logarithm is used to simplify the log exponential sum by employing a correction function along with the maximum operator in the log domain [3]. Although simplified, the correction function remains a nonlinear exponential function. The manner in which the correction function is calculated is critical to the performance and complexity of the decoder. Several methods have been proposed to simplify its computation which gives a tradeoff between complexity and performance [4]–[6].

This paper presents a simplified hybrid algorithm for sub optimal Log-MAP. The algorithm employs linear fitting methods as well as minimal bitwise shift operations to compute the correction function that will be used in the calculations of state metrics and ultimately the LLR. In this algorithm, only linear multiplication, addition, comparator, and minimal number of bit shifts are required to obtain near Log-MAP performance.

The rest of the paper is organized as follows: in Section II, a brief review of the original MAP, Log-MAP as well as Max-Log-MAP algorithm are presented. Section III reviews existing

methods for the approximation of the correction function. Next, we present the novel hybrid approximation in Section IV. In Section V, the simulation results of the hybrid Log-MAP in comparison with other approximation methods is presented. The paper is finally concluded in Section VI.

## II. THE MAP, LOG-MAP, AND MAX-LOG-MAP ALGORITHMS

In this section, the derivation of the MAP algorithm will not be detailed but the results of the algorithm will be stated. For details on the derivation, see [1] and [2]. Following this, a review of the Log-MAP as well as the introduction to the  $\max^*$  operator in the Log-MAP algorithm and the Max-Log-MAP algorithm will be presented.

### A. The Encoder

The MAP algorithm operates in blocks of binary input (information) data sequence represented by  $\vec{d} = (d_1, \dots, d_N)$ . The encoder consists of two identical recursive systematic convolutional (RSC) encoders with  $M$  memory elements. We will assume that each of the RSC encoder has a rate of  $1/2$  with two outputs, which is the sequence of systematic bits  $\vec{x}^s = (x_1^s, \dots, x_N^s)$ , and the parity bit sequence,  $\vec{x}^p = (x_1^p, \dots, x_N^p)$ . The outputs are punctured and transmitted over the channel. The corresponding received sequences are  $\vec{y}^s$  and  $\vec{y}^p$  or for brevity,  $\vec{y} = (\vec{y}^s, \vec{y}^p)$  which refers to the pair of received systematic and parity bit received.

### B. Maximum A Posteriori (MAP) Algorithm

Let the state of the encoder at time  $k$  be  $S_k$ , where it can take on values between 0 and  $2^M - 1$ . The bit  $d_k$  represents the transition from step  $k - 1$  to  $k$ . The goal of the MAP algorithm is to provide the LLR,  $\Lambda(d_k)$  of the a posteriori probability (APP) of  $d_k = 1$  to the APP of  $d_k = 0$  as follows:

$$\Lambda(d_k) = \ln \frac{\sum_{S_k} \sum_{S_{k-1}} \gamma_1(y_k, S_{k-1}, S_k) \cdot \alpha_{k-1}(S_{k-1}) \cdot \beta_k(S_k)}{\sum_{S_k} \sum_{S_{k-1}} \gamma_0(y_k, S_{k-1}, S_k) \cdot \alpha_{k-1}(S_{k-1}) \cdot \beta_k(S_k)}. \quad (1)$$

The forward state metric  $\alpha_k$  can be expressed as:

$$\alpha_k(S_k) = \frac{\sum_{S_{k-1}} \sum_{i=0}^1 \gamma_i(y_k, S_{k-1}, S_k) \cdot \alpha_{k-1}(S_{k-1})}{\sum_{S_k} \sum_{S_{k-1}} \sum_{i=0}^1 \gamma_i(y_k, S_{k-1}, S_k) \cdot \alpha_{k-1}(S_{k-1})} \quad (2)$$

and the backward state metric  $\beta_k$  is given as:

$$\beta_k(S_k) = \frac{\sum_{S_{k+1}} \sum_{i=0}^1 \gamma_i(y_{k+1}, S_k, S_{k+1}) \cdot \beta_{k+1}(S_{k+1})}{\sum_{S_k} \sum_{S_{k+1}} \sum_{i=0}^1 \gamma_i(y_{k+1}, S_k, S_{k+1}) \cdot \alpha_k(S_k)} \quad (3)$$

The branch transition probabilities,  $\gamma_i$  are given as:

$$\gamma_i((y_k^s, y_k^p), S_{k-1}, S_k) = \exp\left[\frac{1}{2} x_k^s (L_e(x_k^s) + L_c y_k^s) + L_c y_k^p x_k^p\right]; \quad (4)$$

where the channel reliability value,  $L_c = \frac{2}{\sigma^2}$  with  $\sigma^2$  being the noise variance and  $L_e$  is the *extrinsic* information which serves as the *a priori* information.

### C. Log-MAP Algorithm

The MAP algorithm is too complex for practical implementation in a real system. To avoid complicated operations, the entire MAP algorithm can be computed in the log domain. By taking the logarithm of  $\alpha_k(S_k)$ ,  $\beta_k(S_k)$  and  $\gamma_k(S_{k-1}, S_k)$ , the MAP algorithm reduces to addition and multiplication operations. However, the computation of forward state metric,  $\tilde{\alpha}_k(S_k) = \ln \alpha_k(S_k)$  involves the log exponential sum which is complicated to implement in hardware:

$$\begin{aligned} \tilde{\alpha}_k(S_k) = & \ln\left(\sum_{S_{k-1}} \sum_{i=0}^1 e^{\ln \gamma_i((y_k^s, y_k^p), S_{k-1}, S_k) + \ln \alpha_{k-1}(S_{k-1})}\right) \\ & - \ln\left(\sum_{S_k} \sum_{S_{k-1}} \sum_{i=0}^1 e^{\ln \gamma_i((y_k^s, y_k^p), S_{k-1}, S_k) + \ln \alpha_{k-1}(S_{k-1})}\right); \end{aligned} \quad (5)$$

Simplifying (5) gives the general form:

$$F(x_1, x_2, \dots, x_n) = \ln\left(\sum_{i=1}^n e^{x_i}\right); \quad (6)$$

where  $n$  is a function of the encoder states. Consider the Jacobian logarithm for two variables given as:

$$\begin{aligned} \max^*(x_1, x_2) = & \ln(e^{x_1} + e^{x_2}) \\ = & \max(x_1, x_2) + \ln(1 + e^{-|x_1 - x_2|}) \\ = & \max(x_1, x_2) + f_c(x) \end{aligned} \quad (7)$$

where  $f_c(x)$  is the correction function, and  $\max(x_1, x_2)$  is the maximum of the function's two arguments. The computation of the exponential term of  $f_c(x)$  in  $\max^*(x_1, x_2)$  greatly increases the complexity of the Log-MAP algorithm. A simple way to obtain (6) with  $n > 2$  is to recursively perform the  $\max^*$  operator as follows [6]:

$$F(x_1, x_2, \dots, x_n) = \max^*(x_n, \max^*(x_{n-1}, \dots, \max^*(x_3, \max^*(x_2, x_1))) \dots). \quad (8)$$

Similar forms of the log-exponential-sum occurs for the calculation of the backward state metric,  $\beta_k(S_k)$  and the LLR,  $\Lambda(d_k)$ . It is clear that in order to improve the performance of the Log-MAP algorithm, a simpler implementation to the correction function must be found. One of the earliest significant finding is to present the correction function as a single dimension look up table [3]. However, table sequences require storage, and additional memory units will have to be added to the Log-MAP decoder rendering the increase in area, power, and a drop in the overall speed [5]. Hence, another sub-optimal solution to this would be to implement the correction function with a simpler approximate as will be detailed in Section III.

### D. Max-Log-MAP Algorithm

With the Max-Log-MAP algorithm the  $\max^*$  operation is loosely approximated using

$$\max^*(x_1, x_2) \approx \max(x_1, x_2). \quad (9)$$

The Max-Log-MAP simplifies the Log-MAP algorithm by simply omitting the correction function  $f_c(x)$  altogether. The performance for the Max-Log-MAP algorithm gives up to a 10% performance drop [7] when compared to the Log-MAP. The Max-Log-MAP algorithm is the least complex of all the existing methods but offers the worst BER performance. This creates a need to complement the Max-Log-MAP algorithm with a simple implementation of the correction function in order to improve performance.

## III. SUB OPTIMAL LOG-MAP ALGORITHMS:

### APPROXIMATION TO THE CORRECTION FUNCTION.

This section gives a brief review of existing algorithms which approximates the correction function in order to achieve a simple implementation yet improved performance as compared to Max-Log-MAP.

#### A. Constant Log-MAP Algorithm.

In this algorithm proposed by [4], the correction function  $f_c(x)$  is approximated with the following rule:

$$f_c(x) = \begin{cases} \frac{3}{8}, & -2 \leq x < 2 \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

The constant Log-MAP algorithm offers a simple implementation in hardware but with trade off in performance.

#### B. Linear Log-MAP Algorithm.

In [5], the author suggests a linear approximation to the correction function by employing the MacLaurin Series expansion. It is observed that the correction function is effective when  $f_c(x)$  is around zero. Therefore, the Maclaurin series can be exploited to approximate the correction function about zero. By neglecting Maclaurin's series order two and above, the approximation for the correction term is given as:

$$f_c(x) \approx \max(0, \ln 2 - \frac{1}{2}x). \quad (11)$$

This approximation offers better performance than the Constant Log-MAP algorithm and requires only a simple linear implementation.

### C. Multistep Log-MAP Algorithm.

A more accurate and elegant solution to approximate the correction function is given by [6]. The approximation to the correction term as suggested by [6] is given as:

$$f_c(x) \approx \frac{\ln 2}{2^{\lfloor x+0.5 \rfloor}}. \quad (12)$$

Where  $\lfloor x+0.5 \rfloor$  denotes the largest integer that is smaller or equal to  $x+0.5$ . The correction term given here is a more accurate yet simple approximation to the correction function. Note that division by 2 can be easily done in digital systems by implementing  $\lfloor x+0.5 \rfloor$  number of binary shifts. The algorithm employs shift registers storing the constant  $\ln(2)$  to perform the division in [8]. However, in order to facilitate fast computation, a high speed shift register is needed for this algorithm.

## IV. NOVEL HYBRID LOG-MAP APPROXIMATION

In our novel development, our approximation has the advantage over the linear and multistep Log-MAP algorithm in [5] and [6] respectively in terms of accuracy as well as hardware simplicity. The hybrid approximation is proposed as:

$$f_c(x) \approx \begin{cases} 0.6512 - 0.3251x & \text{for } x < 1.5 \\ \frac{0.1635}{2^{\lfloor 0.5x \rfloor}} & \text{otherwise} \end{cases} \quad (13)$$

When plotted against the exact correction function in Fig. 1, the hybrid approximation proves to be a better fit to the correction term than [5] and [6]. The hybrid approximation is divided into two regions i.e.,  $|x| < 1.5$  and  $|x| \geq 1.5$ .

In the region of  $|x| < 1.5$  the hybrid algorithm employs a linear polynomial fit. The accuracy of the approximation is verified through the goodness of fit test against the exact Log-MAP curve with parameters of SSE: 0.05058, R-square: 0.9836, Adjusted R-square: 0.9835, and RMSE: 0.01842. The polynomial fit is acceptably accurate as indicated with a confidence interval of 95%.

Inspired by [6], a good approximation can be achieved by adjusting Equ. (12). The constant  $\ln 2$  is replaced with the constant 0.1635 where it is the corresponding value for  $|x| = 1.5$  for the linear region. The replacement gives a double advantage. The replacement with the constant 0.1635 requires lesser number of bitwise shifts for larger values of  $|x|$  and this reduces the number of shift operations needed to perform computation for  $f_c(x \geq 1.5)$ . It is observed that  $f_c(x)$  can be approximated to zero for values of  $|x| > 4$ . With this adjustment, only a maximum number of 3 shifts (i.e., 3 extra shift operations) would suffice to give a good approximation and thus reducing the number of shift operations as compared to [8]. A good approximation using the multistep approximation can continue to be achieved by adjusting the number of integer shifts to  $\lfloor 0.5x \rfloor$  to accommodate the change in the replacement of the constant 0.1635 in the multistep region. Note that the

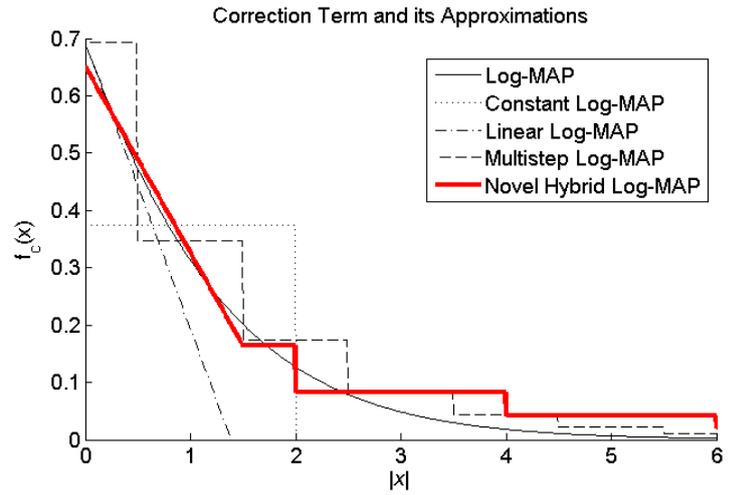


Fig. 1. Approximations to the correction function in Equ. (1).

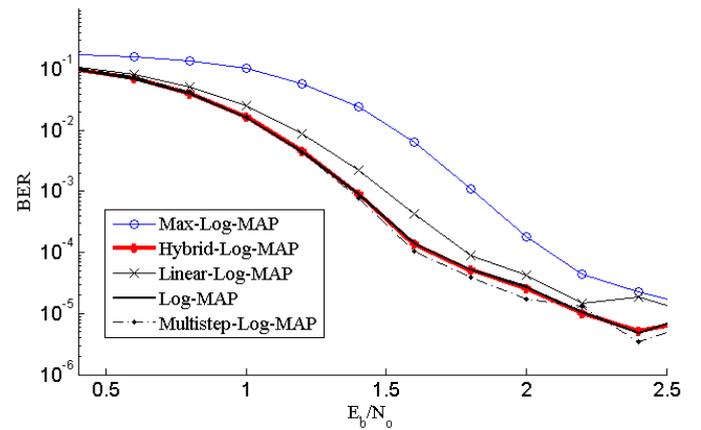


Fig. 2. BER performances for 16 states Turbo decoder with different correction functions.

calculation for  $\lfloor 0.5x \rfloor$  can be easily achieved by bit shifting as well. To facilitate fast computation, the value of  $0.5x$  can be performed by implementing a dedicated logic circuit to readout values of  $0.5x$  which provides a fast combinational logic solution as compared to the shift register method where it is sequential.

## V. SIMULATION PARAMETERS AND RESULTS

Two 16 state parallel concatenated Log-MAP turbo decoder with generator polynomial  $g[23, 33]$ , implemented in 5 iterations, with overall rate of  $1/2$ , and a random interleaver of size 1000 bits was used in the simulation over the AWGN channel for 10,000 bits modulated using BPSK.

The BER performance for the HLM algorithm including Log-MAP, Max-Log-MAP, Linear Log-MAP, and Multistep Log-MAP are presented in Fig. 2. The HLM algorithm is shown to have the closest performance to the exact Log-MAP solution. The HLM algorithm outperforms other algorithm especially in lower SNRs shown in Fig. 3. This is because the Log-MAP algorithm is sensitive to the SNR [8]. At high SNR,

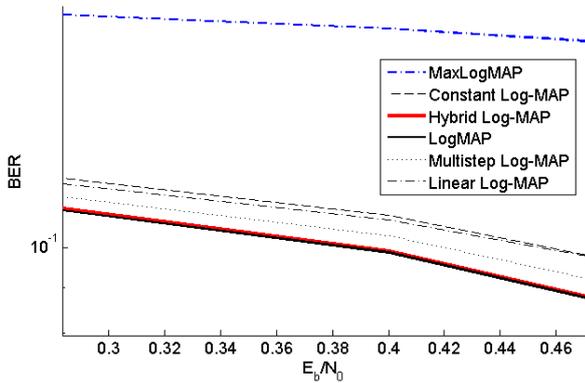


Fig. 3. BER performances at low  $E_b/N_0$ .

the performance of the turbo code approaches the Max-Log-MAP performance and does not rely heavily on the correction function. However, if the SNR is low, the decoder considers the *a priori* or *extrinsic* information from the previous decoder more. The *a priori* is Gaussian distributed with increased number of iteration [9] which causes an increased distribution of argument  $|x|$  in the correction function to regions close to zero where it is most effective. Due to the good fit of the HLM for smaller arguments of  $|x|$ , the HLM algorithm performs better than other approximations in low SNR. The sensitivity of the Log-MAP algorithm to SNR is also more pronounced in encoders with more memory elements [8].

## VI. CONCLUSION

The exact Log-MAP algorithm involves computationally intensive operations in order to attain the ideal performance. Neglecting the correction function, the Log-MAP algorithm reduces to the simple Max-Log-MAP algorithm. However, this rough approximation gives a capacity loss, and hence the correction term will have to be included or approximated which gives us the suboptimal Log-MAP algorithm. The novel HLM algorithm is a suboptimal Log-MAP solution that achieves nearly identical performance to the Log-MAP algorithm. The hybrid approximation offers a simple implementation on hardware involving shift registers, multiplications, comparators, and addition operations. In addition, we also show that the hybrid algorithm outperforms existing Log-MAP based algorithm.

## ACKNOWLEDGMENT

The authors would like to thank Tunku Abdul Rahman College and its staff for useful discussion and feedback.

## REFERENCES

- [1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," *IEEE International Conference on Communications*, May 1993.
- [2] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Information Theory*, vol. 20, pp. 284 – 287, Mar 1974.

- [3] P. Robertson, E. Villebrun, and P. Hoeher, "A comparison of optimal and sub-optimal MAP decoding algorithms operating in the log domain," *IEEE International Conference on Communications*, vol. 2, pp. 1009–1013, Jun 1995.
- [4] W. J. Gross and P. G. Gulak, "Simplified MAP algorithm suitable for implementation of turbo decoders," *Electronic Letters, IET Journal*, vol. 34, pp. 1577 – 1578, Aug 1998.
- [5] S. Talakoub, L. Sabeti, B. Shahrava, and M. Ahmadi, "A linear log-MAP algorithm for turbo decoding and turbo equalization," *IEEE International Conference on Wireless And Mobile Computing, Networking And Communications*, vol. 1, pp. 182 – 186, Aug 2005.
- [6] Hao Wang, Hongwen Yang, and Dacheng Yang, "Improved log-MAP decoding algorithm for turbo-like codes," *IEEE Communications Letters*, vol. 10, pp. 186 – 188, Mar 2006.
- [7] Lin Zhang and Shun-Zheng Yu, "A simplified log-MAP turbo decoder by fitting method," *IEEE International Conference on Advanced Communication Technology*, vol. 2, pp. 854–857, Jul 2005.
- [8] M. A. Khalighi, "Effect of mismatched SNR on the performance of log-MAP turbo detector," *IEEE Trans. Vehicular Technology*, vol. 52, pp. 1386 – 1397, Sept 2003.
- [9] M. Fu, "Stochastic analysis of turbo decoding," *IEEE Trans. Information Theory*, vol. 51, pp. 81 – 100, Jan 2005.

# Evaluating Transceiver Power Savings Produced by Connectivity Strategies for Infrastructure Wireless Mesh Networks

Pragasen Mudali, Murimo B. Mutanga and Matthew O. Adigun  
 Department of Computer Science  
 University of Zululand  
 South Africa  
 Email: pmudali, madigun@pan.uzulu.ac.za

Ntsibane Ntlatlapa  
 Meraka Institute  
 Council for Scientific and Industrial Research  
 South Africa  
 Email: nntlatlapa@csir.co.za

**Abstract**—Infrastructure Wireless Mesh Networks (I-WMNs) are increasingly used to bridge the digital divide in rural areas around the world. Rural African areas in particular require energy efficient I-WMNs as the nodes comprising the I-WMN backbone network may be battery-powered in the absence of reliable power supplies. A key requirement for the proper functioning of the I-WMN backbone is that network connectivity be maintained. Two main types of connectivity strategies exist in the literature and the more practical Critical Number of Neighbors (CNN) method is focused upon. Three CNN-based connectivity strategies are evaluated via simulation to determine their effect on transceiver power savings when applied to the I-WMN backbone. The evaluation shows that these strategies are capable of cumulative transceiver power savings (in excess of 10%) and that the capacity for transceiver power savings largely corresponds to the position of a node relative to the (imaginary) network center. However, the evaluated connectivity strategies were found not to increase the network lifetime due to the nature of the network topologies created by these strategies. This particular result is however dependent upon the node energy model employed and further experiments with differing energy models are required to confirm this finding.

**Index Terms**—wireless mesh networks; connectivity; power savings; network lifetime; topology control

## I. INTRODUCTION

Wireless Mesh Networks (WMNs) are increasingly used as both an inexpensive alternative to broadband provisioning in urban areas and as a primary method for broadband provisioning in rural areas. The most common form of WMN deployment consists of a two-tier architecture comprising an access and a backbone network. This type of WMN is commonly referred to as an Infrastructure WMN (I-WMN). Client devices connect to the I-WMN backbone which is typically self-organizing and self-configuring. These backbone nodes, comprising Mesh Points (MPs), Mesh Access Points (MAPs) and Mesh Portals (MPPs), collaborate to maintain network connectivity and deliver traffic to the intended destinations. (see Figure 1).

Despite the stationary nature of the I-WMN backbone, maintaining network connectivity is made difficult by the transient nature of wireless links, making them unreliable

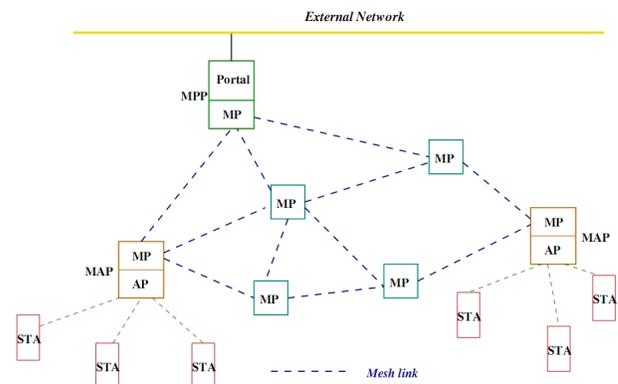


Fig. 1. Infrastructure WMN Architecture [12]

when deployed [1], [2], [3]. Network connectivity is traditionally assured by ensuring that each device in the I-WMN backbone utilizes its maximum transceiver power. The use of maximum transceiver power is disadvantageous, resulting in high levels of interference, increased contention for the shared transmission medium, a reduction in network capacity and unnecessary transceiver power consumption.

Conventional wisdom holds that WMNs do not suffer from power constraints [4], but in the rural African context where electrical mains power is often unreliable or more often non-existent, such an assumption is easily disproved. The I-WMN backbone is often battery- or solar-powered [5] and thus, in the rural African context, any power savings are welcomed such that the operational lifetime of the network may be maximized. Operating a network at maximum transceiver power output in this scenario is thus an ill-afforded luxury. The rural African context also constrains I-WMN deployments (and their associated QoS mechanisms) to those that are as autonomous as possible due to the lack of technical expertise in these areas.

As a result of the inefficiencies associated with the use of maximum transceiver power outputs, several studies have been undertaken to devise strategies for optimal network

connectivity (where network connectivity is maintained with minimal transceiver power outputs). These connectivity strategies have demonstrated that they possess the ability to create interference-efficient network topologies [6], [7] as well as to provide route redundancy in some cases [8], [9], [10]. These connectivity strategies have been shown to produce cumulative transceiver power savings but, to the best of our knowledge, the effect of these transceiver power savings on the network lifetime has not been evaluated.

In this paper, various connectivity strategies based on the Critical Number of Neighbors approach are evaluated via simulation to determine the relationship between transceiver power savings and the network lifetime. The selected connectivity strategies have been subjected to an indoor I-WMN testbed evaluation [11] and the study presented in this paper attempts to validate the transceiver power output pattern reported in [11].

The evaluation reported in this paper indicates that the selected connectivity strategies are able to produce cumulative transceiver power savings. The extent of the power savings produced by individual backbone nodes is largely dependent upon the location of the node relative to the (imaginary) center of the backbone network. The evaluation also suggests that cumulative transceiver power savings do not automatically translate into corresponding extensions of network lifetime.

The remainder of this paper is organized as follows. Section 2 provides a review of the connectivity present in the literature. Section 3 details the simulation setup and measurement methodologies employed in this study whilst Section 4 contains the performance evaluation of the selected connectivity strategies. Section 5 concludes the paper.

## II. CONNECTIVITY STRATEGIES APPLICABLE TO WIRELESS MESH NETWORKS

Connectivity in the backbone of infrastructure WMNs can be achieved by using one of three possible approaches. The first approach specifies that each node utilizes its maximum transceiver power. The second approach determines the minimum transmission range, also dubbed the Critical Transmitting Range (CTR), required to maintain a connected network and the transceiver power output for all nodes is adjusted to sustain this transmission range. The third approach determines the optimal number of neighbors to be maintained in order to ensure network connectivity.

The CTR approach results in the homogeneous assignment of transceiver powers and may not minimize total transceiver power output. This approach is highly susceptible to the effect of outlying nodes that force a high common power level (or equivalently, transmission range) [13]. Using the CTR to achieve network connectivity may be done in one of three ways, each with its own disadvantages. The first technique requires that a central node determine the appropriate CTR and this value is subsequently broadcast throughout the network. Each node then automatically adjusts its own transceiver power output. The second technique also requires that the CTR is determined at a central location but the network nodes

are manually adjusted to maintain this transmission range, as described in [14]. The third technique requires that all nodes broadcast their positions and the CTR is subsequently determined locally at each node, generating high messaging overheads. Thus, the practicality of the use of the CTR approach becomes limited when mobile nodes or dynamic network sizes are taken into account.

An alternative approach is for each backbone node to maintain an optimal number of one-hop neighbors, also referred to as the Critical Number of Neighbors (CNN). This approach may result in heterogeneous transceiver power outputs, potentially maximizing transceiver power savings. In addition, the CNN is less affected by the distribution and position of network nodes so there is no need to assume a uniform or homogeneous backbone node distribution or a GPS-enabled device. Lastly, maintaining connectivity via a CNN potentially eliminates human intervention (especially when a proactive routing protocol is employed) which is of fundamental importance if true autonomous configuration is to be realized in WMNs.

The CNN approach possesses the advantage of being distributed in nature and relying on locally-available information. This approach is also the most likely to lead to autonomous power and topology control mechanisms that are able to produce cumulative transceiver power savings whilst maintaining network connectivity. Thus, the CNN approach forms the basis for the work reported in this paper.

Prior research has produced CNN values that are both independent and dependent of the network size (total number of nodes), which are discussed below.

### A. Network-Size-independent CNN

The work in [15] proposed a CNN of 6 which was later adjusted to 8 in [16]. The work in [17] suggested that the transmission range be dynamic and adjusted at the beginning of every transmission and the modelling of the adaptive transmission strategy resulted in each node having an optimal CNN of 3. CNN values of 8 and 6 were also proposed in [18]. It must be noted that these constant CNN values were derived for the optimization of packet forwarding strategies and that network connectivity was not explicitly considered.

Works that have taken network connectivity into account have also produced size-independent CNN values. In particular the work in [19] shows that the CNN converges to 9 as the network size approaches  $\infty$ . This result is shown to hold when the connectivity requirement is relaxed such that at least 95% of the nodes find themselves in the giant component of the original network.

### B. Network-Size-dependent CNN

Works taking network connectivity into account have also provided ranges within which the CNN can be found. The first such work was described in [20] which proposed that the CNN could be found within the range expressed in 1,

$$2.186 < CNN < 10.588 \quad (1)$$

Subsequent research has seen a continuous tightening of the upper- and lower-bounds, firstly in [21] and subsequently in [22]. The work done in [22] found that the network is asymptotically disconnected with probability 1 as  $n$  increases if each node is connected to less than  $0.074 \log(n)$  nearest neighbors and that the network is asymptotically connected with probability 1 as  $n$  increases if each node is connected to more than  $5.1774 \log(n)$  nearest neighbors, where  $n$  refers to the number of network nodes. This result was shown in [7] to be valid for square deployment regions containing both sparse and dense ad hoc networks. A further tightening of the upper-bound derived in [22] was obtained in [10] resulting in connectivity being assured with high probability if a maximum of  $2.718 \log(n)$  neighbors are maintained (shown in 2),

$$CNN < \alpha 2.718 \log(n) \quad (2)$$

where any real number  $\alpha > 1$  and  $n$  is the number of nodes in the backbone network and provided that  $n < 10000$ . Experiments conducted in [22] suggest that the critical value of  $\alpha$  may be close to 1. The connectivity bounds were further improved in [23] for both directed and undirected graphs, but it has been shown in [24] that the only way to guarantee full network connectivity in an ad hoc network is to ensure worst-case connectivity where each node is connected to every other network node. These size-dependent CNN-based connectivity strategies, except for the worst-case strategy, have been shown to create connected networks with increasing probability as the network size increases to infinity but to the best of our knowledge, their effectiveness in extending the network lifetime has not been previously evaluated.

### III. SIMULATION SETUP AND MEASUREMENT METHODOLOGY

Details of the simulation tools and measurement methodology employed are presented in this section.

#### A. Simulation Setup

The Atarraya simulation tool [25] was used to generate the network topologies used in the simulation. This tool was chosen for its emphasis on power control simulations and was modified to support I-WMNs. Various network sizes, ranging from 20 to 120 nodes were simulated with uniform node distribution. The simulated network areas were scaled to ensure a constant node density. The Atarraya simulator was also modified to record node positions and the resultant transceiver power levels assigned by the selected connectivity strategies.

In order to determine the resultant network lifetimes when employing the various connectivity strategies, the network topologies and the resultant transceiver power levels assigned to individual nodes by the connectivity strategies, were imported into the ns-2 simulation tool [26]. ns-2 (version 2.34) was found to possess better support for node energy models<sup>1</sup> and Application Layer traffic thus allowing for the analysis of

<sup>1</sup>The default energy model employed in ns-2 is used in this study.

TABLE I  
SIMULATION DETAILS

Simulation Time	100 seconds
Network Size	20–120 nodes
Network Area	300m x 300m – 1000m x 1000m
Routing Protocol	OLSR
Traffic type	CBR with 90% of nodes as traffic sources
Traffic rate	4 pkts per second with a max. of 1000 pkts
Max. transceiver range	100m
Initial energy	1.0 Joule
Transmit Energy	0.6W
Receive Energy	0.3W

the QoS achieved by the various connectivity strategies. The QoS data is, however, not reported in this paper. Additional simulation details can be found in Table I.

#### B. Measurement Methodology

Justifications for the reported experiments are provided in addition to highlighting how the various evaluation metrics were recorded.

1) *Network Connectivity*: This experiment aims to establish the effectiveness of the connectivity strategies in maintaining a connected backbone network. Network connectivity was determined by the number of entries in the routing tables of each backbone node. The OLSR routing protocol creates and maintains an entry for each possible destination node and the presence of an entry signifies that a route to the destination exists. Network connectivity is assured when all nodes can potentially communicate with all other network nodes, ensuring  $n^2 - n$  possible routes where  $n$  refers to the number of backbone nodes.

2) *Transceiver Power Savings*: The aim of this experiment is to determine the magnitude of transceiver power savings produced. Cumulative transceiver power savings are determined by the difference between the summation of the maximum power level of each node and the summation of the assigned power levels of all backbone nodes.

3) *Transceiver Power Assignment*: This experiment aims to establish whether a relationship exists between the assigned power level and a node's position in the network. The position of the imaginary network center is determined. Subsequently, the distance between every backbone node and the network center is calculated and plotted against the assigned transceiver power level.

4) *Network Lifetime*: The aim of this experiment is to determine whether the transceiver power savings that have been achieved will result in an extension to the network lifetime. Network lifetime is defined as the elapsed duration until the first node exhausts its energy supply. The number of alive nodes is determined and plotted against the simulated time.

### IV. PERFORMANCE EVALUATION

The connectivity strategies defined by (Xue, Kumar) [22], (Wan, Yi) [10] and Blough [19] have been chosen for evaluation. The Xue, Kumar and Wan, Yi strategies exemplify

TABLE II  
NETWORK CONNECTIVITY

Network Size	Src-Dest Pairs (Max. Power)	Src-Dest Pairs (Xue, Kumar)	Src-Dest Pairs (Wan, Yi)	Src-Dest Pairs (Blough)
20	380	380	380	380
40	1560	1560	1543	1560
60	3540	3539	3501	3527
80	6320	6317	6273	6303
100	9900	9898	9847	9882
120	14280	14277	14196	14256

adaptive CNNs whilst the Blough strategy is representative of a fixed CNN strategy.

#### A. Network Connectivity

The network is fully connected at max. transceiver power, where the number of source-destination (src-dest) pairs is  $n^2 - n$ . Table II shows that the Xue, Kumar strategy was best able to maintain network connectivity as there was little observed difference in the number of available src-dest pairs for all network sizes. The Xue, Kumar strategy benefits from maintaining an adaptive CNN that is based on the size of the backbone network.

The Wan, Yi strategy also maintains an adaptive CNN but the CNN is approximately half that maintained by the Xue, Kumar strategy. Thus, the Wan, Yi strategy does not provide the same degree of connectivity.

The Blough strategy maintains a fixed CNN for a large range of network sizes. This strategy is able to maintain full network connectivity for the smaller network sizes, but does not perform as well as the Xue, Kumar strategy due to a lower CNN being maintained at the higher network sizes.

#### B. Transceiver Power Savings

As shown in Figure 2, all 3 connectivity strategies being evaluated were found to produce cumulative transceiver power savings. The extent of the transceiver power savings produced can be found in Table III. Table III shows that the power savings produced by the adaptive CNN strategies (Wan, Yi and Xue, Kumar) remained fairly constant as the network size increased whilst the fixed-CNN Blough strategy produced increases in power savings. The Wan, Yi strategy produced the greatest magnitude of transceiver power savings due to the lower CNN being maintained. The Xue, Kumar strategy outperformed the Blough strategy at the smaller network sizes but the reverse began to occur at the largest network size. This particular phenomenon is explained by the CNNs required to be maintained by the Xue, Kumar and Blough strategies respectively. At a network size in the range [20 : 100] the CNN maintained by the Xue, Kumar strategy is less than that required by the Blough strategy. At greater network sizes, due to the fixed nature of the Blough strategy, the adaptive Xue, Kumar strategy maintains a greater CNN thus increasing the transmission range assigned to the nodes, thereby limiting the magnitude of transceiver power savings produced.

TABLE III  
PERCENTAGE POWER SAVINGS ACHIEVED

Network Size	Xue, Kumar	Wan, Yi	Blough
20	33.5	48.4	11.7
40	34.3	46.6	17.3
60	35.4	45.7	21.2
80	33.3	46.5	26.1
100	34.3	44.4	33.5
120	34.1	47.9	38.7

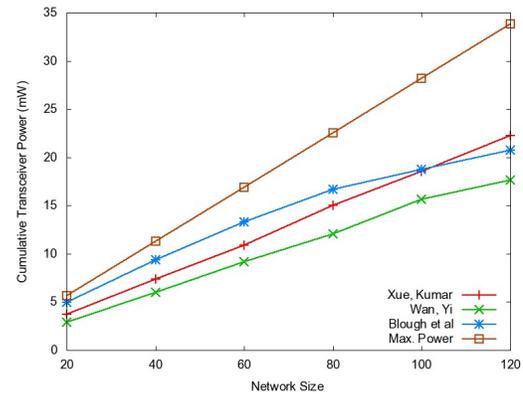


Fig. 2. Cumulative Transceiver Power Output

#### C. Transceiver Power Assignment

A closer look at the cumulative transceiver power savings produced by all three connectivity strategies reveals an interesting trend. Figure 3 depicts the relationship between the transceiver power levels assigned to individual nodes and the distance of these nodes from the (imaginary) network center.

The evaluation has found that the nodes closer to the network center produced significantly greater transceiver power savings than nodes at the network edge. This phenomenon was observed for all the connectivity strategies and across all the network sizes under evaluation. Edge nodes suffer from situations where fewer candidate neighbors exist and these neighbors are not evenly distributed within the node's transmission range but are rather loosely concentrated in a particular direction.

The correlation between the node position relative to the network center and the resultant transceiver power output is not exclusive to uniformly distributed nodes. This phenomenon has been previously observed in a clustered environment [27] as well as in a WMN testbed with arbitrary node distribution [11].

#### D. Network Lifetime

Figure 4 depicts the impact of the selected connectivity strategies on the network lifetime of a 120-node network and it can be seen that the application of the connectivity strategies does not result in extensions to the network lifetime. All three connectivity strategies result in the first node failing before the corresponding first node failure in the Max. Power scenario.

The reduction in network lifetime attributed to the three connectivity strategies can be explained by the effect that

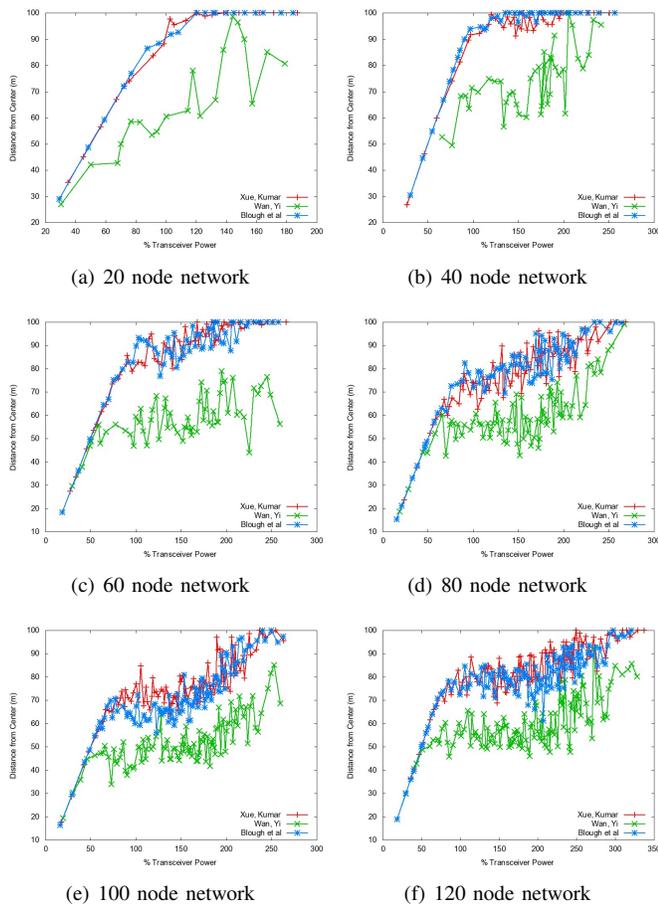


Fig. 3. Transceiver Power Output vs. Distance from Network Center

the reduction in node transceiver powers has on the resultant path lengths between source and destination nodes. Lower transceiver powers require data to traverse more hops to reach the intended destination. Energy is spent, at each intermediate node (hop), during packet reception and re-transmission thus negating the effect of transceiver power savings. In particular, nodes that still maintain maximum transceiver power outputs undertake greater packet forwarding responsibilities, thus depleting their energy sources at faster rates than before.

The three connectivity strategies do however lengthen the duration until the last node depletes its energy supply. This situation occurs since nodes with lower transceiver powers are overlooked as intermediaries for packet forwarding if nodes with higher transceiver powers are available and the nodes with lower transceiver powers conserve their energy sources for longer durations thus realizing lifetime extension gains. The Wan, Yi strategy produces the greatest number of alive nodes at the end of the simulated time due to the smallest CNN being maintained whilst the Xue, Kumar strategy produced the least number of alive nodes due to its maintenance of the highest CNN.

It must however be noted that the reduction in network lifetime caused by the evaluated connectivity strategies is not yet determined to be a universal consequence of the application

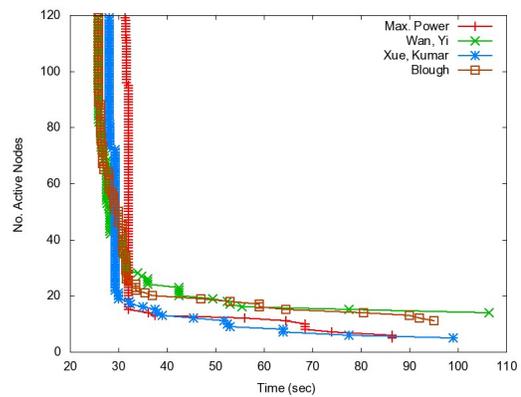


Fig. 4. Network Lifetime Achieved by the Connectivity Strategies

of these strategies as several factors prove influential. Network lifetime is highly dependent upon, amongst others:

- the resultant transceiver power assignment which is dictated by the network connectivity strategy being employed
- the energy consumption model being used
- the routing protocol and routing metric being utilized
- the prevalent traffic conditions

## V. CONCLUSION

I-WMN deployments are being employed to bridge the digital divide in rural areas around the world. The rural African context requires energy-efficient I-WMNs since it is highly likely that the nodes comprising these networks will be battery-powered.

The key requirement for the I-WMN backbone is the maintenance of network connectivity and several connectivity strategies exist in the literature. In this paper, three CNN-based connectivity strategies have been evaluated via simulation to determine their ability to produce transceiver power savings.

The evaluation has indicated that cumulative transceiver power savings in excess of 10% can be attained and that the ability of a node to produce these transceiver power savings is dependent upon the position of the node relative to the (imaginary) center of the network. Despite the achievement of cumulative transceiver power savings, corresponding extensions of network lifetime were not achieved. On the contrary, decreases in network lifetime were recorded due to the overburdening of nodes that maintained the use of the maximum transceiver power and the longer path lengths that are a side-effect of transceiver power reductions.

## REFERENCES

- [1] W. Allen, A. Martin, and A. Rangarajan, "Designing and Deploying a rural ad hoc community Mesh Network Testbed," Proc. IEEE Conference on Local Computer Networks, Nov. 2005, pp. 740-743.
- [2] H. Lundgren, K. Ramachandran, E. Belding-Royer, K. Almeroth, M. Benny, A. Hewatt, A. Touma, and A. Jardosh, "Experiences from the Design, Deployment and Usage of the ucsh meshnet Testbed," IEEE Wireless Communications, vol. 13, no. 2, pp. 18-29, 2006.

- [3] J. Camp, J. Robinson, C. Steger, and E. Knightly, "Measurement Driven Deployment of a Two-Tier Urban Mesh Access Network," Proc. 4th International Conference on Mobile Systems, Applications and Services, Jun. 2006, pp. 96–109.
- [4] I. F. Akyildiz and X. Wang, "Wireless Mesh Networks," Wiley: Chichester, 2009.
- [5] F. O. Aron, T. O. Olwal, A. Kurien, and M. O. Odhiambo, "Energy Efficient Topology Control Algorithm for Wireless Mesh Networks," Proc. International Wireless Communications and Mobile Computing Conference, Aug. 2008, pp. 135–140.
- [6] P. Gupta and P. R. Kumar, "Critical Power for Asymptotic Connectivity," Proc. IEEE Conference on Decision and Control, Dec. 1998, pp. 1106–1110.
- [7] P. Santi and D. M. Blough, "The Critical Transmitting Range for Connectivity in Sparse Wireless Ad Hoc Networks," IEEE Transactions on Mobile Computing, vol. 2, no. 1, pp. 25–39, 2003.
- [8] M. D. Penrose, "On  $k$ -connectivity for a Geometric Random Graph," Random Structures and Algorithms, vol. 15, no. 2, pp. 145–164, 1999.
- [9] C. Bettstetter, "On the Minimum Node Degree and Connectivity of a Wireless Multihop Network," Proc. ACM International Symposium on Mobile Ad Hoc Networking and Computing, 2002, pp. 80–91.
- [10] P. Wan and C. Yi, "Asymptotic critical Transmission Radius and critical Neighbor Number for  $k$ -connectivity in wireless ad hoc networks," Proc. ACM International Symposium on Mobile Ad Hoc Networking and Computing, May 2004, pp. 1–8.
- [11] P. Mudali, T. C. Nyandeni, N. Ntlatlapa, and M. O. Adigun, "A Test-Bed Evaluation of Connectivity Strategies for Infrastructure Wireless Mesh Networks," Proc. Southern African Telecommunications and Network Applications Conference, Sep. 2010. Available at: <http://www.satnac.org.za/proceedings/2010/engineering.htm>
- [12] X. Wang and A. O. Lim, "IEEE 802.11s Wireless Mesh Networks: Frameworks and Challenges," Ad Hoc Networks, vol. 6, no. 6, pp. 970–984, 2008.
- [13] V. Kawadia and P. R. Kumar, "Principles and Protocols for Power Control in wireless ad hoc networks," IEEE Journal on Selected Areas in Communications, vol. 23, no. 1, pp. 76–88, 2005.
- [14] P. Valera, P. W. Q. Lee, Y. F. Wong, W. K. G. Seah, H. P. Tan, and H. Ju, "An experimental study on Connectivity and Topology Control in Real Multi-hop Networks," Proc. International Wireless Internet Conference, Nov. 2008, pp. 1:1–1:8.
- [15] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," Proc. National Telecommunications Conference, Dec. 1978, pp. 4.3.1–4.3.5.
- [16] H. Takagi and L. Kleinrock, "Optimal Transmission Ranges for Randomly Distributed Packet Radio Terminals," IEEE Transactions on Communications, vol. 32, no. 2, pp. 246–257, 1984.
- [17] B. Hajek, "Adaptive Transmission Strategies and Routing in Mobile Radio Networks," Proc. Seventeenth Annual Conference on Information Sciences and Systems, Mar. 1983, pp. 373–378.
- [18] T. Hou and V. O. K. Li, "Transmission Range control in Multihop Packet Radio Networks," IEEE Transactions on Communications, vol. 34, no. 1, pp. 38–44, 1986.
- [19] D. Blough, M. Leoncini, G. Resta, and P. Santi, "The  $k$ -neighbors Approach to Physical Degree Bounded and Symmetric Topology Control in ad hoc networks," Istituto di Informatica e Telematica, Pisa, Technical Report IIT-TR-142003, 2003.
- [20] P. Hall, "On continuum percolation," The Annals of Probability, vol. 13, no. 4, pp. 1250–1266, 1985.
- [21] T. Philips, S. Panwar, and A. Tantawi, "Connectivity properties of a packet radio network model," IEEE Transactions on Information Theory, vol. 35, no. 5, pp. 770–777, 1989.
- [22] F. Xue and P. Kumar, "The Number of Neighbors Needed for Connectivity of Wireless Networks," Wireless Networks, vol. 10, no. 2, pp. 169–181, 2004.
- [23] P. Balister, B. Bollobas, A. Sarkar, and M. Walters, "Connectivity of random  $k$ -nearest-neighbour graphs," Advances in Applied Probability, vol. 37, no. 1, pp. 1–24, 2005.
- [24] P. Santi, "Topology Control in Wireless Ad Hoc and Sensor Networks," Wiley: Chichester, 2005.
- [25] M. A. Labrador and P. M. Wightman, "Topology Control in Wireless Sensor Networks: With a Companion Simulation Tool for Teaching and Research," Springer, 2009.
- [26] Network Simulator - ns-2, Available at: <http://www.isi.edu/nsnam/ns/> // Last accessed: 13 Feb. 2011.
- [27] P. Mudali, N. Ntlatlapa, and M. O. Adigun, "Evaluating Transceiver Power Savings Produced by Topology Control in Rural Wireless Mesh Networks," Proc. World Wireless Research Forum 25, Nov. 2010.

# Spectrum Access during Cognitive Radio Mobiles' Handoff

Emna Trigui, Moez Esseghir, Leila Merghem Boulahia  
 Autonomic Networking Environment, ICD/ERA, CNRS UMR STMR 6279  
 University of Technology of Troyes (UTT)  
 12, rue Marie Curie, 10010 Troyes Cedex, France  
 {emna.trigui, moez.esseghir, leila.merghem\_boulahia} @utt.fr

**Abstract** — Cognitive radio is an emergent technology in wireless networks that aims to improve spectrum's use by allowing opportunist access. Recent research activities related to cognitive radio consider static terminals, neglecting the impact of user's mobility. In this paper, we are interested in the concept of mobility in cognitive radio networks. Thus, we propose an algorithm of spectrum handoff for mobile cognitive radio users. Our algorithm presents a decentralized approach using multi-agent systems. Each terminal is managed by an agent that enables it to negotiate and cooperate with neighboring users. During the handoff, agents recover sensing information then choose appropriate band to switch if spectrum handoff is necessary. Finally, agents use negotiation and cooperation methods to insure a more efficient spectrum sharing.

**Keywords**- Cognitive radio, Dynamic spectrum access, Mobility, Spectrum sharing, Spectrum handoff.

## I. INTRODUCTION

Recent evolution of wireless technologies is creating greater demand in terms of spectrum resources. To overcome this problem, researchers have used a new paradigm known as cognitive radio (CR) [1] that senses the nearby spectrum and tries to utilize it opportunistically. Indeed, there are unused spectrum portions (spectrum holes), which can be utilized in order to increase the number of users and to better distribute the available resources [2].

In CR networks [3], there are two types of users: licensed or primary users (PUs), and unlicensed or secondary users (SUs). PUs can access the wireless network resources according to their license. SUs are equipped with CR capabilities to opportunistically access the spectrum.

Researches on CR networks are mainly focused on detection and allocation of available spectrum resources, leaving terminal mobility issues mostly unexplored. In fact, mobility makes spectrum management problem more complex notably because of handoff management.

Due to the above facts, our work focuses on mobility management and spectrum sharing for mobile CR terminals during the handover. This latter occurs when mobile terminal switches from one network point of attachment to

another one [4]. In CR context, it can come with spectrum handoff, which is the change of the used spectrum frequency. The goal of managing handoff is mainly to keep alive ongoing sessions with the requested quality of service (QoS). In the rest of the paper "spectrum handoff" [5] and "spectrum mobility" [6] will be used interchangeably.

The remainder of this paper is organized as follows. Section II presents related works on spectrum and mobility management in CR networks. Next, Section III describes the considered scenario. Section IV details our proposed solution, depicts the suggested handoff algorithm for mobile CR terminals and gives an illustration of the resulting dynamic spectrum distribution. Finally, Section V concludes the paper.

## II. RELATED WORKS

Since the last decade, a large amount of literature already exists on CR. Consequently, several dynamic spectrum access approaches have been proposed [6], [7], [8], mostly addressing the issues of spectrum sensing and spectrum allocation while there is no much research effort to address the problems of spectrum handoff.

The spectrum sensing [7], [9] is a fundamental step to detect the presence of PU in CR networks. There are different ways in, which CR users are able to perform spectrum sensing. These ways are classified into two categories: Non cooperative and cooperative spectrum sensing. *Non-cooperative spectrum sensing* occurs when a CR acts on its own and self-configures according to the signals it can detect. *Cooperative spectrum sensing* uses a central station to receive reports of signals from a variety of radio users. CR cooperation reduces problems of interference where a CR user cannot hear a primary user because of issues such as shading from the PU. Besides, channel miss detection phase a problem that may occur during sensing phase and it depends necessarily on the selected sensing algorithm. In our work, we do not address spectrum sensing issues as our research is focused on spectrum sharing and mobility.

Dynamic spectrum allocation and sharing [10] exploit temporal and spatial traffic statistics to share more

efficiently the underutilized spectrum. Game-theory approach [11] is the mostly used for spectrum sharing. Bargaining, auctions and multi-agent systems are also increasingly used.

Broadly, research works on CR have been concentrated on the case of static networks without taking into account mobility and handover aspects. Although the mobility-based handoff mechanisms have been extensively investigated in wireless, cellular and heterogeneous networks [4], [12], [13], it is still an open research issue for CR networks.

Nevertheless, some works like [14], [15] have used CR concept to improve mobility management in traditional cellular networks. For example in [15], the proposed approach enables changes in the base station's parameters to meet the new services requirements in modern wireless cellular systems. These changes are performed using agents that manage cells via negotiation, learning, reasoning, and identification strategies. The principal aim of this solution is to reduce interference, HO delay and blocking probability. However, it is only suitable for traditional cellular networks with a centralized management system.

Our objective in this paper is to achieve an optimal dynamic spectrum sharing and an efficient spectrum mobility management by considering the handover in the CR networks. Therefore, next, we propose a typical scenario and a spectrum handoff algorithm for mobile CR users.

### III. SCENARIO

Most scenarios previously discussed in the literature in CR context are restricted to the spectrum management between fixed nodes (absence of HO). In this work, we consider the scenario of a mobile node, as shown in Fig. 1.

Mobile cognitive radio terminal (MCT) moves from location A towards location B through a set of areas, where the space is supposed to be distributed in zones each having its own characteristics (frequencies, number of users, etc.).

The MCT uses initially a spectrum portion already assigned to it in the departure zone. This allocation was made after detection and decision phases according to the MCT resources requirements. Assigned spectrum portion can be shared with PU or SUs.

When moving from one zone to another one, the MCT can meet the following scenarios:

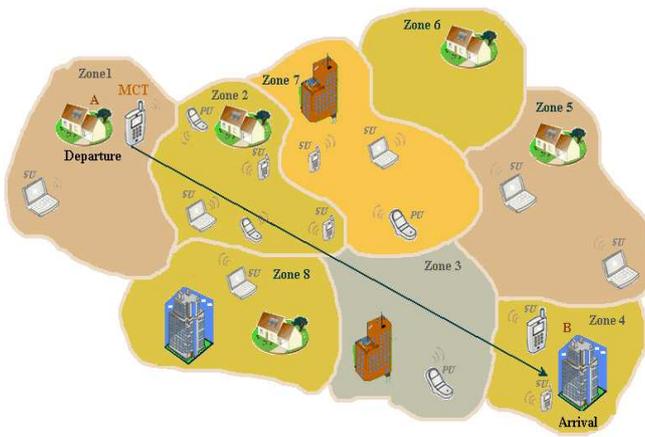


Figure 1. Scenario of mobile cognitive radio node

- The spectrum portion used by the MCT no longer guarantees the QoS required in the new zone.
- The spectrum portion is partially occupied and negotiation/ cooperation are possible. This happens in the following cases:
  - The band is occupied by a PU.
  - The band is occupied by one or several SUs.
  - PU and a set of SUs coexist in this band.
- Other spectrum portions ensuring a better QoS are available in the new zone.

### IV. PROPOSED SOLUTION

In this section, we first present the basic behavior of the MCT. Then we propose an algorithm for dynamic spectrum sharing and handoff for the MCT when switching zones. Finally, we give an example of the spectrum distribution to illustrate our solution.

#### A. The basic behavior of mobile cognitive radio terminal

The state diagram in Fig. 2 details the MCT behavior within the same zone.

During the detection phase, the CR node senses its surrounding radio environment to find spectrum holes. This detection process continues till discovering available spectrum portion (at least one). Then, the CR node starts the decision phase to choose the appropriate band based on the requested QoS by the running applications on MCT. During this phase, the CR node observes and characterizes bands and chooses on FIFO bases the first one that satisfies its applications' needs. By using FIFO strategy when selecting spectrum portion, we aim to minimize the decision process duration. If none of free channels is suitable (for instance, insufficient bandwidth or QoS not guaranteed), the CR node starts again the spectrum sensing.

Once the appropriate channel is selected, the CR node may share it with other users. Two possible scenarios can

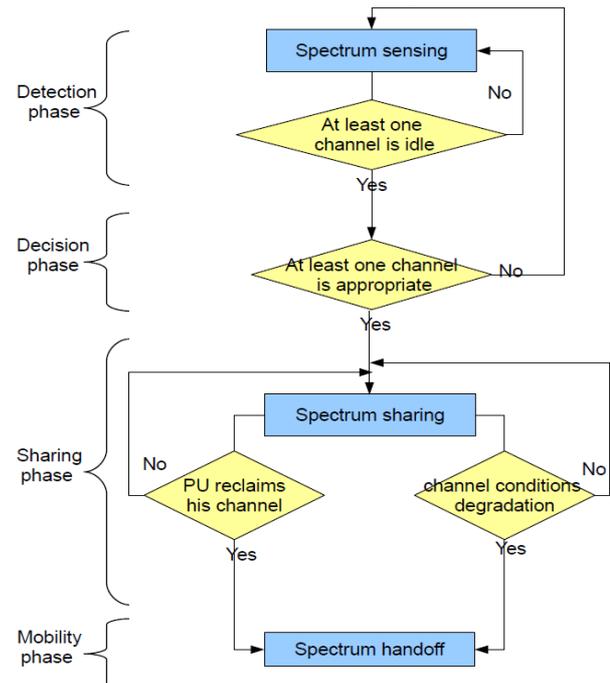


Figure 2. State diagram of cognitive radio node behavior

lead to spectrum handoff: (1) a coexisting PU would reclaim its radio spectrum resources and thus, the MCT has to move immediately to another available portion since a PU has always the priority; (2) QoS degradation due to interference can also lead to spectrum mobility.

### B. Spectrum handoff algorithm for MCT

The node mobility imposes new challenges, which include the topology change and breaking the continuity of services during HO. Hence, the MCT requires a specific behavior when switching to new zones (section 3 details eventual scenarios). It should know in prior new spectrum conditions and should react towards new circumstances. Different mechanisms to ensure continuity of service and efficient spectrum management are needed. For these reasons, we propose an algorithm of spectrum access that is executed by the MCT when Handoff occurs (algorithm 1).

In the rest of the paper, we use the following notations:

- $QoS(MCT)$  : Quality of service required by MCT.
- $spectrum_{i(zone\ j)}$  : Current spectrum portion  $i$  used by the MCT in zone  $j$ .
- $spectrum_{i(zone\ j+1)}$  : Spectrum portion  $i$  that should be occupied by the MCT in the new zone  $j+1$ .

We assume that the MCT activates its handoff decision algorithm as soon as it comes close to a new zone in order to anticipate a possible handover. According to the information recorded about its new environment, the MCT updates its knowledge base with, among others, spectrum conditions.

Our proposed solution uses multi-agent systems where each CR node is equipped with an agent. Every Agent is autonomous and manages its spectrum resources needs in a decentralized way. It interacts with other users to insure an effective spectrum management. The MCT will then negotiate with the PU and cooperate with SUs.

Negotiation with PU means a discussion indented to produce a contract between the MCT and the PU. The PU can accept or refuse to share its licensed spectrum portion with the requesting SU.

Cooperation means that MCT and SUs collaborate together to share the spectrum. SUs will check whether the new distribution of spectrum including new arriving MCT still satisfies their network access needs.

If in the new zone, the current used spectrum is totally occupied or if it does not guarantee required QoS then the MCT has to execute a spectrum handoff. Otherwise, if the spectrum is totally idle, it keeps using it.

On the other hand, if the current used spectrum is partially occupied in the new zone, it is necessary to check, which kind of existing users (PU or SU) are there. The MCT can distinguish between a PU and an SU either by using a particular sensing algorithm [16] or by exchanging (or broadcasting) messages with new zone  $s'$  users.

If a PU is present then the MCT must request its approval for spectrum sharing. In case of disagreement, MCT has to look for another portion. When the spectrum is partially used only by SUs then they should verify the sharing feasibility with the MCT.

---

### Algorithm 1: Spectrum Mobility management for MCT during Handoff

---

```

If ( $spectrum_{i(zone\ j+1)}$  = fully occupied)
  Or ( $spectrum_{i(zone\ j+1)}$  does not guarantee QoS(MCT))
  Then  $spectrum_{i(zone\ j+1)} \leftarrow$  Choose_new_portion() // Spectrum Handoff
  Else
    If ( $spectrum_{i(zone\ j+1)}$  is totally idle) // Band is totally free
      Then
         $spectrum_{i(zone\ j+1)} \leftarrow$   $spectrum_{i(zone\ j)}$ 
        // Continue to use the same spectrum
      Else
        If ( $PU \in spectre_{i(zone\ j+1)}$ )
          // If a PU is using a portion of the spectrum
          Then
            If negotiation (PU, MCT) = acceptance
              Then spectrum_sharing(PU, MCT)
              Else  $spectrum_{i(zone\ j+1)} \leftarrow$  Choose_new_portion()
            End If
          Else // If there are one or several SUs in the spectrum
            If cooperation (SUs, MCT) = acceptable
              // free spectrum portion is sufficient
              Then spectrum_sharing (MCT, SUs)
              Else  $spectrum_{i(zone\ j+1)} \leftarrow$  Choose_new_portion()
            End If
          End If
        End If
      End If
    End If
  End If

```

---

**Choose\_new\_portion()** function makes the same tests for another spectrum hole detected by the sensing algorithm in order to choose a new portion to be utilized by the MCT in the new zone. Returns a portion that meets MCT access needs without disrupting other available users (PU or SU) in the same zone.

---

Through sensing, the MCT can find more than one appropriate spectrum portions in the new zone. However, our algorithm avoids MCT unnecessary spectrum handoff assuming that the current used spectrum is still available in the new visited zone and provides QoS requirements.

In addition, our algorithm insures maintaining PU QoS. Indeed, TCM cannot use spectrum unless it receives PU acceptance. PU agreement depends on many factors like PU QoS, disposal of unutilized appropriate spectrum resources, price, etc. In other words, PU would not accept to share its spectrum with a TCM if this will degrade PU QoS.

### C. Example of spectrum distribution

To better understand our proposed algorithm (algorithm 1) an example of spectrum distribution is depicted in Fig. 3.

Initially, the MCT is located in zone1 and uses spectrum portion 1. When starting the handover from zone 1 to zone 2, the MCT discovers that the spectrum portion 1 in use is unavailable in zone 2. The spectrum portion 1 is either fully used by PUs and/ or SUs or no longer guarantees the required QoS. Thus, the MCT switches to an idle spectrum

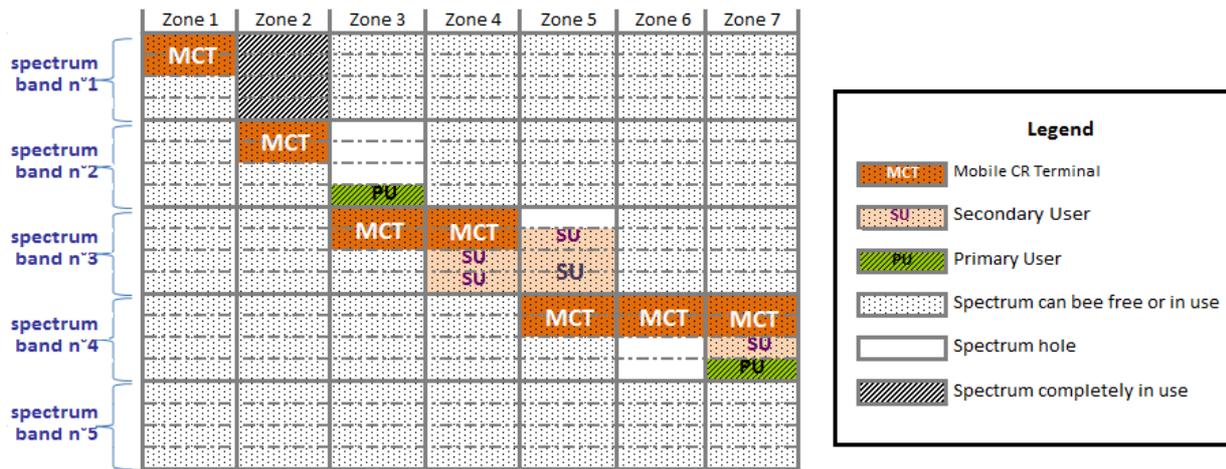


Figure 3. Example of dynamic spectrum distribution

band, the portion 2. When moving from zone 2 to zone 3, the MCT meets a PU, which is partially using the spectrum portion 2 in zone 3. The MCT starts a negotiation with the PU for spectrum sharing. In our selected scenario, the PU disagrees sharing and the MCT runs another spectrum handoff to acquire portion 3 (idle in zone 3).

During the handover from zone 3 to zone 4, the MCT detects the presence of SUs. The MCT cooperates with them for spectrum sharing. As the available amount of band meets the MCT request without affecting the other SUs, the MCT continues to use spectrum portion 3 in zone 4. When arriving to zone 5, remaining band is not sufficient bringing the MCT to a spectrum handoff by switching to portion 4. MCT continues to operate in the same band within zone 6 as it is totally idle and in zone 7 since the PU accepts here to share its spectrum portion.

V. CONCLUSION

In this paper, we have addressed the spectrum management issues in CR networks and showed the limitations considering the mobility. We proposed then a spectrum management algorithm that combines terminals mobility with negotiations for dynamic spectrum sharing in CR networks. We have then illustrated our proposal through an example of spectrum distribution. Our solution relies on decentralized and multi-agent approach and uses negotiation and cooperation mechanisms in order to ensure dynamic and efficient spectrum access.

As future works, we will choose well-defined negotiation algorithms and define channels selection parameters more precisely. We will also study (in more details) agent’s interactions to refine our decision process. We will subsequently evaluate the performances of our solution through simulations and mathematical modeling.

ACKNOWLEDGMENT

This work is partly supported by the Ministry of Higher Education and Research of France.

REFERENCES

[1] A. Fette, “Cognitive radio technology”, Elsevier, USA, 2006  
 [2] Fcc, Et Docket N° 03-222 Notice of proposed rulemaking and order, December 2003

[3] J. Mitola, “Cognitive radio: An integrated agent architecture for software defined radio”, PhD thesis, Royal Institute of Technology (KTH), Sweden, May 2000  
 [4] B. Abuhaija and K. Al-Begain, , “An algorithm for multi service continuity in multi radio access cellular networks”, 12th International Conference on Computer Modelling and Simulation, Cambridge, 24-26 Mars 2010, vol.9, p 558–566  
 [5] A. Lertsinsrubtavee, N. Malouch, and S. Fdida, “Spectrum Handoff Strategies for Multiple Channels Cognitive Radio Network”, ACM CoNEXT Conf., Philadelphia, 2010, USA  
 [6] F. Akyildiz, W-Y. Lee, M.C. Vuran, and S. Mohanty, “Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey”, International Journal of Computer and Telecommunication Networking, vol. 50, n° 13, September 2006, pp. 2127–2159  
 [7] S. Haykin, “Cognitive radio: Brain-empowered wireless communications”, IEEE Journal on Selected Areas in Communications, vol. 23, February 2005, pp. 201–220  
 [8] Q. Zhao and B.M. Sadler, “A survey of dynamic spectrum access: signal processing, networking, and regulatory policy”, IEEE Signal Processing Mag., vol. 4, n°3, 2007, pp.79-89  
 [9] T. Ycek and H. Arslan, “A survey of spectrum sensing algorithms for cognitive radio applications”, IEEE Comm. Surveys & Tutorials, vol. 11, n°1, pp. 116–160, 2009  
 [10] U. Mir, L. Merghem-Boulahia, and D. Gaïti, “On using multiagent systems for spectrum sharing in cognitive radios networks”. Book chapter, accepted in Formal and Practical Aspects of Autonomic Comp. and Networking: Specification, Development and Verification, IGI Global, 2010  
 [11] Z. Ji and K.J.R. Liu, “Dynamic spectrum sharing: a dynamic spectrum sharing: a game theoretical overview”, IEEE Communications Magazine, 2007, pp. 88–94  
 [12] F. Akyildiz, W-Y. Lee, M.C. Vuran, and S. Mohanty, “Mobility management in next-generation wireless systems”, IEEE, vol. 87, 1999, pp. 1347–1384  
 [13] A. Ahmed, “Continuity of Service in Heterogeneous Wireless Networks with Multi-agent Systems”, PhD thesis, University of Technology of Troyes, October 2010  
 [14] E.V. Belmega, S. Lasaulce, and M. Debbah, “Decentralized handovers in cellular networks with cognitive terminals”, 3rd International Symposium on Communication, Control, and Signal Processing (ISCCSP-2008), Malta, 2008, pp. 219–224  
 [15] J. Raiyn, “Toward cognitive radio handover management”, Annual IEEE 19<sup>th</sup> International Symposium on Personal, Indoor and Mobile Radio Communication (PIMRC ‘08), Cannes, France, 15-18 September 2008, pp. 1–5  
 [16] T. Newman and T. Clancy, “Security threats to cognitive radio signal classifiers”, in Virginia Tech Wireless Personal Communications Symposium, June 2009.

# Efficiently Using Bitmap Floorplans for Indoor Navigation on Mobile Phones

Martin Werner

*Mobile and Distributed Systems Group  
Ludwig-Maximilians-Universität München, Germany  
Email: martin.werner@ifi.lmu.de*

**Abstract**—Pedestrian navigation applications and especially indoor navigation applications need to describe a way in a form that is easy to understand, remember and follow. While in the case of outdoor vehicle navigation the distance in meter together with a turn hint is a good description such a description is merely useless for the indoor area. Pedestrians do not have a good understanding of distances and turns are not always well-defined in the indoor area. Landmarks seem to bring the solution to this problem. With this paper we define an algorithm to identify relevant landmarks and a general algorithm enhancing the visualisation quality of an indoor way using bitmap floorplans. This opens the door for automatically generated waypoint graphs with bad visualisation properties to be used for a multimedia indoor navigation application. Furthermore the algorithms are implemented on a smartphone and results on the final performance are given. The main contribution of this paper is the reformulation of some indoor navigation tasks as image processing tasks.

**Keywords**-Navigation; Image Processing;

## I. INTRODUCTION

In the last decade, many new multimedia services have been designed. These include numerous applications from the field of location based services. Furthermore the ongoing trend towards mobile computing and the immense development in the field of cellular phones leads to more and more context-information that can actually be used. While it is relatively easy to provide location based services for the outside area it is relatively difficult to do the same for the indoor area. For the indoor area there are many problems, that are solved outdoors. The first problem is the availability of digital map data. While for most outdoor location based services the map functionality offered by major Internet companies (e.g., Google Maps) is enough, there is no comparable service for the indoor area. There are several reasons behind: The complexity of indoor maps would be much higher than outdoors (different floors, differences in the treatment of free space). Moreover the contents of indoor maps is often protected by intellectual property rights of architects. Finally, there is no cheap positioning technology available, which makes indoor location based services attractive.

Nevertheless indoor navigation is a very promising technology. People want to have technological support for indoor orientation, which is comparable to the outdoor situation. Moreover there are several buildings that are either very

complex (industrial buildings) or not known to the majority of guests (foreign airports, exhibitions) or not known exactly enough for safety services (firefighter, police, etc.). The most difficult task for indoor navigation is of course the positioning of the users. But in the last decade many promising technologies are under development, which will solve this problem completely in the near future. These include classical signal-strength methods based on Wireless LAN [1], [2] or more specialised approaches based on UWB [3] as well as advanced statistical treatments of this measurement data [4], [5].

Cellular phones are becoming more and more powerful in terms of calculational power as well as in terms of sensing capabilities. A complete integration of all data that a modern cellular phone can sense will lead to indoor positioning with acceptable accuracy in the future.

Often indoor navigation has been implemented in a relatively small area due to the focus on positioning technology. In such small settings it is not really important to have efficient algorithms for several tasks. However large-scale indoor navigation is an upcoming topic [6].

With this paper we want to show how the reformulation of some subproblems of indoor navigation into image processing tasks is possible and that modern smartphones are able to run this type of algorithms in acceptable time. Note that the results of these algorithms are useful not only for indoor navigation but also for other context-aware indoor location based services and ubiquitous computing trends.

We are focussing on two very complex tasks an indoor navigation system has to accomplish. The first task is the transformation of a shortest way into an augmented form, which is easy to understand, remember and follow. Essentially we are talking about how to transfer a line-strip into a sequence of textual instructions and about how to make several decision points in the navigation solution easy to remember. In this paper we present a parallelisable image processing algorithm, which finds all landmarks that can be used to augment a given way. A good treatment of how landmarks can be used to aid indoor navigation is [7]. We are not talking about how to actually use this landmark information subset for textual instruction generation, because this would go beyond the scope of this paper and is really indoor navigation specific, while the other algorithms have a more general scope.

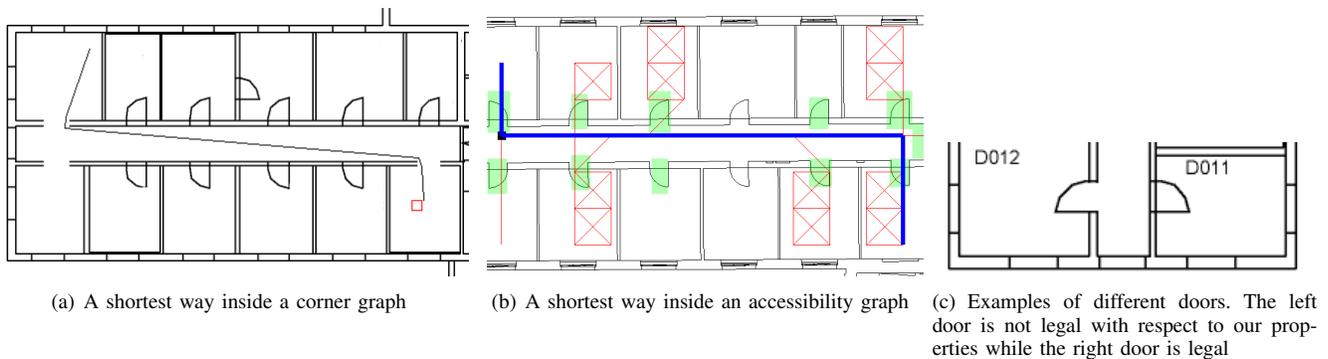


Figure 1. Examples of waypoint graphs and problems

The second task, which we accomplish by a sequence of image processing tasks, is to enhance the visualisability of a given way. Navigation in general is often based on a waypoint graph [8], [9]. A waypoint graph consists of a set of waypoints and an edge in between those waypoints, which are directly reachable from each other (usually on a straight line of walking). As the impact of the graph size on the performance of shortest path algorithms is high, people have studied algorithms generating small graphs. However the reduction of the number of waypoints in a graph leads to fewer ways inside the graph. Hence the shortest way inside the graph is not a way, that a human being would choose. One of the smallest and most efficient graphs is for example the corner graph. The corner graph is a graph, which contains all corners of the building (map) as vertices and an edge between two corners if and only if the direct line inbetween these two points is inside the building and free (walkable) space. The resulting ways tend to scrape along walls as can be seen in Figure 1(a). Another well-behaved form of a navigation graph consists of a subset of a grid of waypoints in navigation space defined by some properties (such as a minimal distance to a wall and being inside the building) and edges between waypoints if there is free space inbetween. A shortest way inside such a graph can contain some flaws as can be seen in Figure 1(b). In this example the relatively large grid size leads to the effect, that the chosen way is too near to the left wall. Note, that a finer grid does not help because a finer grid finally leads to a situation comparable to the corner graph. This type of problem is exactly what we want to remedy with the second algorithm in this paper.

When it comes to indoor navigation it is essential to fix a model of the surroundings. It is important to define exactly in terms of available data, what is meant when we are talking for instance about a room or a door. In the following chapter we give possible definitions for this in terms of images. Of course there are other views (especially concerning GIS databases), which can be more flexible, but our choice is made on the background that good GIS data does seldom

exist for the indoor area while a basic floorplan (possibly scanned from a building blueprint) is almost always available. Furthermore cellular phones have more difficulties with handling vector GIS-data and the associated queries than with handling and manipulating bitmaps. Moreover the standardisation of bitmap file formats allows for flexible and sustainable treatment of indoor navigation data.

In the following section, we fix an environmental model in terms of bitmap floorplans, in Section III we explain the problem of finding relevant landmarks and describe an algorithm solving this problem. In Section IV we describe an algorithm to enhance the visualizability of ways, which also clarifies turning points useful for textual description of the way. We then explain an implementation of this algorithm for mobile phones and give experimental results on the performance. In Section VI we finalize this paper with a conclusion.

## II. THE INDOOR ENVIRONMENTAL MODEL AND ITS ASSOCIATION WITH BITMAPS

For this paper we want to describe now exactly our environmental model and how it is setup with different bitmaps. Starting with a building (or site) in some reference system, we define the bitmap projection by first projecting the building information into an orthogonal coordinate system (if it is not given in an orthogonal coordinate system) and map the bounding box of the building to a bitmap by a choice of pixel size. In our experiments, pixel sizes of  $0.15\text{m} \times 0.15\text{m}$  seem to work fairly well.

For this bitmap floorplan we start with some definitions, which do not exactly resemble the definitions of common agreement.

*Definition 1:* An **area** is a subset of available pixels.

Note that an area need not be connected or otherwise have properties, which the term area describes in other contexts.

*Definition 2:* The **floodfill-closure** of a point  $(x, y)$  is the area inside the bitmap that would be filled by a usual 4-neighbour-floodfill operation at  $(x, y)$  with a colour that is not used elsewhere in the bitmap. The 4-neighbour-floodfill operation is defined to fill a pixel if it is of the same

colour as the start pixel and in this case continues with the neighbouring pixels above, left, right and below.

*Definition 3:* An **area** is called floodfill-connected if and only if it is the floodfill-closure of one (and hence any) of its points.

*Definition 4:* A **room** is a floodfill-connected area.

*Definition 5:* The **outside-space** is the floodfill-closure of the pixel coordinate (0,0).

*Definition 6:* A room  $R_1$  is **inside** another room  $R_2$  if and only if  $R_1$  is in the convex closure of  $R_2$ .

Of course this definition does not recover the usual term of a room is inside another room. But as it is unlikely that a room that is contained in another room in our sense is not reachable easily this discrepancy to the real world is not severe.

With these definition in place we formulate some properties that a floorplan should have for the application of our algorithms below:

- *Closed-Building-Property:* The building is closed by black lines and completely surrounded by white space. This essentially means, that the outside space resembles the usual definition of what is outside a building.
- *Doors-Are-Rooms-Property:* A door inside the building is drawn such that it itself is a room in the sense of Definition 4. An example of a legal door and an illegal door is given in Figure 1(c).
- *Walkable-Space-Is-Known:* There is an oracle telling us whether a straight line between two points lies in walkable space (which is defined to be the space where a human being can walk).

For the rest of the paper, we assume that we are given a bitmap - simply called floorplan in the sequel - where these properties hold.

### III. LANDMARK SEARCH

Pedestrian navigation results need a different presentation form as compared to vehicle navigation results. While for a vehicle navigation system based on GPS the typical errors of the positioning system do not have much influence on the identification of a turn and the orientation of the vehicle is known by the orientation of the street, this is not true for the indoor area. We usually have positioning systems with low accuracy and different possible turns within this accuracy. Typically the positioning errors do not allow the distinction of two doors, which are directly next to each other. In this situation, we want to augment a navigation solution with semantical information such that it is easy to remember and follow. For this the concept of a landmark is often used. Landmarks are objects in the surroundings having a local uniqueness and being eye-catching. All classical signs are landmarks in this sense. But even shops and plants can serve as landmarks. With landmark information the problems of coarse positioning can be reduced. If landmarks are drawn into a map in form of a pictogram (e.g., the logos of the

shops) the relation between the proposed way and those shops can be easily remembered and used for difficult way decisions. This role of a landmark is best explained by the following textual instruction, which can be generated if good landmark information is available: "Before the post office turn left. Then you will see a red sculpture in 300m distance." As you can see from this sentence, landmark information can be used to make explicit the position of a turn in relation to semantical information (as opposed to geometric information). Landmark information can also be used to enhance the confidence in having a good orientation as you can see from the second sentence.

The difficult task is now to reduce the set of landmarks (which is usually very big) to the set of landmarks that are visible from the way to facilitate more complex election algorithms for the actual integration of landmark information into visual and textual representations of the way.

The following algorithm is a good symbiosis of a search technology with a geometric enhancement technology. The results of this algorithms is an image containing a set of visible landmarks identified by a colour convention. It is possible to very quickly extract this information as a list of visible landmarks or to directly integrate the graphical result into the visualisation pipeline. One could for example highlight the visible space, draw a pictogram over visible landmark positions and so on.

#### A. Landmark Search By Image Processing

With the following algorithm we solve the problem of finding relevant landmarks out of a list of landmarks visible from within a way. Common algorithms to solve this problem are more or less searching for landmarks by checking whether a given landmark is visible. As there is no good geometric ordering of landmarks (i.e., reducing the search space by a distance limit will miss good landmarks in long rooms) it is not easy to do such a search efficiently. This type of search problem also shows up in other problems of ubiquitous and context-aware computing.

A landmark in the sense of the following algorithm consists of a pixel coordinate and a connected information (identification in a database, etc.). The input of the algorithm consists of

- A set of landmarks
- A floorplan
- A way (given as a list of points forming a line-strip)

The configurable parameters influencing this algorithms are

- The maximal viewing distance
- The length threshold used during tessellation of the way

1) *Step 1: Prepare Landmark Map:* The very first step is to overlay our floorplan with drawn landmark locations. Therefore we use a colour palette mapping a colour (that is not used in the floorplan) to the identification data. This mapping is symbolised in the following pseudo-codes by the function `landmark_to_colour(landmark l)`.

```

void draw_landmarks()
copy (floorplan, landmark_map)
for each landmark l in landmark_set{
putpixel(landmark_map,
         position,
         landmark_to_colour(l))
}

```

This step is of course general and the result can be cached for subsequent applications of this algorithm.

2) *Step 2: Tessellation of the Way:* As it is computationally very expensive to calculate the set of pixels, which are visible from a line-strip, we approximate this set of pixels by the set of pixels visible from the points of a tessellation of the line-strip. To obtain this tessellation, we keep inserting middlepoints between two subsequent points until the distance between all points is shorter than the *tessellation length*.

```

global tessellation_length
void tessellate(linestrip l)
for each segment s of l{
if (s.length() > tessellation_length)
{
    s.split()
    return tessellate(l)
}
}

```

3) *Step 3: Calculate the mask bitmap:* In this step we calculate a mask, which resembles the set of pixels visible from the way. This is done by preparing the mask bitmap to be of the same size as the floorplan and filled with black. We then use a radial floodfill operation starting at each point of the tessellated way and copying every examined pixel into the mask bitmap.

```

void calculate_mask()
for p in tessellated_way{
    radial_flood_fill(p);
}

```

The radial floodfill algorithm is an algorithm, which fills out the area surrounding a point as long as there is a direct line between each point and the starting point. Note that the resulting area need not be convex.

4) *Step 4: Multiply the landmark map with the mask:* In this step, we stamp the visible area out of the landmark map. For each black pixel in the mask bitmap, we black out the same pixel in the landmark map.

```

void mask_out()
for each (x,y) in landmark_map{
if (mask_map(x,y) != black)
    result(x,y) = landmark_map(x,y)
}

```

This results in a bitmap containing exactly the visible landmarks together with their geometric location. It is now up to the rest of the visualisation pipeline how to work further. One could quickly scan the image and get a list of visible landmarks or one could just overlay all landmark pixels with a pictogram assigned to the landmark.

Though this algorithm seems to be very complex, it has some beneficial properties. First of all it is a formulation of the landmark problem in terms of basic image processing. Secondly its result is near to a complete visualisation of the landmarks and thirdly it is highly parallelisable and designed to be run on dedicated graphics hardware. Examples for an application of this algorithm are given in Figure 2.

#### IV. POST-PROCESSING WAYS

Waypoint graphs are either large or do not contain nice-looking ways. And even if a waypoint graph contains all nice-looking ways, these ways will not be the shortest. So for an indoor navigation system it is difficult to find ways, which are short and at the same time have a specific quality with respect to visualisation. As the efficiency of search algorithms is tightly coupled with the number of vertices and edges inside the graph, it is common that people try to have small waypoint graphs. Shortest ways in such graphs tend to scrape along walls or seem otherwise unnatural.

With our algorithm, we want to post-process such ways to obtain a relatively nice visualisation with acceptable computational overhead.

##### A. An Algorithm for Post-Processing Ways for Better Visualisation

Therefore we propose the following algorithm, which essentially is a series of image processing operations. The input of the algorithm consists of

- A way (given as a list of points forming a linestrip)
- A collision map

The configurable parameters influencing this algorithms are

- The maximal length a point may move during the algorithm
- The length threshold used during tessellation of the way
- The valuation choosing the best movement in a set of possible movements

1) *Step 1: Tessellation of the Way:* For this algorithm we need a tessellation of the way just as for the previous algorithm. The reader is referred to Section III-A2 for details.

2) *Step 2: Move the tessellation points:* For each point in the tessellation of the way determine a set of points, where we could move this point. Therefore we grow a circle until this circle collides and if this circle collides, we move the middlepoints away from the collision points until we get stuck. Hence we find out the position where - in a limited



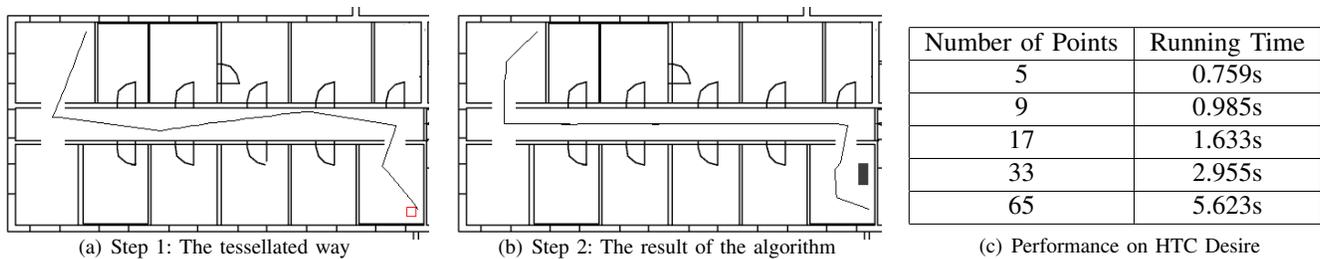


Figure 3. The Post-Processing Way algorithm and its performance running on HTC Desire (complete screen buffer)

(a Java bitmap of the exact pixel size of the screen), which is then passed to our implementation of this algorithm via Java natives. In Java natives, we are performing the image processing as described on a 16-bit-per-pixel bitmap (using essentially 5 bits per color). Therefore we implemented a fast and stable bitmap manipulation library. The tessellation is being performed in Java. Experimental performance results for the Post-Processing Ways algorithm are given in Figure 3(c).

As you can see, for moderate numbers of tessellation points the running time of the algorithm is quite acceptable. The algorithm has to be run only once for each navigation result. As the effective screen resolution of a full-screen application (not drawing over the status bar) on the device is 480x725, the number of tessellation points to consider will not exceed 20 points. As mobile devices are able to run this type of algorithms natively, we are able to provide full navigation functionality with navigation graphs, which have relatively bad visualisation properties such as corner graphs.

## VI. OUTLOOK

The feasibility of this algorithm has been shown with corner graphs as well as with grid-based accessibility graphs. The corner graph example from the introduction has been used in the description of the algorithm in Section IV.

As all algorithms in this paper perform a specific independent task (e.g., radial floodfill or growing circles) on many elements (the points of the tessellation), they are perfectly suited for implementation on modern many-core architectures such as NVIDIA CUDA [10] or the upcoming platform independent OpenCL standard [11]. In these many-core environments hundreds of cores are ready to perform a specific task on a stream. The only precondition for using such architectures is that the parallelised operation is completely independent from each other in the sense that the ordering of execution is arbitrary. This type of execution does of course only apply to server-hosted navigation systems unless these many-core architectures are available on smartphones.

This work allows us to follow a new philosophy for indoor navigation graph generation. As the visualisation quality of the graph is less important if our algorithms are applicable,

we can drop this requirement and generate and readily use minimal graphs such as the corner graph for navigation.

## REFERENCES

- [1] Y. Chen and H. Kobayashi, "Signal strength based indoor geolocation," in *Proceedings of IEEE International Conference on Communications (ICC '02)*, vol. 1, 2002, pp. 436–439.
- [2] F. Evennou and F. Marx, "Advanced integration of wifi and inertial navigation systems for indoor mobile positioning," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 164–164, uary.
- [3] C. Falsi, D. Dardari, L. Mucchi, and M. Z. Win, "Time of arrival estimation for uwb localizers in realistic environments," *EURASIP J. Appl. Signal Process.*, vol. 2006, pp. 152–152, uary.
- [4] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, feb. 2002.
- [5] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation and tracking," in *IEEE Transactions on Signal Processing*, vol. 50, no. 2, 2002, pp. 425–437.
- [6] P. Ruppel, F. Gschwandner, C. K. Schindhelm, and C. Linnhoff-Popien, "Indoor navigation on distributed stationary display systems," *Computer Software and Applications Conference, Annual International*, vol. 1, pp. 37–44, 2009.
- [7] A. Butz, J. Baus, A. Krüger, and M. Lohse, "A hybrid indoor navigation system," in *IUI*, 2001, pp. 25–32.
- [8] P. Tozour, "Search space representations," *AI Game Programming Wisdom 2*, vol. 2, pp. 85–102, 2004.
- [9] K. Yu, "Finding a natural-looking path by using generalized visibility graphs," in *PRICAI 2006: Trends in Artificial Intelligence*, ser. Lecture Notes in Computer Science, Q. Yang and G. Webb, Eds. Springer Berlin / Heidelberg, 2006, vol. 4099, pp. 170–179.
- [10] NVIDIA, *NVIDIA CUDA Programming Guide 2.0*. NVIDIA, 2008.
- [11] K. O. W. Group, "The opencl specification, version 1.0.29," Web site: <http://khronos.org/registry/cl/specs/opencl-1.0.29.pdf> [Last accessed: 13 October 2010], 2008.

## Near Field Sensing in Wireless LANs

Wolfgang Narzt and Heinrich Schmitzberger

Department of Business Informatics – Software Engineering

Johannes Kepler University of Linz

Altenberger Straße 69, 4040 Linz, Austria

{wolfgang.narzt, heinrich.schmitzberger}@jku.at

**Abstract** — Miscellaneous WLAN-based localization methods focus on preferably precise location estimation and tracking. However, the infrastructure of a WLAN system can also be utilized for relative spatial assessment within discrete zones, not using a reference coordinate system. This paper proposes a WLAN-based localization prototype that is capable of unambiguously distinguishing detached interaction zones, allowing applications to imitate near field communication (NFC). Operation is accomplished by customary mobile devices not equipped with NFC technology but a WLAN interface. A prototypical payment scenario at cash-desks in a store proves the concept, where customers are required to securely show their electronic store card, i.e., bring their mobile device into a near, non-interfering zone to their cashier. The paper provides measured figures revealing clearly distinguishable interaction zones in WLAN environments.

**Keywords** – Near Field Sensing; Interaction Zones; Wireless LAN.

### I. INTRODUCTION

Near Field Communication (NFC) has evolved as the preferred technology for applying contactless interaction services at near distances. It is therefore considered an indispensable paradigm in mobile computing environments [1]. Several commercial fields of application profitably make use of NFC, e.g., access control systems, payment systems, or time measurement systems (e.g., using RFID chips attached to the shoes of runners at marathon events). However, employing NFC implicates disadvantages in terms of non-restrictive mobile interaction: people are either required to carry a smart card for each individual application or a commercially hardly available NFC-equipped mobile phone, which accordingly diminishes broader utilization.

The alternative for near range communication, also based on radio frequency signals, is Bluetooth [2]. Although, the penetration of Bluetooth is already widespread and has found its way into a majority of state-of-the-art mobile phones, it has hardly won recognition concerning public mobile applications, primarily due to security and privacy concerns of its users. Likewise, Bluetooth offers a low bandwidth for service applications and restricts the number of clients to be served by specification.

In order to overcome these impairments regarding low penetration and security issues, we propose a complemen-

tary option for contactless interaction at near distances based on Wireless LAN [3]. This technology is widely spread and considered as mature and trustworthy. Latest investigations on WLAN localization [4] have revealed the ability of the Wi-Fi medium to unambiguously recognize the presence of mobile devices at very near distances. Moreover, WLAN offers a broad service bandwidth and comprises elaborate security concepts. Hence, it is an eligible and powerful alternative to NFC in terms of contactless interaction at spatial proximity [5] [6].

The paper is structured as follows: Section 2 deals with selected points of state-of-the-art methods and technology. Section 3 gives an insight into the proposed WLAN architecture for discrete zones separation. Section 4 illustrates the applicability of the architecture in the frame of a use-case scenario. Section 5 provides figures and measured results and finally, section 6 concludes the paper and prospects future work.

### II. RELATED WORK

Near field sensing is a popular method for determining the proximity of a mobile user to the known location of one or several sensors and has been widely studied on the basis of various radio technologies. Hightower and Borriello's survey on location sensing [7] revealed three distinct approaches for inferring the proximity of an object or a person to a sensor resulting from "a physical phenomenon with limited range" that is either physical contact, the contact of an ID tag (e.g., credit card, RFID tag, etc.) with a sensor device whose location is known, or the being in range of one or more access points in a wireless cellular network. Alternatively interpreted regarding the underlying distance, a separation of the user's vicinity to a known location into three distinct zones of proximity can be concluded, i.e., a direct zone, a near zone, and distant zone. In this work we emphasize a different classification using a near zone and several graded distant zones for interaction (cf. Section 5). Indulska and Sutton refine the term proximity sensor in [8], which according to them describes a sensor that locates an entity or device as being within a region. In order to accurately estimate a mobile user's position however a setup of a number of overlapping proximity sensors is needed. Given a region with sufficiently overlapping sensor ranges, the location can then be triangulated. At its core, the approach accentuated in this work

can be seen as a variant of a proximity sensor system. A main claim however is to reliably distinguish a greater amount of interaction zones. To this end, the proposed system utilizes peer-to-peer communication to cooperatively narrow down the radio proximity ranges and consequently conclude on more discrete zones.

The term proximity sensing itself has become popular in the course of Bluetooth technology research. As in this work, most contributions in the Bluetooth context concentrate on signal strength measurements utilizing stationary beacons [9] or PCs as sensor stations [10]. In either way, location estimation is achieved by detecting the close proximity of a mobile device to a beacon or a sensor station making use of the limited radio communication range of the Bluetooth technology. Another important aspect that we are also emphasizing in this paper has been demonstrated in [11]. Hay and Harle use a localization approach that realizes the tracking of mobile phones without installing additional software. The acceptability of our presented system strongly depends on such usability considerations in a real-life setup. Consequently, our system avoids any form of user-side software modification by solely depending on a web browser for service consumption.

In the context of localization accuracy, ultrasound technology is considered the most precise sensor technology for determining the location of an emitter, allowing accuracy results of 10 centimeters and below [12]. In [13], the Active Bat system has been presented as one of the first systems to utilize ultrasound sensor infrastructure for indoor localization using a time difference of arrival (TDOA) algorithm to track a user carrying an emitter tag. To refine the potential whereabouts of its clients, the back-end system constructs a bounding region for each Bat emitter in relation to the radio zone covered by the nearest sensors. Similar to the WPE approach this system combines sensor readings to a fingerprint for each zone to classify the covered environment into separable regions. Another usage of spatial regions was demonstrated by the Relate system [14]. Mobile peers equipped with an ultrasound sensor were used to study the incorporation of proximity aspects into the user interface and present a toolkit API for mobile applications.

In the last decade of research, the most prominent basis platform for near field sensing and interaction was RFID. The LANDMARC system [15] proposes stationary deployed RFID readers as sensors to determine the position of active tags within range. For positioning refinement, the received signal strength is compared to the measurements of reference tags deployed at known positions. A use case for RFID localization related to the application presented in this paper was discussed in [16]. In their work, sequentially deployed RFID readers provide queue length estimation that senses the proximity of tags passing by. In this paper we suggest to apply proximity information in a different way. Derived from the proximity to one or more stationary sensors, we associate mobile users with certain interaction zones. These zones determine the user's interaction interface with the back-end server systems.

Localization in the context of WLAN technology does typically no longer involve near field sensing, especially because the accuracy achieved by other approaches for location estimation allows a much finer location resolution (e.g., using particle filters [17]). WLAN provides a higher range of signal dispersion as the Bluetooth technology for instance, meaning that the communication range between two stations comes up to 100 meters in indoor environments. Applying a straightforward Cell-of-Origin (COO) algorithm does not narrow down the client's location notably. The NearMe Wireless Proximity Server [6] addresses this issue on a proximity sensing basis by applying a peer-to-peer technique. Instead of computing absolute location information of mobile clients, the system determines the proximity of two mobile users by mutually exchanging lists of Wi-Fi signatures (i.e., lists of access points and clients signal strengths). The similarities in the signatures help estimating the distance. Similar to the system presented in this paper, NearMe does not rely on a training phase since it uses relative location instead of absolute location (e.g., WGS-84 coordinates). A similar neighborhood reasoning localization approach has been described in [18]. They improve WLAN-based position estimation with ZigBee sensor readings that compensate dynamically appearing signal interferences (as provoked by passers-by). Though such interferences do not affect close proximity detection achieved by the WPEs, we use a comparable technique to improve distant zones separation.

### III. WLAN-BASED NEAR-FIELD SENSOR NETWORK

The main design objective for near-field sensing in Wireless LANs was to implement a system operating without any client pre-requisites but a WLAN interface and a mobile internet browser for service access. Commercial availability of WLAN in public places and its integration into modern mobile phones suggest these goals. In addition, sophisticated encryption already realized on common WLAN infrastructure, security measures and the bandwidth of the 802.11 standard allow for elaborate applications (e.g., multimedia web applications) as opposed to NFC or Bluetooth.

In our approach, all the processing components are settled in the WLAN infrastructure, aiming at reliably separating spatial zones of interaction. Customized networked WLAN sensors (i.e., off-the-shelf access points running a modified software kernel) utilize proximity recognition by evaluating signal peaks of connected devices. The client is not requested to emit special tracking signals – the system uses the client's communication traffic originated from (web-)service consumption for proximity detection purposes. In contrast to absolute location estimation the system does not need a preceding training phase.

Figure 1 illustrates the system architecture of the near-field sensor network. The nucleus of our system is contained within conventional access points, further referred to as "sniffers". They are executing an altered Linux operating system and customized proximity detection software. In our setup we use Linksys WRT610N access points with

a 533MHz system processor and two separate WLAN interfaces covering the 2.4GHz and the 5GHz frequency band). The 2.4GHz band (802.11bgn) is used for proximity detection and service provisioning. The 5GHz band (802.11an) acts as backbone network for the sniffers.

The hardware platform is capable of concurrently running the proximity detection algorithms, a web server and a database in the background, incorporating service provider functionality in the sniffers. Optionally, our setup supports interfacing with a back-end server to ease the integration into existing service infrastructure at potential deployment sites.

As shown in Fig. 1 a sniffer consists of 4 components:

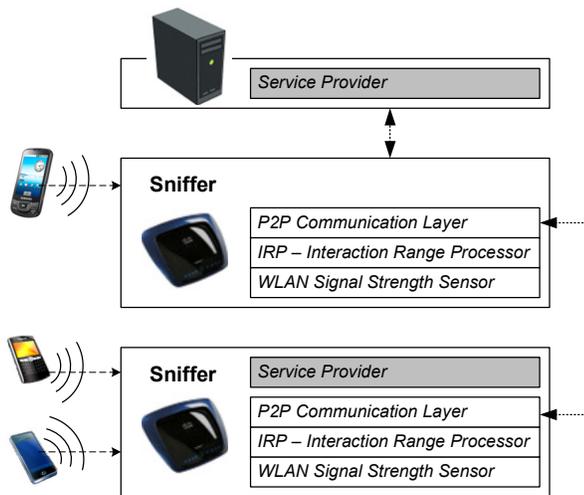


Figure 1. System Architecture

1. A *WLAN signal strength sensor* is realized as a low-level daemon process that queries the interface driver in raw packet monitoring mode for RSSI (Received Signal Strength Indicator) measurements of the mobile clients. Its purpose is to supply the second component layered above with measurement data in real time.
2. The *interaction range processor (IRP)* uses these data to separate spatial regions into distinct interaction zones. As our proposed setup demands for reliable zone separation, the IRP uses unambiguous peak values to determine the respective interaction zone of a mobile client.
3. In order to allow sniffer cooperation the system entails a *peer-to-peer communication layer*. During an initial discovery phase each sniffer executes a simplified voting algorithm using broadcasts on the backbone network. The first appearing sniffer is assumed the master peer, which waits for other sniffers to appear on the network until the configuration application is triggered. The master peer acts as central instance hosting the configuration for defining the sensor network topology and additional parameters that represent the setup environment. After configuration, each cooperating sniffer continuously reports live proximity measurements to the master peer, which acts as front-end and determines

the relative location of inquiring clients on the basis of the sensor input delivered by the sniffer network.

4. Finally, the front-end application offered by the *service provider* differentiates the clients' locations into interaction zones by applying a set of topology depending separation patterns. The system distinguishes the near interaction zone (i.e., signal strength measurements of -25dBi and higher) and several distant interaction zones graded by signal strength thresholds. Depending on the amount of cooperating sniffers and the characteristics of the setup environment the granularity of distant zones can be refined. A more detailed discussion on refining these interaction zones is given in Section 5.

Signal strength fluctuation provoked by people in the line of sight between sensor and the inquiring client is compensated by using a stationary control signal emitter placed behind the region of interest [19] (see Fig. 2). The control signal is steadily broadcasted by a WLAN-enabled device (e.g., an ordinary access point or a mobile phone) and measured at each sniffer. The initial signal strength value of the control signal is stored as a reference at each sniffer, enabling adjustments for client signal measurements during live operation.

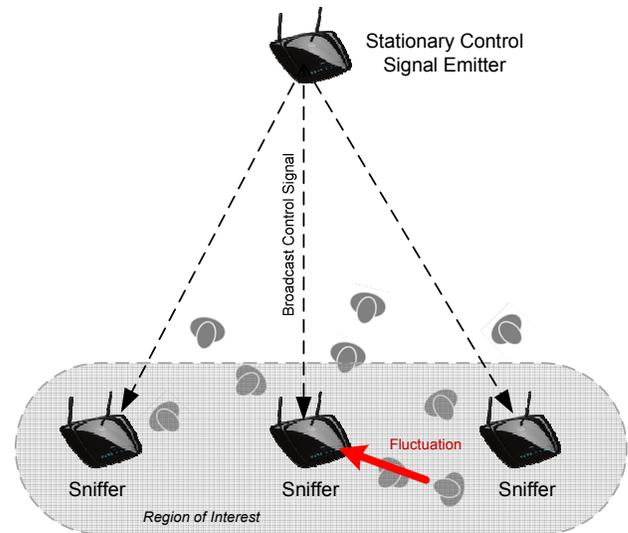


Figure 2. Stationary Control Signal Emitter

#### IV. USE CASES

Our system has prototypically been implemented for a supermarket cash desk scenario under laboratory conditions: four cash desks have been arranged in parallel with a distance of about 1.5m to each other. Our application prototype implements an electronic store card utilizable on the clients' mobile phones, which can be "shown" to the cashier on a single button click and automatically associated to the correct cash desk and purchase. The challenge in this setup is to confidently detect the correct desk by WLAN depending on the client's proximity when the client presses the button on his mobile phone. Moreover, it must

be assured that several clients in a queue at the same cash desk are handled correctly even when they simultaneously press their buttons. The setup provides for a sniffer at every cash desk mounted at spatial proximity to the cashier. The clients are requested to hold their mobile phones close to the appropriate sniffer and press a button in order to initiate network traffic, which can be used to explicitly determine physical closeness. This further triggers an authentication process to ensure correct association of client and service.



Figure 3. Proximity interaction

Figure 3 shows snapshots of the prototypical arrangement: a Linksys access point is used as the sniffer and detects physical closeness of an off-the-shelf mobile phone (here: Motorola Defy operating on Android 2.2). The browser component of the phone enables the user to consume the provided service of the sniffer recognizable by an authorization screen for “showing” the user’s store card, which only then offers an enabled button when the phone is close enough to the sniffer (a detailed discussion on results is given in section 5). For confirming the button click near the access point every cash desk is equipped with a

screen showing the customers’ identification data through the sniffer service.

We have arranged this setup in four parallel lines in order to simulate a supermarket cash desk scenario with customers being simultaneously served at the four desks and interfering in the queues. Figure 4 exemplarily illustrates that two customers in different lines and at different proximity to the cash desk are handled correctly, i.e., they are only then identified when their mobile device close enough to the access point when pressing the authentication button. This near zone is intended to manage security related interaction (e.g., exchanging customer identification data).



Figure 4. Parallel interaction at cash desk scenario

Beyond operations at very close distances (i.e., in the near interaction zone), the sniffer sensor network is capable of distinguishing further discrete interaction zones (cf. Section 5) enhancing the variety of applications that can be set up upon, e.g., for non critical operations characterizing a semi-close area around the sniffers. At the far distance zone the system could advert to latest offerings and common vendor services. In the vicinity of the checkout lines customers may be reminded of cross-checking their shopping list, by means of a web-service provided by the supermarket, which customers may fill out at home. Enqueued in a checkout line the customers’ waiting time could be shortened e.g., by participating in a (yet anonymous) quality survey rewarded with credits. These credits can finally be encashed right away in the near interaction zone, where the customer is identified for the first time (cf. Figure 5).

In general, the sniffer approach contributes to an innovative interaction paradigm in mobile computing environments, where people are able to trigger electronically controlled actions just at spatial proximity without the needs of glimpsing at displays, typing, clicking or pressing buttons (cf. [20]). Usually, human attentiveness is required by

conventional interaction metaphors via display and/or keystroke at the place of event in order to open a gate, buy a ticket, start or stop an engine, etc. However, attentiveness for pressing a button or glimpsing at a display may occasionally be unavailable when the involved person must not be distracted from performing a task (e.g. while driving in a car) or is handicapped through wearable limitations (e.g. gloves, protective clothing) or disability. As the sniffer on the one hand is capable of discretely detecting physical proximity and on the other hand includes a customizable service provider component it is possible to automatically trigger those actions just at physical closeness of a person, i.e., dismissing displays and keypads in order to ease human computer interaction.

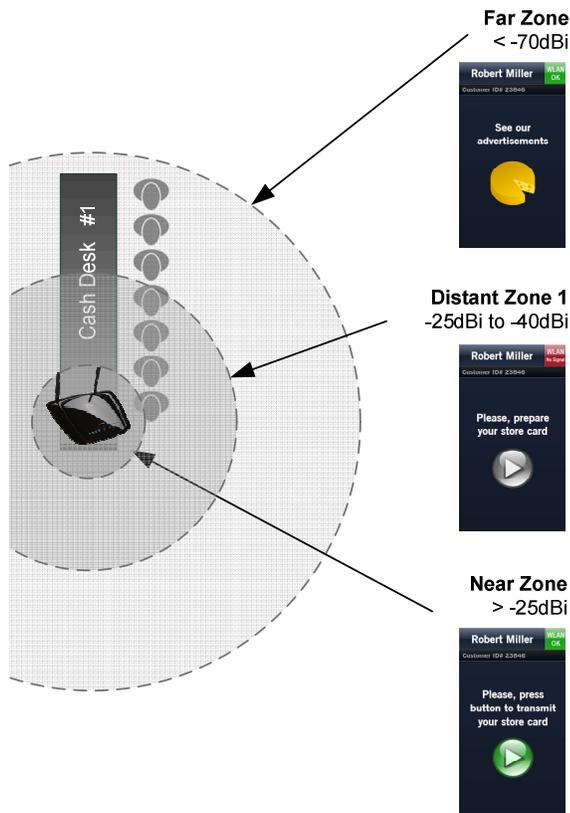


Figure 5. Interaction zones

### V. RESULTS

The majority of WLAN localization systems deal with accurate location estimation and user tracking in indoor environments to supply location-based services with absolute coordinates of the users' current whereabouts. This work emphasizes the usage of discrete interaction zones for application scenarios that benefit from clearly separated zones that can be associated with different functionality (e.g., consumer interaction in the supermarket, public display interaction, access control systems or elderly care scenarios).

Laboratory-based experiments revealed that signals emitted at distances < 30cm (LOS and NLOS) can reliably be differentiated from those sent out beyond. Weak signals transmitted at distances > 15m also show significant measurement characteristics. Consequently, one single sniffer can robustly determine three zones: (i) the near zone identified by signal strength measurements greater than -25dBi, (ii) the far zone identified by signals less than -70dBi referring to distances of > 15m (indoors and NLOS), and (iii) the distant zone for measurements in between these two extremes.

Table 1 lists the measured signal strengths obtained by the sniffer sensor arrangement in the setup described in Section 4 (four cash desk lines) using a HTC Desire smart phone. The quadruples in the table columns refer to the measurements taken by the four sniffers [sniffer<sub>0</sub>, sniffer<sub>1</sub>, sniffer<sub>2</sub>, sniffer<sub>3</sub>]. The highlighted entries mark the respective sniffer assigned to the cash desk line. The bold-faced values in the near zone column illustrate distinct measurement peaks allowing a unique classification.

TABLE I. MEASUREMENT RESULTS

	0m ( <i>near zone</i> )	3m ( <i>dist. zone 1</i> )	9m ( <i>dist. zone 2</i> )
Line0	[ <b>-04</b> , -34, -39, -40]	[-29, -35, -37, -36]	[-45, -41, -41, -38]
Line1	[-32, <b>-05</b> , -29, -32]	[-35, <b>-31</b> , -46, -35]	[-38, <b>-40</b> , -46, -40]
Line2	[-36, -26, <b>-11</b> , -27]	[-39, -35, <b>-34</b> , -29]	[-46, -43, <b>-41</b> , -41]
Line3	[-44, -36, -26, <b>-08</b> ]	[-38, -36, -37, <b>-33</b> ]	[-42, -47, -44, <b>-41</b> ]

Even though the measurements related to distant zone 1 and 2 seem decisive regarding their associated sniffer, the signal strength values within this range tend to fluctuate in the order of ±10dBi mainly due to multipath propagation, attenuation provoked by people in the LOS and emitter characteristics of different WLAN chipsets. In order to robustly separate the two distant zones these fluctuations must be compensated. Hence, we use the collaboratively obtained average value of the measurements to mitigate signal variability. Since the strength of the WLAN signal decreases logarithmically, the system is able to reliably separate four interaction zones in the course of our sketched setup arrangement.

Concerning response times the system is dependent on a recurring signal emission (e.g., small UDP packets) of the mobile device. In our setup we have chosen an average transmission interval of 1.5 seconds, which is slow enough not to overload the wireless network backbone and fast enough to apply for the use-case. Generally, the transmission interval (and therefore the response time) is selectively adaptable to specific use-cases.

In this context, scalability of the system is coherent to the physical limits of WLAN. Adding one access point likely increases accuracy regarding the distant and the far zones. Besides, the setup of the cash-desk scenario restrains the users in sequential lines and therefore provides constructional boundaries for the number of clients to be served by one sniffer.

## VI. CONCLUSION AND FUTURE WORK

WLAN-based localization mainly focuses on accuracy aspects concerning absolute positioning as an indoor alternative to GPS. In this work we present a WLAN setup that can be used as an alternative to Near Field Communication utilizing a proximity-based mechanism to determine relative spatial associations of mobile users. To this end, we have developed a network of wireless proximity sensors, i.e., either detached or collectively applicable entities associating mobile devices with discrete interaction zones. In the course of a prototypical cash desk setup we have robustly distinguished four interaction zones providing specific customized services (e.g., store card authorization, advertisement delivery, electronic shopping list, etc.). Our system is instantly operable without any training effort and users can interact without any prerequisites on the client-side but a WLAN interface.

Funded results presented in [19] confirm that the accuracy of indoor localization benefits from spatial variability, i.e., the reflection, diffraction or absorption of the WLAN signal by stationary obstacles (such as furniture, walls, doors and alike) leading to unique characteristics of each potential location spot. Given such characteristics typically found in real-life environments the number of distinguishable interaction zones is likely to increase, but has not been verified, yet. In this context, further investigation has to be conducted on filter patterns for separating the zones combined with the arrangement of the sniffer sensors (e.g., parallel, circle, square, radial, etc.) and also on analytical error estimation and on power consumption issues compared to Bluetooth or NFC in order to prove the sufficiency of our proposed model.

## REFERENCES

- [1] A. Ferscha, M. Hechinger, R. Mayrhofer, M. dos Santos Rocha, M. Franz, and R. Oberhauser, "Digital aura," in *Advances in Pervasive Computing. A Collection of Contributions Presented at the 2nd International Conference on Pervasive Computing (Pervasive 2004)*, vol. 176. Vienna, Austria: Austrian Computer Society (OCG), April 2004, pp. 405–410.
- [2] G. Lenzini, "Trust-based and context-aware authentication in a software architecture for context and proximity-aware services," in *Architecting Dependable Systems VI*, ser. Lecture Notes in Computer Science, R. de Lemos, J.-C. Fabre, C. Gacek, F. Gadducci, and M. ter Beek, Eds. Springer Berlin / Heidelberg, 2009, vol. 5835, pp. 284–307, 10.1007/978-3-642-10248-6\_12.
- [3] C. Holzmann, M. Hechinger, and A. Ferscha, "Relation-centric development of spatially-aware applications," in *Proceedings of the 17th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE 2008)*. Rome, Italy: IEEE CS Press, June 2008.
- [4] H. Schmitzberger and W. Narzt, "Leveraging wlan infrastructure for large-scale indoor tracking," in *Proceedings of the Sixth International Conference on Wireless and Mobile Communications (ICWMC 2010)*, Valencia, Spain, IEEE Computer Society, 2010, pp. 250–255.
- [5] M. B. Kjaergaard, G. Treu, P. Ruppel, and A. Küpper, "Efficient indoor proximity and separation detection for location fingerprinting," in *MOBILWARE '08: Proceedings of the 1st international conference on MOBILE Wireless MiddleWARE, Operating Systems, and Applications*. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2007, pp. 1–8.
- [6] J. Krumm and K. Hinckley, "The nearme wireless proximity server," in *UbiComp 2004: Ubiquitous Computing*, ser. Lecture Notes in Computer Science, N. Davies, E. Mynatt, and I. Siio, Eds. Springer Berlin / Heidelberg, 2004, vol. 3205, pp. 283–300.
- [7] J. Hightower and G. Borriello, "A survey and taxonomy of location sensing systems for ubiquitous computing," University of Washington, Department of Computer Science and Engineering, Seattle, WA, UW CSE 01-08-03, August 2001.
- [8] J. Indulska and P. Sutton, "Location management in pervasive systems," in *ACSW Frontiers '03: Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003*. Darlinghurst, Australia, Australia: Australian Computer Society, Inc., 2003, pp. 143–151.
- [9] S. Chawathe, "Beacon placement for indoor localization using bluetooth," in *Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, 2008, pp. 980–985.
- [10] J. J. M. Diaz, R. d. A. Maues, R. B. Soares, E. F. Nakamura, and C. M. S. Figueiredo, "Bluepass: An indoor bluetooth-based localization system for mobile applications," in *Computers and Communications (ISCC), 2010 IEEE Symposium on*, 2010, pp. 778–783.
- [11] S. Hay and R. Harle, "Bluetooth tracking without discoverability," in *Proceedings of the 4th International Symposium on Location and Context Awareness*, ser. LoCA '09. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 120–137.
- [12] S. Zhou, H. Feng, and R. Yuan, "Error compensation for cricket indoor location system," in *Parallel and Distributed Computing, Applications and Technologies, 2009 International Conference on*, 2009, pp. 390–395.
- [13] O. Woodman and R. Harle, "Concurrent scheduling in the active bat location system," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference*, 2010, pp. 431–437.
- [14] G. Kortuem, C. Kray, and H. Gellersen, "Sensing and visualizing spatial relations of mobile devices," in *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*. New York, NY, USA: ACM, 2005, pp. 93–102.
- [15] L. Ni, Y. Liu, Y. C. Lau, and A. Patil, "Landmark: indoor location sensing using active rfid," in *Pervasive Computing and Communications, 2003. (PerCom 2003). Proceedings of the First IEEE International Conference on*, 2003, pp. 407–415.
- [16] D. Sanders, S. Mukhi, M. Laskowski, M. Khan, B. Podaima, and R. McLeod, "A network-enabled platform for reducing hospital emergency department waiting times using an rfid proximity location system," in *Systems Engineering, 2008. ICSENG '08. 19th International Conference on*, 2008, pp. 538–543.
- [17] J. Hightower and G. Borriello, "Particle filters for location estimation in ubiquitous computing: A case study," in *Proceedings of the UbiComp 2004*, 2004.
- [18] L.-w. Chan, J.-r. Chiang, Y.-c. Chen, C.-n. Ke, J. Hsu, and H.-h. Chu, "Collaborative localization: Enhancing wifi-based position estimation with neighborhood links in clusters," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, K. Fishkin, B. Schiele, P. Nixon, and A. Quigley, Eds. Springer Berlin / Heidelberg, 2006, vol. 3968, pp. 50–66.
- [19] M. Youssef, M. Mah, and A. Agrawala, "Challenges: device-free passive localization for wireless environments," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2007, pp. 222–229.
- [20] W. Narzt and H. Schmitzberger, "Location-triggered code execution — dismissing displays and keypads for mobile interaction," in *UAHCI '09: Proceedings of the 5th International on Conference Universal Access in Human-Computer Interaction. Part II*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 374–383.

# Spectrum Sensing Measurement using GNU Radio and USRP Software Radio Platform

Rozeha A. Rashid, M. Adib Sarijari, N. Fisal, S.  
K. S. Yusof, N. Hija Mahalin  
Faculty of Electrical Engineering  
Universiti Teknologi Malaysia  
81310 UTM Johor Bahru, Johor, Malaysia  
{adib\_sairi, rozeha| sheila| sharifah|  
kamilah}@fke.utm.my, nhija.m@gmail.com

A. C. C. Lo  
Wireless & Mobile Communications Group  
Delft University of Technology  
The Netherlands  
A.C.C.Lo@tudelft.nl

**Abstract**— Spectrum utilization can be significantly improved by adopting cognitive radio (CR) technology. Such radios are able to sense the spectral environment and use this information to opportunistically provide wireless links that meet the user communications requirements optimally. To achieve the goal of cognitive radio, it is a fundamental requirement that the cognitive user (CU) performs spectrum sensing to detect the presence of the primary user (PU) signal before a spectrum is accessed as to avoid harmful interference. Therefore, two probabilities are of interest; the probability of detection,  $P_d$  and the probability of false alarm,  $P_{fa}$ . In this paper, we investigate sensing performance implemented on real-time testbed of GNU Radio and USRP Software Defined Radio (SDR) communication platform operating at 2.48 GHz with a bandwidth of 4 MHz. Energy detector utilizing 1024 FFT bin is the sensing mechanism used in the experimental set-up. The acquired experimental results of  $P_d$  and  $P_{fa}$  are duly analyzed and verified to be comparable to the curve of the theoretical framework for line-of-sight indoor environment. It is observed that at a target  $P_{fa}$  of 5%, the optimal decision threshold for PU detection is -39 dB. The plot of measured number of samples needed for a desired  $P_d$  for various received signal levels, representing various signal-to-noise (SNR) conditions, is also included. At SNR of 0 dB and a target Quality of Service (QoS) set at  $P_d$  of 90%, it is found out that the required sensing time for our GNU Radio USRP based CR system is equal to 31.59ms.

**Keywords**- Spectrum Sensing; Probability of Detection; Probability of False Alarm; GNU Radio; USRP.

## I. INTRODUCTION

A recent spectrum occupancy measurement shows that a significant portion of the spectrum allocated to licensed services show little usage over time, with concentration on certain portions of the spectrum while a significant amount of the spectrum remains unutilized [1]. A new communication paradigm to exploit the existing wireless spectrum opportunistically is necessary to overcome limited available spectrum and inefficiency in spectrum utilization.

Originally introduced by Mitola, [2], Cognitive Radio (CR) technology allows unlicensed or cognitive users (CUs) to take advantage of the spectrum holes by intelligently identifying and using them, and possibly releasing them when required by the primary users (PUs). Hence, it is a

fundamental requirement that the CU performs spectrum sensing to detect the presence of PU signal and also locate unoccupied spectrum segments, the spectrum holes, as accurately and quickly as possible.

Various approaches have been proposed for spectrum sensing such as matched filter, energy detection, feature detection and more recently, wavelet detection methods [3]. In local sensing, each CU senses the spectrum within its geographical location and makes a decision on the presence of PUs based on its own local sensing measurements. The matched filter (MF) is an optimum coherent detector. However it requires a prior knowledge on the behavior (modulation) of the received signal. Energy detection (ED) is a non-coherent detection method that uses the energy of the received signal to determine the presence of primary signals. This simple method is able to gather spectrum-occupancy information quickly. However, its sensing capability is vulnerable to noise. Cyclostationary detector exploits the inherent periodicity in the received signal to detect primary signals with a particular modulation type by implementing a two-dimensional spectral correlation function (SCF) rather than the one-dimensional power spectral density (PSD) of the energy detector. Its spectrum-sensing performance is robust to noise-like signal. However this method demands excessive Analog-Digital Converter (ADC) requirement and signal processing capabilities, thus accompanying a large amount of power consumption [4]. In this research, energy detector is chosen due to the assumption that CU has limited information on the primary signal (i.e. only the local noise power is known). Hence energy detector is optimal. In addition, IEEE802.22 standard on cognitive radio has spectrum sensing via energy detection in its provision.

The flexibility of Software Defined Radio (SDR) makes it a well suited candidate for the implementation of cognitive features [5]. One of the first research works on implementation of spectrum sensing is reported in [3] using SDR platform called Berkeley Emulation Engine 2 (BEE2). The sensing mechanisms utilized are the energy and cyclostationary detectors. However, the measurements and analysis are based on the captured local data. While in [6], Dynamic Spectrum Access (DSA) mechanism is developed

using an SDR platform consisting of USRP and GNU radio. GNU Radio is an open source software toolkit which consists of a huge numbers of signals processing blocks library (i.e modulators, filters, amplifiers and etc). This signal processing blocks can be linked together for building and deploying the baseband part of the dedicated radio [9], [10]. USRP is the hardware link between this baseband signal blocks and the real environment. Its function is to change the analog value of the spectrum to the digital domain and to change the digital domain signal to analog value [7], [8]. The overall block diagram is shown in Figure 1.

This research has two goals; (1) is to develop a DSA based CR system using GNU Radio and USRP, by combining the works of [3] and [6], in addition to having our own designed spectrum management and decision and (2) to do measurement and analysis based on the captured local data. However, this paper only highlights the use of GNU radio and USRP as the testbed for the implementation and measurement of ED sensing scheme and the determination of these parameters from the local data; sensing threshold, sensing time, probability of false alarm ( $P_{fa}$ ) which is the probability of wrong detection of PU when it is absent, and the probability of detection ( $P_d$ ) that is the probability of correct detection of a PU.

The paper is organized as follows: Section II presents a background on cognitive radio technology. While Section III defines channel sensing hypotheses and introduces the sensing performance metrics, the probabilities of detection and false alarm. Section IV discusses energy detector block diagram of of GNU Radio. The experimental set-up and preliminary results are presented in Section V. Finally, Section VI provides conclusion and future works.

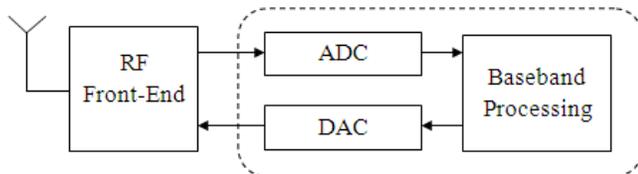


Figure 1. Software defined radio block diagram

## II. COGNITIVE RADIO TECHNOLOGY

The vital objective of the CR is to achieve the best accessible spectrum through cognitive capability and reconfigurability. In other words, CR also embodies awareness, intelligence, learning, adaptivity, reliability and efficiency. Cognitive cycle consists of three major steps as follows [1],[2]:

- a) Sensing of RF stimuli which involves the detection of spectrum holes to facilitate the estimation of channel state information and prediction of channel capacity for use by the transmitter.
- b) Cognition/spectrum management which controls opportunistic spectrum access and capturing the best available spectrum to meet user

communication requirements. Cognitive radios should decide on the best spectrum band to meet the Quality of Service (QoS) requirements over all available spectrum bands by managing functions such as optimal transmission rate control, traffic shaping and routing.

- c) Actions to be taken can be in terms of re-configurable communication parameters such as transmission power, modulation and coding.

Those three tasks form a cognitive cycle as shown in Figure 2.

In this research work, DSA based CR system is developed using USRP and GNU radio as hardware and software platforms, respectively. The proposed design consists of four main functional blocks which are spectrum sensing, spectrum management, spectrum decision and data transmission [12]. However, the contribution of this paper is limited to spectrum sensing that concerns with the sensing parameters such as the sensing threshold, probability of false alarm, probability of detection and the sensing time, which are all decided based upon locally measured data. The analysis on the captured data is used to assess the platform characteristics in terms of sensitivity and best performance in local environment. The results are further utilized to set the desired Quality of Service (QoS) for the system.

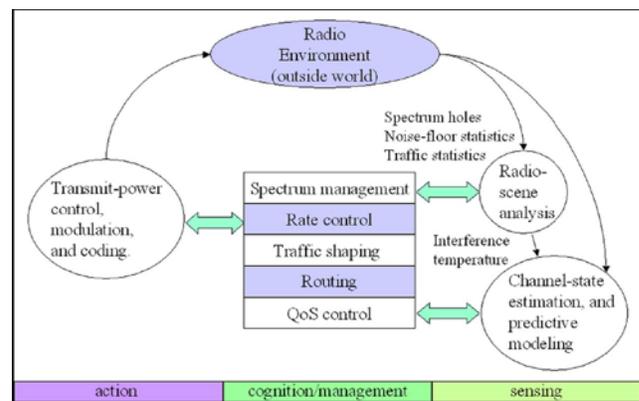


Figure 2. Basic cognitive cycle [2]

## III. CHANNEL SENSING HYPOTHESES

The sampled received signal,  $X[n]$  at the CU receiver will have two hypotheses as follows:

$$\begin{aligned} H_0: X[n] &= W[n] && \text{if PU is absent} \\ H_1: X[n] &= hW[n] + S[n] && \text{if PU is present} \end{aligned} \quad (1)$$

where  $n = 1, \dots, N$ ;  $N$  is the number of samples and  $h$  is the gain of channel that is assumed to be 0 under hypothesis  $H_0$  and 1 under hypothesis  $H_1$ . The noise  $W[n]$  is assumed to be additive white Gaussian (AWGN) with zero mean and variance  $\sigma_w^2$ .  $S[n]$  is the PU's signal and is assumed to be a random Gaussian process with zero mean and variance  $\sigma_x^2$ .

Using energy detector, the decision based on Neyman-Pearson criterion will be

$$Y = \frac{1}{N} \sum_{n=1}^N (X[n])^2 \quad (2)$$

where  $Y$  is the output of the energy detector which serves as the test statistic. Taking  $\gamma$  as the threshold to decide whether signal is present or not, the performance of energy detector can be characterized by a resulting pair of  $(P_{fa}, P_d)$  as the probabilities that the CU's sensing algorithm detects a PU under  $H_0$  and  $H_1$ , respectively.

$$P_{fa} = P(Y > \gamma / H_0) \quad (3)$$

$$P_d = P(Y > \gamma / H_1) \quad (4)$$

If the noise term is assumed to be circularly symmetric complex Gaussian, using central limit theorem, Gaussian distribution approximation for the probability density function (PDF) of  $Y$ , it can be derived from (3) and (4) [13];

$$P_{fa} = Q\left(\left(\frac{\gamma}{\sigma_w^2} - 1\right)\sqrt{N}\right) \quad (5)$$

$$P_d = Q\left(\left(\frac{\gamma}{\sigma_w^2} - SNR - 1\right)\sqrt{\frac{N}{2SNR+1}}\right) \quad (6)$$

where signal-to-noise ratio ( $SNR$ ) is taken as  $\frac{|h|^2 \sigma_s^2}{\sigma_w^2}$  and  $Q(\cdot)$  denotes the generalized Marcum Q-function. The challenge of the local spectrum sensing is to reliably decide on the two hypotheses to achieve high  $P_d$  for good protection of PU and low  $P_{fa}$  to provide satisfactory access for CUs.

#### IV. GNU RADIO ENERGY DETECTOR

The energy detector is known as a suboptimal detector, which can be applied to detect unknown signals as it does not require a prior knowledge on the transmitted waveform as the optimal detector (matched filter) does. Figure 3 depicts block-diagram of an energy detector. The ADC is used to convert the received signal to the digital domain. Then the square magnitude of the digitized signal is calculated by using the Fast Fourier Transform (FFT) and magnitude square function. To make the measurement more accurate,  $N$  numbers of samples is taken and the average value of the samples is used to make the decision whether signal is present or not by comparing it with the threshold.

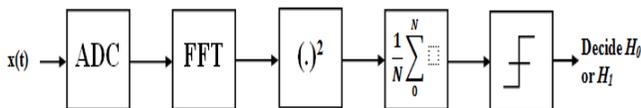


Figure 3. Block diagram of an energy detector

In this research, the GNU Radio energy detector flow graph design is a modification from the GNU Radio's spectrum analyzer which is available in the GNU Radio

package. Figure 4 shows the flow graph of the energy detector used for this research. The Radio frequency (RF) signal is captured and down converted to the baseband frequency (DC) by the RFX2400 USRP front end [8] and then passed to the USRP Motherboard. The daughterboard is necessary since the ADC speed cannot cope with the high frequency RF signal in the air (i.e. 2.4GHz) as the ADC speed is only 64MS/s.

The speed of the digitized signal will be scaled down by the digital down converter (DDC) which is implemented inside the FPGA. The DDC value is programmable with even numbers in the range of 4 to 256. The baseband signal needs to be down converted again by the DDC because the speed of the signal have to be reduced to be transferred to the GNU Radio installed in Personal Computer (PC) through the USB 2.0 bus which has the bottleneck speed at 32MHz.

In GNU Radio, the transferred signal from the USRP which is in the format of a stream of data will first be converted to I & Q format (Vector) by a block called *gr.stream\_to\_vector*. Its task is to take a stream of items as its input and convert it into a stream of blocks containing *nitens\_per\_block* as its output [9]. In this research, *nitens\_per\_block* is equal to the size of our FFT which is 1024. Then, this signal will be pushed into the GNU radio FFT block, the *gr.fft\_vcc*. The term *vcc* at the end of the block indicates that the input of the signal processing block is in complex vector type and the output is also a complex vector [10]. In the FFT block, windowing technique is used to optimize the FFT result. Windowing is a technique used to shape the time portion of the sampled signal. This is to minimize edge effects that will result in spectral leakage in the FFT spectrum and increases the spectral resolution the frequency domain result [11].

The complex output of the FFT block will then be connected to the complex magnitude block named *gr.complex\_to\_mag*. This block takes a complex number as its input and gives the magnitude (in float format) of this number as output. Then, the result of this block will be converted from the ADC value which represents the voltage to the dB value by using *gr.nlog10\_ff* block. Lastly, the result in dB will be sent to a sink block called *gr.message\_sink*. The sink block will then pass the value of the power spectral to the Python layer. Collecting and averaging  $N$  samples of FFT will be done in Python, and then the decision on the presence of PU will be made based on the result of the sensing and predetermined threshold.

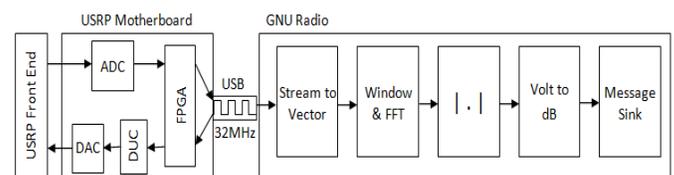


Figure 4. GNU Radio ED Block Diagram

V. ENERGY DETECTOR EXPERIMENTAL RESULTS

The section provides the real-time testbed implementation to evaluate and verify the theoretical results on the performance and limitations of the GNU Radio energy detector. In particular, we measured the achievable probabilities of detection and false alarm as a function of sample size for various signal levels based on a predetermined threshold.

A. Testbed Description

The testbed used in the experiments is the Universal Software Radio Peripheral (USRP) as the Software Defined Radio (SDR) platform and GNU Radio as the SDR software implementation.

USRP consists of two main boards; the daughter board and the mother board. The mother board consists of four 12-bit Analog to Digital Converter (ADC) with sampling rate of 64MS/s, four 14-bit Digital to Analog Converter (DAC) with speed of 128MS/s, two Digital up Converter (DUC) to up-convert the baseband signal to 128MS/s before translating them to the selected output frequency, a programmable USB 2.0 controller for communication between USRP and GNU Radio and an FPGA for implementing four Digital Down Converter (DDC) and high rate signal processing. The daughterboard is acting as the RF front-end of the SDR.

GNU Radio consists of signal processing blocks library and the glue to tie these blocks together for building and deploying SDRs [5],[7],[9]. The signal processing blocks are written in C++ while Python is used as a scripting language to tie the blocks together to form the flow graph. Simplified Wrapper and Interface Generator (SWIG) is used as the interface compiler which allows the integration between C++ and Python language. Figure 5 shows the structure of GNU Radio and USRP SDR. The USRP will digitize the inflow data from the air and passing it to the GNU Radio through the USB interface. GNU Radio will then further process the signal by demodulating and filtering until the signal is translated to a packet or a stream of data.

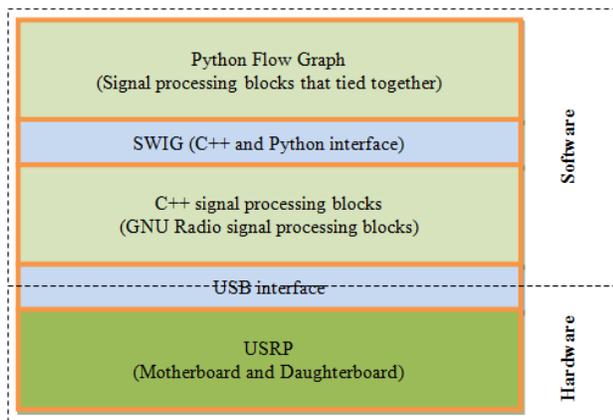
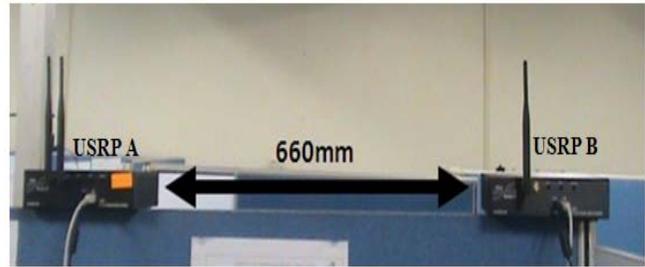


Figure 5. GNU Radio and USRP Structure

B. Energy Detection Implementation

Figure 6 shows the experimental setup consisting of two USRPs, one laptop and one PC. PC with USRP A acts as the receiver while the laptop and USRP B acts as the transmitter. Daughter boards used for these experiments are RFX2400 which can cover frequencies from 2.3GHz to 2.9GHz. There is a neighboring access point to consider which can interfere with the USRP frequencies. Therefore, the USRP center frequency is set at 2.48GHz to avoid the interference and jamming with the said access point operating at 2.4 GHz band.



(a) USRP hardware platform



(b) Block diagram of overall setup

Figure 6. Experimental setup

The energy detector is implemented using 1024 point FFT with a fully parallel pipelined architecture for the fastest speed. Each block of FFT outputs is averaged and stored inside a buffer.  $N$ -numbers of the averaged FFT block will be collected and then averaged again to acquire the final result that will be used to make the decision on the presence of PU.  $N$  value is programmable and it will be set based on the selected  $P_d$ .

C. Experimental Results

We tested a GMSK signal centered at 2.48GHz carrier frequency. In the experiment to determine  $P_{fa}$ , the measurement is carried out when there is no signal transmission from PU and CU as shown in Figure 7. From this figure, it can be observed that the highest noise spike is around -37.0dB and more noise spike is recorded after -41.0dB. Therefore, in this work, the threshold is in the range of -37.0dB to -41.0dB. Thus, the curve for  $P_{fa}$  versus sample size is obtained and we estimated the detection threshold to meet the target  $P_{fa}$ . Then, we applied the threshold to the captured data when the PU signal is present, as shown in Figure 8, and computed the probability of detection,  $P_d$ . Each of the detection measurement is repeated 1000 times to obtain an accurate estimation of  $P_{fa}$  and  $P_d$ .

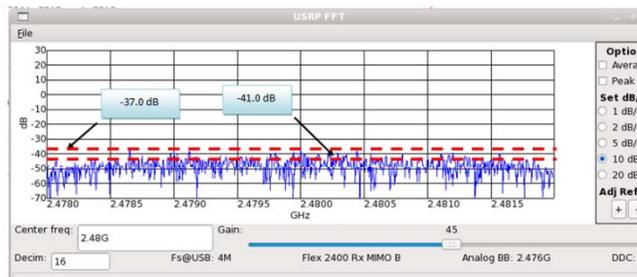


Figure 7. Absence of PU as observed using GNU Radio spectrum analyzer at 2.5GHz

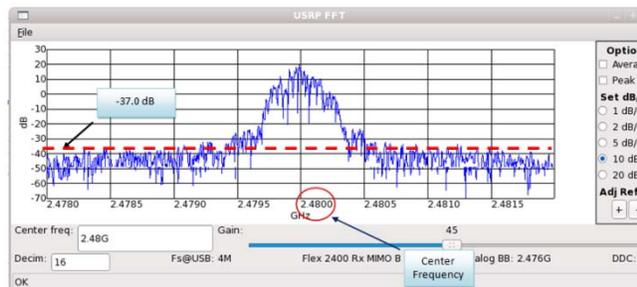


Figure 8. Presence of PU as observed using GNU Radio spectrum analyzer at 2.8GHz

Figure 9 shows the performance of  $P_{fa}$  against sample size,  $N$ . This graph will be used to determine the sensing threshold based on the desired  $P_{fa}$ . For instance, if a target  $P_{fa}$  of 5% is chosen, the threshold curves that intersect with the  $P_{fa}$  value of 0.05 will be considered. It should be noted that higher number of samples size will cause longer sensing time to the CR system. In this work -39.0 dB is chosen as the sensing threshold to decide on the presence of PU for  $P_d$  measurement since it crosses  $P_{fa}$  value of 0.05 and hence, satisfies the desired  $P_{fa}$  of 5%.

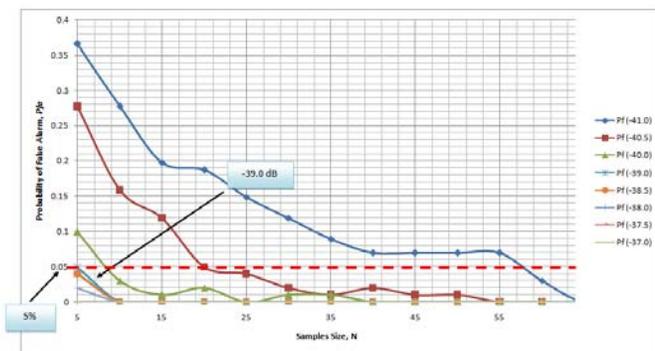


Figure 9. Probability of False Alarm versus Number of Sample

With the set  $P_{fa}$  of 5%, we do the measurement for  $P_d$  by sweeping the signal level from -37.0dB to -39.0dB and the obtained results are plotted in Figure 10. This graph is used to determine the required  $N$  for the desired  $P_d$ . At a predetermined noise threshold of -39.0dB, power received ( $P_r$ ) of -39 dB, -38 dB and -37 dB can be translated to Signal-to-Noise ratio (SNR) values of 0, 1 and 2 dB,

respectively. The higher the  $P_d$ , sample size needed will increase and hence, the longer the sensing time. For instance, for a target  $P_d$  of 90%, at least 35 samples are needed at SNR equals to 0 dB. Longer sensing time will reduce the data transmission time, and thus will result in a lower overall throughput.

Figure 11 shows the theoretical results performance of  $P_d$  against sample size under related SNR conditions. It can be observed that that the experimental results in Figure 10 follow the curve of the theoretical framework as in Fig. 11, especially for target  $P_d$  of 90% which gives sample size needed of 35 at SNR of 0 dB. The difference between the theoretical result and experimental result might be caused by the limitations of the SDR hardware used, the real room environment which differs from the considered theoretical AWGN channel and formula approximation.

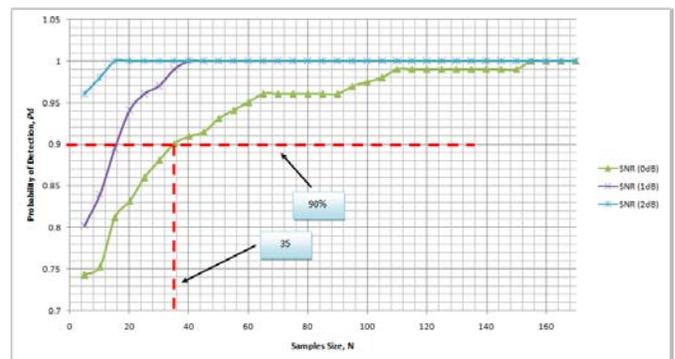


Figure 10. Probability of Detection vs. Number of Sample for Pfa=5%

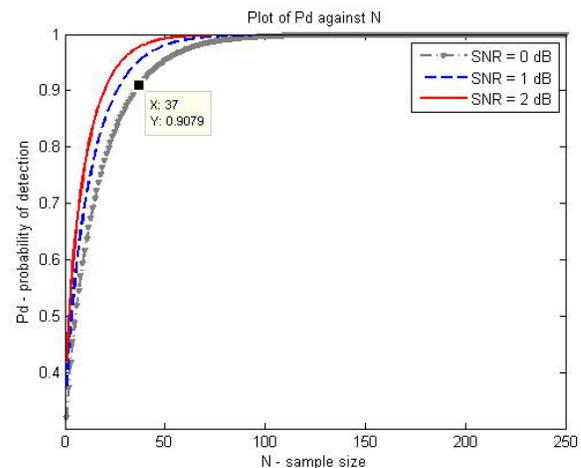


Figure 11. Theoretical Probability of Detection vs. sample size for Pfa=5% under various SNR conditions

In this experimental set-up, the sampling time,  $t$  used is 0.9025ms. This sampling time is obtained by using time stamping on the running designed CU sensing system of Figure 4. Since the experimental sensing performance closely matches that of the theoretical framework at  $P_d$  of 90%, the target Quality of Service (QoS) for  $P_d$  is set at the

value. Hence, as observed in Figure 10, the required number of samples,  $N$ , will be 35. By using (7);

$$T_s = tN \quad (7)$$

the sensing time,  $T_s$  for our GNU Radio USRP based CU system is derived to be 31.59ms. The frame structure for our system is given in Figure 12 where  $T_f$  is the frame period and  $T_f - T_s$  is for data transmission.

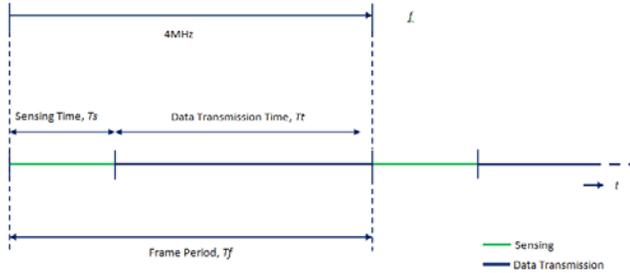


Figure 12. Frame Structure for CU System

According to [14], the frame period,  $T_f$  can be approximated using (8);

$$T_f \approx -\beta_0 \left( 1 + W_{-1} \left( -\exp \left( -\frac{\beta_0 + T_s}{\beta_0} \right) \right) \right) \quad (8)$$

where  $\beta_0$  is the off time of the PU which in this work equals to 650 ms and  $W_{-1}$  is the negative branch of the Lambert's W function[14]. Substituting all the values in the formula will give the  $T_f$  equals to 224.25 ms and hence, transmission time is derived to be equal to 192.66 ms. It can be concluded from the frame structure that sensing time has a direct impact to the data throughput. Longer sensing time will reduce the transmission time, resulted in lower throughput and vice versa.

## VI. CONCLUSION

Two main performance metrics for spectrum sensing were studied; probability of false alarm ( $P_{fa}$ ) and probability of detection ( $P_d$ ).  $P_{fa}$  is used to determine the threshold of the CR DSA system while  $P_d$  is used to determine on how much samples are needed by the CR DSA to meet the desired performance. The bigger the sample size, the longer the sensing time needed for the CU system, hence reducing the throughput. Future works will include attempts to minimize the sensing time using bio-inspired technique for decision making in achieving target QoS.

## ACKNOWLEDGEMENT

The authors wish to express their gratitude to Research Management Center (RMC), Universiti Teknologi Malaysia for the financial support of this project.

## REFERENCES

- [1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next Generation/Dynamic Spectrum Access/Cognitive Radio Wireless Networks: A Survey". The International Journal of Computer and Telecommunications Networking vol. 50, pp. 2127 – 2159, 2006.
- [2] J. Mitola, "Software radios-survey, critical evaluation and future directions". IEEE National Telesystems Conference, pages 13/15-13/23, 19-20 May 1992.
- [3] D. Cabric, S. M. Mishra, R. W. Brodersen, "Implementation Issues in Spectrum Sensing", Asilomar Conference on Signal, Systems and Computers, November 2004
- [4] R. A. Rashid, N. M. Aripin, N. Faisal and S. K. S. Yusof, "Sensing Period Considerations in Fading Environment for Multimedia Delivery in Cognitive Ultra Wideband", International Conference on Signal and Image Processing Applications (ICSIPA 2009), Kuala Lumpur, 18-19 Nov 2009
- [5] N. Manicka, "GNU Radio Testbed", Thesis for Master of Science in Computer Science, University of Delaware, Spring 2007.
- [6] A. Crohas, "Practical Implementation Of A Cognitive Radio System For Dynamic Spectrum Access", Master of Science in Electrical Engineering Thesis. Notre Dame, Indiana, July 2008
- [7] M. Ettus, Universal software radio peripheral. <http://www.ettus.com>.
- [8] USRP RFX diagrams, <http://gnuradio.org/redmine/wiki/gnuradio/UsrpRfxDiagrams>
- [9] GNU Radio API list, <http://gnuradio.org/doc/doxygen/index.html>
- [10] D. Shin, "A Dictionary of the GNU Radio blocks", 21 August 2005.
- [11] P. Courtois. "Windowing: Optimizing FFTs Using Window Functions". National Instruments Tutorial, 3 September 2008.
- [12] M. A. Sarijari, R. A. Rashid, N. Faisal, A. C. C. Lo, S. K. S. Yusof, N. H. Mahalin, "Dynamic Spectrum Access Using Cognitive Radio Utilizing GNU Radio and USRP", 26<sup>th</sup> Wireless World Research Forum (WWRF26), Doha, Qatar, 11-13 April, 2011
- [13] H. V. Poor, An Introduction to signal detection and estimation, 2<sup>nd</sup> ed. Springer-Verlag, New York, 1994
- [14] Y. Pei, A. T. Hoang, Y. C. Liang, "Sensing-Throughput Tradeoff in Cognitive Radio Networks: How Frequently Should Spectrum Sensing be Carried Out?", IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications, 2007. PIMRC 2007. Athens. 3-7 Sept. 2007. pp. 1 - 5.

# Secure, Privacy-Preserving, and Context-Restricted Information Sharing for Location-based Social Networks

Michael Dürr, Philipp Marcus, and Kevin Wiesner

*Mobile and Distributed Systems Group*

*Department for Informatics*

*Ludwig-Maximilians-University Munich*

*80538 Munich, Germany*

*Email: {michael.duerr, philipp.marcus, kevin.wiesner}@ifi.lmu.de*

**Abstract**—Today, *Online Social Network (OSN) providers like Facebook or Google provide their users with web-based Location-based Services (LBSs) like Facebook Places or Google Latitude. To facilitate proactive LBS, most OSN providers also offer Location-based Social Network Services (LB-SNSs) for mobile platforms like the iPhone, Android, or Symbian. A multitude of LB-SNSs focus on proximity detection of nearby friends. However, at present almost non of these implementations guarantees a sufficient degree of information security and privacy as well as location integrity for its users. In this paper, we address these shortcomings in two ways. We define a set of requirements we deem indispensable to enable LB-SNSs that facilitate secure and privacy-preserving information sharing and guarantee for location integrity at the same time. We then present an extension to our decentralized OSN architecture Vegas which incorporates these requirements and therefore supports the secure, privacy-preserving, and location-restricted provision of LB-SNSs.*

**Keywords**—*Online Social Networks; Proximity Detection; Location Integrity; Privacy; Security.*

## I. INTRODUCTION

At present we observe a remarkable increase of *Location-based Social Network Services (LB-SNSs)* for mobile platforms like Android, Blackberry, or the iPhone operating system. One reason for this development represents a steadily increasing participation in Online Social Networks (OSNs) like *Facebook Places, Foursquare, Gowalla, or Google Latitude*. However, it is the implementation of such services for mobile platforms that provokes the rocketing success of such services. Powerful and easy-to-use location APIs for mobile phones as well as their precise positioning technologies like GPS can be used to implement a multitude of Location-based Services (LBSs). This also facilitates proximity detection, a frequently implemented concept in the field of proactive LBS. An example are buddy trackers that generate an event in case a friend is located within a predefined distance.

To detect proximity of friends, a user has to reveal information about his current location. Unfortunately, most people are not aware of the implications and threats caused by the distribution of location information. However, recent

studies [1], [2] have shown that, as people gain a deeper knowledge about how LB-SNSs work, their demand for security and privacy increases significantly. In association with LB-SNSs that focus on proximity detection of nearby persons, one has to consider the following aspects:

1) *Friendship Relevance*: Although some OSN providers like Facebook support fine-grained privacy settings, most of their users do not understand how to properly apply them. Once enabled friends automatically get access to location information although the relevance of this relationship can change over time. Therefore, location information can become accessible to a multitude of persons that are considered outdated friends for a long time.

2) *Third Party Providers*: Most OSN providers take the role of a trusted third party. However, even in case OSN providers promise secure and confidential application of personal data, they cannot guard against security holes caused by software bugs or data abuse by employees with criminal intent. Becoming the victim of a crime due to the unsolicited dissemination of location information must not be possible at all.

3) *Location Integrity*: At present, a user that participates in a LB-SNS focusing on proximity detection cannot verify that location information has not been forged. This security hole could be exploited for location data mining. For instance, an adversary could distribute forged location information and thereby trigger proximity alerts at another user *A*. Assuming a LB-SNS requires two users *A* and *B* to share their information, the proximity alert at *A* would trigger the distribution of his shared information to *B*. The question that has not been answered yet is the following: How can a user restrict his participation in an LB-SNS such that only persons whose physical proximity he can verify are able to get access to any of his shared information?

In this paper we address these problems in the scope of LB-SNSs that focus on the detection of nearby persons with similar interest and that want to share a certain type of information. An example of such a service is *location-based ride sharing* which allows its subscribers to find each other at a railway station to share a train ticket. Another example

is a *location based coupling* service which allows its users to receive profiles of each other when visiting the same single party.

Generalizing these examples of LB-SNSs, we aim at a solution to share any kind of information within a predefined proximity in a secure, privacy-preserving, and location-restricted way.

The contribution of our paper is twofold. We present *a*) a requirements analysis for secure, privacy-preserving, and location-restricted LB-SNS, and *b*) an extension to our decentralized OSN architecture Vegas [3] which facilitates the implementation of LB-SNS that comply with our requirements for location integrity.

The paper is organized as follows. We present our requirements analysis in Section II and a detailed use case in Section III. Section IV gives a short overview of Vegas and details our extension for LB-SNS provision. Section V discusses our approach. Related work is presented in Section VI and Section VII concludes the paper.

## II. REQUIREMENTS

In our previous work [3], we identified four major requirements a secure and privacy-preserving OSN has to fulfill. These requirements encompass a user's *informational self-determination*, *strong trust relationships* between friends, anywhere and anytime *profile availability*, and transparent *mobility support*. However, to provide for secure, privacy-preserving, and location-restricted LB-SNSs, we must consider further requirements specific to location.

### A. Context-Dependent LB-SNS Provision

LB-SNSs like Facebook Places allow their users to receive push notifications on their mobile device as soon as a friend initiates a *check-in event*. A check-in event is generated each time a friend registers his current location within his Facebook profile. Unfortunately, virtually no LB-SNS (including Facebook Places) offers a mechanism that allows its users to validate location information of friends. This facilitates an easy violation of location integrity as an adversary could easily trigger unsolicited LB-SNS alerts by pretending to be located at an arbitrarily chosen place. We therefore identify the requirement of *Context-Dependent LB-SNS Provision*: It must be possible to restrict LB-SNS provision by location-restricted context information. Communication between two participants must not be activated until both are located within their predefined proximity and mutually proved their physical presence. This proof could be based on frequently alternating context information exclusively available to participants at the given location. As a user can restrict LB-SNS provision to a specific context, an adversary physically absent lacks this information. Hence, he is no longer able to pretend proximity.

### B. User-Selective LB-SNS Provision

A repeatedly criticized aspect of LB-SNSs are inadequate capabilities to restrict access to shared information. Facebook Places represents an extreme example, as it allows even a user to publish private location information of his friends. Hence, a user loses control over his informational self-determination [3] as well. Even in case this feature was disabled, a user cannot selectively restrict access to his location information to certain friends and has no influence on how these friends distribute his location information. To address such problems we define the requirement of *User-Selective LB-SNS Provision*. In order to increase privacy, user must be able to define a certain persons or a subset of persons in advance to restrict the provision of LB-SNS services.

### C. Decentralized LB-SNS Provision

At present, virtually every LB-SNS is operated in a centralized way. Although most providers guarantee compliance with more or less restrictive privacy policies, we observe recurring privacy leaks due to unconscious software development or employees' criminal intent. To eliminate this problem, personal information must be under control of its originator. This also implies that each user should be able to control when to trigger the provision of a LB-SNS. We formulate the requirement for *Decentralized LB-SNS Provision* as unsolicited distribution of personal data by a third party must not be possible. It is obvious that this requirement also helps to achieve compliance with requirement II-A as each user is responsible for the distribution of his shared information.

## III. USE CASE

To better understand our design principles, we describe a sophisticated LB-SNS usage scenario that we reference throughout this paper.

A university organizes a conference that focuses on security and privacy aspects in mobile networks. In conjunction with the obligatory registration process, the conference committee decides to offer a voluntary subscription to their conference LB-SNS. They want to provide interested conference participants with a *location-based business card browser* for mobile devices. The idea of the browser service is to allow conference participants to browse each other's business card information. The browser service should be location-restricted i.e it only displays present participants and dynamically updates visible profile information of recently arriving participants.

To meet the aforementioned requirements we define further constraints:

- The browser service must not be usable anywhere else except at the conference location (req. II-A).

- Business cards are broadcasted automatically as a participant arrives at the conference (req. II-A and req. II-C).
- The visibility of business cards must be restricted to subscribers only (req. II-B).
- Apart from the subscription process, the browser service should work in a decentralized way (req. II-C).

#### IV. SYSTEM ARCHITECTURE

In the following, we present our extension in order to implement secure, privacy-preserving, and location-restricted LB-SNSs. We decided to build our solution on top of Vegas, a decentralized secure and privacy-aware OSN that has been developed on the design principles of our previous work [3]. We first review some parts of Vegas that are relevant to understand the functionality of our approach, before we delve into the details of our design.

##### A. Vegas Design

Vegas represents an OSN architecture that focuses on its users' security and privacy-demands. Its design was motivated by a set of requirements that we identified as inevitable in a secure and privacy preserving OSN. These requirements encompass a user's *informational self-determination*, *strong trust relationships* between friends, anywhere and anytime *profile availability*, and transparent *mobility support*.

The Vegas core concept does not support communication between participants that are not directly connected by an edge of the underlying social graph. This restriction is motivated by a problem we termed *social network pollution*. To give but a few examples of social network pollution, present OSNs offer the possibility for search operations on their social graphs, provide unsolicited friendship recommendations, and offer support for non-authorized linkage of a friend's friends. This causes a multitude of unwanted friendship establishments, i.e., links in the social graph which not necessarily represent a real friendship.

Figure 1 illustrates the fundamental communication principles and components of Vegas. In Vegas each user interacts with the OSN through one or more mobile or stationary clients. In order to support delay-tolerant network communication, we apply an asynchronous message exchange scheme based on the concept presented in [4]. We rely on well known services like email, SMS, or instant messaging which can be exploited to implement the *exchanger* instance. An exchanger represents the abstract concept of a message queue which is used to transmit messages or any other kind of content. Any two Vegas friends  $A$  and  $B$  are aware of one or more such exchanger addresses of each other. In addition, since a user cannot be expected to be permanently online, we introduced the *datastore* component. A datastore represents the abstract concept of a user-writable storage space with world-readable access. A datastore can be simply

implemented through some simple-to-administer and cost-free web space. Each user provides one or more datastores to place his user profile encrypted and signed for each of his friends. As we will see in the next section, any kind of

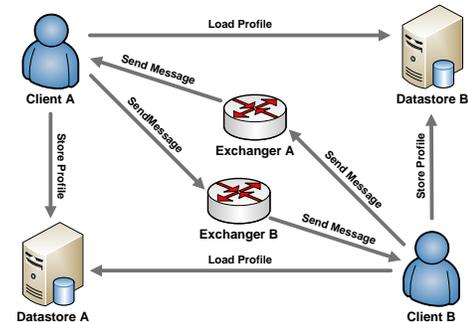


Figure 1. In Vegas any piece of information exchanged between users is encrypted and signed. Each user maintains one or more client instances and performs encrypted messaging over one or more exchanger instances. A user publishes individual profiles for each friend at one or more datastores.

information flow in Vegas is subject to cryptographic operations. Every message as well as each profile are separately encrypted for each friend.

##### B. Vegas Operation

In a nutshell, Vegas messaging and profile distribution works as follows: Any two Vegas friends  $A$  and  $B$  generate a unique public key pair which must not be applied for messaging and profile generation except in the context of  $A$  and  $B$ . In the following, we term such a key pair a *link-specific* key pair. As user  $A$  holds a unique key pair  $K_{A \rightarrow X_i}^- / K_{A \rightarrow X_i}^+$  ( $i \in 1, \dots, n$ ) for each of his  $n$  friends  $X_1, \dots, X_n$ , a key pair represents nothing else than a directed edge in the overall social graph. The notion of a key  $K_{A \rightarrow X_i}^{[-|+]}$  means that this key is a private/public key generated by  $A$  for exclusive communication with  $X_i$ .  $A$  utilizes  $X_i$ 's public key  $K_{X_i \rightarrow A}^+$  to encrypt messages as well as profile information intended for  $X_i$ . In order to allow  $X_i$  to map a received message to its originator  $A$ , a fingerprint of  $A$ 's public key  $K_{A \rightarrow X_i}^+$  is included into each message sent to  $X_i$ . In case  $A$  wants to send a message to  $X_i$ ,  $A$  applies  $X_i$ 's public key  $K_{X_i \rightarrow A}^+$  to encrypt the message content. After signing the message with  $K_{A \rightarrow X_i}^-$ ,  $A$  sends this message to one of  $X_i$ 's exchangers. Now  $X_i$  can fetch this messages and identify sender  $A$  through his attached public key fingerprint. Since  $X_i$  is the only user that knows about the mapping of the included fingerprint,  $X_i$  represents the only user that is able to map this fingerprint to the identity of  $A$ . In case  $A$  considers the mapping of a public key to  $X_i$ 's identity compromised,  $A$  can trigger a key refresh operation in order to replace all former key pairs shared with  $X_i$ .

We apply the same Vegas operations for the placement and update of profile information which we use to send

messages. In case user  $A$  wants to update his profile,  $A$  re-encrypts the corresponding information for each of his friends  $X_1, \dots, X_n$  and places this information at the corresponding datastores. As  $A$  utilizes the same key material to encrypt messages as well as profile information for  $X_i$ ,  $X_i$  can simply perform an interval-based read operation at  $A$ 's data store in order to receive pending updates. As an alternative,  $A$  can trigger such an update by sending each of his friends a corresponding notification.

It should be stressed that a user  $A$  can cancel a friendship with user  $X_i$  by simply deleting the link-specific key pair. Hence, key revocation does not involve complex maintenance and distribution of key revocation lists.

### C. Directory Buddies

Our architecture already includes the basic functionality necessary to support reactive LBSs comparable to those provided by Gowalla, Foursquare, Google Latitude, or Facebook Places. Such LBSs can be easily implemented by extending profile descriptions with location information and introducing location-dependent profile updates. To enable sophisticated LB-SNS applications like a location-based ride sharing, location-based coupling, or location-based business card browsing, we have to extend Vegas as it disallows the distribution of information to other users except Vegas friends.

To guarantee compliance with the requirements formulated in section II, we decided to extend Vegas by the concept of *directory buddies*. A directory buddy represents a special kind of Vegas friend that supports the establishment of obfuscated connections between all participants of a LB-SNS. Although not a requirement, it is likely that a directory buddy is not operated by a single person but an institution or a company that can provide additional contents. In essence, it is the task of a directory buddy to provide for anonymous connections between foreign persons. Figure 2 details how a directory buddy integrates into Vegas. A Vegas user  $X_i$  that wants to subscribe for a LB-SNS  $L$  first has to establish a friendship with a directory buddy  $C$  associated with  $L$ .  $X_i$  and  $C$  require a (semi-) trusted out-of-band (OOB) channel to exchange their public keys ( $K_{X_i \rightarrow C}^+$ ,  $K_{C \rightarrow X_i}^+$ ), their exchanger addresses ( $Ex_{X_i}$ ,  $Ex_C$ ), and their datastore addresses ( $DS_{X_i}$ ,  $DS_C$ ) (1). In case  $C$  does not require detailed profile information of  $X_i$  to support the provision of  $L$ , exchanger and datastore addresses of  $X_i$  need not necessarily be exchanged. A secure exchange of this information is out of scope of this paper. An example for the application of an email-based OOB channel for this purpose can be found in [3]. Considering our use case from section III, the establishment of a friendship with  $C$  corresponds to the subscription process with  $L$ .

A LB-SNS provider determines a certain point in time at which  $C$  initiates its service provision for currently registered users (2). Assuming  $n$  users  $X_1, \dots, X_n$  registered

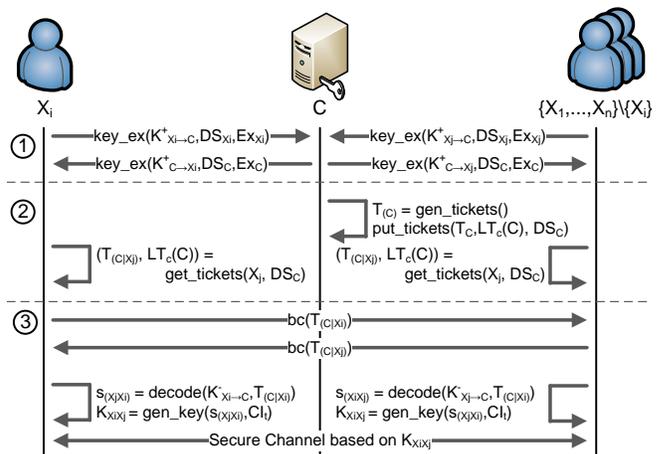


Figure 2. Message exchange between a directory buddy and its subscribers. After users have registered (1) the directory buddy generates all necessary tickets and places them for pickup at his datastore (2). As a users location task activates the LB-SNS, he broadcasts his tickets (3). Each user can decrypt tickets encrypted with his public key and create a secure channel.

for  $L$ ,  $C$  calls the *gen\_tickets()* function to generate a set of tickets  $T_{(C)} = \{t_{(X_i X_j)} | 1 \leq i, j \leq n \wedge i \neq j\}$  for each ordered combination of users  $X_i$  and  $X_j$ . A ticket  $t_{(X_i X_j)} = K_{X_j \rightarrow C}^+(s_{(X_i X_j)} || cs(s_{(X_i X_j)}))$  includes some unique piece of information  $s_{(X_i X_j)}$  (where  $s_{(X_i X_j)} = s_{(X_j X_i)}$ ) which is generated by  $C$  to provide two registered users  $X_i$  and  $X_j$  with a common secret.  $cs(s_{(X_i X_j)})$  represents the checksum of  $s_{(X_i X_j)}$ .  $C$  generates a *location task*  $LT_c(C)$  for each user  $X_i$ . A location task corresponds to the description of some context  $c$  a user  $X_i$  has to observe before  $X_i$  initiates his active participation in the corresponding LB-SNS. Referring to our use case,  $LT_c(C)$  could be defined as the geographic dimensions of the conference building. In this example each participant would receive the same location task. Now  $C$  calls the *put\_tickets()* function to place disjoint subsets  $T_{(C)|X_i} = \{t_{(X_i X_j)} | 1 \leq j \leq n \wedge i \neq j\} \subset T_{(C)}$  and the location task  $LT_c(C)$  individually encrypted based on  $K_{X_i \rightarrow C}^+$  for each user  $X_i$  at datastore  $DS_C$ . Dependent of a preconfigured interval,  $X_i$  calls the *get\_tickets()* function to receive his private subset of tickets  $T_{(C)|X_i}$  and the location task  $LT_c(C)$  from  $DS_C$ .

When  $X_i$  observes the occurrence of  $LT_c(C)$ ,  $X_i$  starts to broadcast  $T_{(C)|X_i}$  (3). For instance, in case of our usage scenario from section III, the context  $c$  corresponds to the arrival at the conference building. Any other registered user  $X_j$  ( $i \neq j$ ) who already observed the occurrence of  $LT_c(C)$  also broadcasts his ticket subset  $T_{(C)|X_j}$ . When  $X_i$  receives a broadcast from  $X_j$ ,  $X_i$  calls the *decode()* function to decode the secret  $s_{(X_i X_j)}$  shared between  $X_i$  and  $X_j$ . In case there exists a ticket  $t_{(X_i X_j)} \in T_{(C)|X_j}$ ,  $X_i$  can extract the secret  $s_{(X_i X_j)}$  by applying his private key  $K_{X_i \rightarrow C}^-$  to  $t_{(X_i X_j)}$ . As  $X_i$  has no previous knowledge about the public key that was used to encrypt a certain ticket,  $X_i$  has to

calculate the checksum of an encrypted secret in order to determine that this secret corresponds to  $s_{(X_i X_j)}$ . Section V discusses this process in detail. It should be stressed that, although  $X_i$  and  $X_j$  now share a common secret,  $C$  is always able to decrypt information which was previously encrypted based on  $s_{(X_i X_j)}$ .

Although  $X_i$  and  $X_j$  cannot trigger their participation in a LB-SNS before they were able to exchange their tickets, i.e., before they are located within the same broadcast domain, the shared secret  $s_{(X_i X_j)}$  does not suffice to give evidence for the occurrence of  $LT_c(C)$ . For instance, a fraudulent user  $M$  could pretend to be located within the broadcast domain physically restricted by  $LT_c(C)$ .  $M$  could connect to the corresponding subnetwork by establishing a VPN tunnel link or driving a wormhole attack. To generate a secret key from  $s_{(X_i X_j)}$  that cannot be determined before  $X_i$  and  $X_j$  share the same broadcast domain and are physically located at an area described by  $LT_c(C)$ , we require an additional context information  $CI_t$  which cannot be determined except at the location described by  $LT_c(C)$ . Referring to our use case, such additional context information could be an identifiable but randomly chosen and frequently alternating essid of the wireless conference network. A shared secret key  $K_{(X_i X_j)}$  could then be generated by the *gen\_key()* function based on the shared secret  $s_{(X_i X_j)}$  and a context information  $CI_t$ . As decryption could fail due to the application of an outdated  $CI_t$  for the generation of  $K_{(X_i X_j)}$ ,  $X_i$  and  $X_j$  temporarily store recently expired keys. This facilitates the decryption of messages that were encrypted with a predecessor of  $K_{(X_i X_j)}$  although  $K_{(X_i X_j)}$  represents the present key.

It is worth mentioning that the generation and distribution of tickets in step (2) of Figure 2 is not necessarily bound to a fixed point in time. This process can be executed incrementally. In case a new participant subscribes with the LB-SNS,  $C$  updates all ticket subsets  $T_{(C|X_i)}$  for each participant  $X_i$  on demand and places them at  $DS_C$ . Eventually,  $X_i$  receives this update in dependence of his pre-configured update interval.

## V. DISCUSSION

Our concept of directory buddies provides the foundation for secure, privacy-preserving, and context-restricted LB-SNSs. In this section we discuss compliance with our requirements from section II. Furthermore we give some notes on performance and security.

### A. Compliance with Our Requirements

In this work we extended the existing OSN architecture Vegas which already meets the requirements to provide for a secure and privacy-preserving OSN [3]. In the following, we only discuss requirements II-A, II-B, and II-C.

As the provision of a LB-SNS is based on a location task  $LT_{(C)}$ , no information is shared before a user observes the

occurrence of  $LT_{(C)}$ . Furthermore, LB-SNS provision is restrict to participants that have knowledge of some predefined additional context information. To share information, two users have to provide each other with an evidence that they fulfill  $LT_{(C)}$ , i.e., they are located within physical proximity. Hence, our approach meets the requirement for Context-Dependent LB-SNS Provision. It should be mentioned, that our requirement for Context-Dependent LB-SNS Provision impedes security attacks as an adversary has to be physically present or at least needs an in situ accomplice.

Each user that wants to participate in a LB-SNS has to subscribe with a directory buddy in advance. This component only serves for the establishment of anonymous links between users which they can use to provide each other with encrypted information, i.e., no identities are revealed. As each link is secured with a separate secret, a user can provide another user with additional information only shared between both of them. Therefore, we achieve User-Selective LB-SNS Provision.

As each subscriber has control over the point in time when to trigger his LB-SNS participation, he has also full control over the provision of shared information. As all information is broadcasted by each user himself, we achieve a completely decentralized, fully meshed communication scheme. Therefore, our solution adheres to the requirement for Decentralized LB-SNS Provision.

### B. Comments on Performance

As mentioned before, our design applies one ticket per anonymous link. In case a user  $X_i$  receives a ticket broadcast  $T_{(C|X_j)}$  from  $X_j$ ,  $X_i$  has to apply his private key  $K_{X_i \rightarrow C}^-$  to each included ticket in order to determine the shared secret  $s_{(X_i X_j)}$ . Hence, the process to decrypt each ticket  $s_{(X_i X_j)}$  for all subscribers  $X_j \in \{X_1, \dots, X_n\} \setminus \{X_i\}$  has a complexity of  $O(n^2)$ . However, we can decrease the complexity to  $O(n)$  if we add an identifier to each ticket. For instance  $C$  could determine a set of unique identifiers for each ticket  $t_{(X_i X_j)}$ , add one identifier to each ticket, and announce relevant identifiers within each subscriber's profile information at  $DS_C$ .

Except the necessity of a single broadcast domain, LB-SNSs are completely independent of the underlying network. Nevertheless it should be mentioned that, in case of an ad hoc network, a LB-SNS could suffer from packet loss due to frequent ticket broadcasts. Dependent of the MAC layer, subscribers should carefully choose a suitable broadcast interval.

### C. Comments on Security

Since tickets are distributed via broadcast, it takes an adversary no effort to replay once received ticket broadcasts. However, this does not represent a security problem as a ticket  $t_{(X_i X_j)}$  cannot be replayed before  $X_j$  sent the ticket for the first time. Since  $X_i$  maintains a list of all decrypted

tickets, replayed tickets are simply ignored as soon as  $X_i$  received  $t_{(X_i, X_j)}$  for the first time. As long as a directory buddy can be trusted, our solution facilitates secure, privacy-preserving, and location-restricted LB-SNSs.

## VI. RELATED WORK

A plethora of research has been conducted in the field of secure and privacy-aware LBS. In the following, we review related work from the area of privacy-aware proximity detection and location-based access control as we consider it the most relevant fields in our context.

Many proposals have been published for privacy-aware proximity detection for LBS. (e.g., [5]–[7]). However, it has been shown [8] that enabling privacy-awareness for LBSs is a challenging task. Recently, Šikšnys et al. presented their *vicinitylocator* [9], a proposal for private and flexible proximity detection in mobile social networks. Their architecture supports proximity detection by checking for inclusion of one user's location inside another user's vicinity. A server instance attempts to map encrypted proximity regions (granules) presented by one user with the granules presented by another one. Although this scheme facilitates a better degree of privacy, proximity detection is still performed server-sided, i.e. a user cannot validate location integrity due to the lack of a common context. Zhong et al. developed different protocols [10] to support privacy-preserving proximity-detection of nearby friends. However, even in case they apply a trusted third-party to avoid situations where friends learn locations of users that must no longer be considered nearby, their approach cannot guarantee location integrity. A paper recently published by Puttaswamy et al. [11] probably shares most similarities with our approach. It describes a decentralized approach to enable privacy-preserving location-based mobile social applications. Mobile devices place their encrypted location data at third-party servers. Participants of the social network may download and decrypt that information in case they share a secret key with the data originator. Although their approach facilitates decentralized proximity detection it does not adhere to our requirement for location integrity. Users can preserve privacy but they cannot prove location integrity.

Some work published by Sastry et al. [12] focuses on the *secure in-region verification problem*. Their approach is related to our location integrity problem, as a prover has to prove its claimed location to a verifier. To address this problem the authors measure emitted ultrasound signals to verify a location. As recent mobile devices are not equipped with ultrasound measurement facilities, their approach will not be applicable in the near future.

## VII. CONCLUSION

We presented our approach for secure, privacy-preserving, and location-restricted location-based services for social networks (LB-SNS) that focus on information sharing with

nearby persons. We first elaborated a set of requirements we deem indispensable to provide for the provision of LB-SNSs. We presented the concept of directory buddies, an extension to our decentralized, secure, and privacy-preserving OSN architecture Vegas and illustrated its application in context of LB-SNSs by exemplifying a location-based business card browser. We showed that this design fully complies with our requirements.

Our future work will focus on the evaluation of our design in a realistic setup. Furthermore, we will investigate novel possibilities to infer context information to provide for location integrity from environmental impacts.

## REFERENCES

- [1] J. Tsai, P. Kelley, L. Cranor, and N. Sadeh, "Location-sharing technologies: Privacy risks and controls," *ISJLP*, vol. 6, pp. 119–317, 2010.
- [2] K. Tang, J. Lin, J. Hong, D. Siewiorek, and N. Sadeh, "Re-thinking location sharing: exploring the implications of social-driven vs. purpose-driven location sharing," in *Proceedings of UbiComp'10*. ACM, 2010, pp. 85–94.
- [3] M. Dürr, M. Werner, and M. Maier, "Re-Socializing Online Social Networks," in *Proc. of CPSCOM'10*. IEEE, Dec 2010.
- [4] M. Werner, "A Privacy-Enabled Architecture for Location-Based Services," in *Proc. of MobiSec '10*, 2010.
- [5] P. Ruppel, G. Treu, A. Küpper, and C. Linnhoff-Popien, "Anonymous User Tracking for Location-Based Community Services," in *Location- and Context-Awareness*, ser. LNCS. Springer Berlin/Heidelberg, 2006, vol. 3987, pp. 116–133.
- [6] A. Küpper and G. Treu, "Efficient proximity and separation detection among mobile targets for supporting location-based community services," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 10, pp. 1–12, July 2006.
- [7] S. Mascetti, C. Bettini, and D. Freni, "Longitude: Centralized Privacy-Preserving Computation of Users Proximity," in *Secure Data Management*, ser. LNCS. Springer Berlin/Heidelberg, 2009, vol. 5776, pp. 142–157.
- [8] A. Khoshgozaran and C. Shahabi, "Blind evaluation of nearest neighbor queries using space transformation to preserve location privacy," in *Proc. of SSTD '07*. Springer Berlin/Heidelberg, 2007, pp. 239–257.
- [9] L. Šikšnys, J. Thomsen, S. Šaltenis, and M. Yiu, "Private and Flexible Proximity Detection in Mobile Social Networks," in *Proc. of MDM 2010*. IEEE, 2010, pp. 75–84.
- [10] G. Zhong, I. Goldberg, and U. Hengartner, "Louis, Lester and Pierre: three protocols for location privacy," in *Proc. of PET '07*. Springer Berlin/Heidelberg, 2007, pp. 62–76.
- [11] K. P. N. Puttaswamy and B. Y. Zhao, "Preserving privacy in location-based mobile social applications," in *Proc. HotMobile '10*. New York, NY, USA: ACM, 2010, pp. 1–6.
- [12] N. Sastry, U. Shankar, and D. Wagner, "Secure verification of location claims," in *Proc. of WiSe '03*. New York, NY, USA: ACM, 2003, pp. 1–10.

## Seamless Multimedia Handoff for Hierarchical Mobile IPv6

Khaled Zeraoulia

Department of Computer Science.  
University of Sciences and Technology Houari  
Boumediene (USTHB)  
Algiers, Algeria  
Khaled.zeraoulia@gmail.com

Nadjib Badache

Research Center on Scientific and Technical  
Information  
CERIST  
Algiers, Algeria  
nbadache@Cerist.dz

**Abstract**—The majority of today’s multimedia applications and content reside on the Internet. Since the Internet Protocol (IP) is the “glue” that holds the Internet together, there is enormous interest in extending IP to mobile multimedia networks. Handover is one of the most important factors that may degrade the performance of real-time applications in mobile multimedia networks. In this paper, we introduce a novel mobility management strategy for mobile IP networks, in which we develop a seamless handover scheme called SHMIPv6 (Seamless Multimedia handoff for hierarchical Mobile IPv6). By integrating MAC and Network layer handovers efficiently, SHMIPv6 can significantly reduce the system signaling cost and handover delay. Also, our performance study shows that SHMIPv6 achieves loss-free packet delivery using an anticipated buffering scheme.

**Keywords**- *Seamless handover; HMIPv6; multimedia applications.*

### I. INTRODUCTION

Mobile IP is the key protocol in providing mobile multimedia applications to mobile devices [1][2]. These applications concern all modern life aspects: peer-to-peer communications, video conferences, remote learning, etc. Several standard bodies such as the IETF [2] are working on the specifications of all IP wireless networks that allow roaming users to access integrated data, voice, and multimedia services. A wide variety of mobility management schemes have been proposed [5][9][12][14][15] working at different layers of the protocol stack. However, these schemes are not suitable for all types of applications. Thus, dedicated schemes capable of acting more semantically must be developed; e.g., in the case of HTTP or FTP applications, handoff latency is not of vital importance (waiting one or two seconds extra when downloading a web page is not critical). But for real-time media, latency and packet losses are extremely important and even a small disturbance can make a media stream unintelligible. The mobility requirements can be satisfied by handover solution based on integrating of MAC and Network layer mechanisms[3][12]. Although the mobile IPv6 offers mobility management capability, however, MIPv6 is not suitable for supporting streaming media with stringent delay and eliminates packet losses requirements [6][7]. Therefore, the MIPv6 handoff needs careful investigation to offer reasonable delays and packet loss for mobile multimedia

application. The rest of this paper is structured as follows: handover, MIPv6 and HMIPv6 are analyzed in Section 2. Section 3 introduces the basic idea of the improved scheme SHMIPv6 in detail. Simulation setup and results of performance comparison are provided in Section 4. Finally, concluding remarks are presented in Section 5.

### II. THEORETICAL BACKGROUND

#### A. Handover Management

The process by which a Mobile Node (MN) changes to a new subnet is called handover. Handover in packet networks is administratively costly due to the number of signaling messages involved and the change of state in participating nodes. Although Mobile IP is designed for mobility management in IP networks, it causes a high latency and signaling overhead during handover. Therefore, advanced mobility mechanisms improving Mobile IP are desirable to perform efficient handovers. Also, appropriate Quality of Service (QoS) support is needed for mobility-enhanced Internet Protocol (IP) in order to meet mobile user’s expectations.

##### 1) MAC Layer Handover (Handover Layer 2)

The handover preparation procedure begins when MN moves into the overlapping radio coverage area of two adjacent subnets, it needs to perform a Handover Layer 2 (MAC Layer) to bring to an end the association with the old Access Point and re-associate with new one [5]. This will require some steps such as detection, authentication and re-association with the new Access Point. Only, after these procedures will finish, higher layer protocols can proceed with their signaling procedure, such as layer 3 router advertisements. Once the MN finishes Handover Layer 2 and receives the router advertise from the Router, it should begin to obtain a new router address.

##### 2) Anticipated Handover

Handover is initiated when either the MN or the previous Router have predictive information about the next point of attachment to which the MN will move to Neighbor Discovery for IP Version 6 (IPv6)[11]. Nodes (MN and Routers) use Neighbor Discovery (ND) protocol [11] to determine the link-layer addresses for neighbors known to reside on attached links. Nodes also use Neighbor Discovery to find neighboring routers that are willing to forward packets on their behalf. Finally, nodes use the

protocol to actively keep track of which neighbors are reachable and which are not, and to detect changed link-layer addresses. If the MN has such information, or it chooses to force a handover to a new subnet, it sends a Router Solicitation for Proxy (RtSolPr) to the previous Router, and receives a Proxy Router Advertisement (PRtAdv) in response, providing the MN with the L2 (MAC Layer) information, such as the subnet prefix, link quality, measured bandwidth and available attachments status required for the MN to establish a new Care-of-Address (CoA) on the new subnet [12]. When previous Router receives an indication from L2 that the MN will be moving or RtSolPr indicating that the MN wants to move, the previous Router exchanges messages with new router in order to obtain or validate the new CoA for the MN. The previous Router sends a Handover Initiate (HI) message to the new router. The HI message contains the requested new CoA on the new subnet.

### B. Mobile IPv6 overview

The IETF is in the midst of designing Mobile IPv6 [2]. The true goal of MIPv6 is to offer an end-to-end IPv6 operability between mobile devices. It should be mentioned here that there is a standard for Mobile IPv4 [10]. The basic principle for both MIPv4 and MIPv6 is that all mobile nodes have a permanent IP address on a “home” network. When a mobile node (MN) roams to another subnet, it must first acquire a temporary (CoA) on that network. The next step is to send a binding update (BU) back to a special router on the home network named the Home Agent (HA). A BU associates the CoA with the permanent IP address of the MN. When any other Correspondent Node (CN) sends an IP packet addressed to the permanent IP address of the MN, the HA intercepts the packet and, using the BU, tunnels the packet to the CoA of the MN. At this point, the MN knows the address of the CN by looking at the source address of the packet header. Therefore, the MN does not have to tunnel reply packets to the CN through the HA but can send them directly to the CNs destination address. The MN can elect to send a BU to the CN so that the CN can send optimally routed packets to the MN instead of having to go through the HA. How does MIPv6 discover that the MN has moved to a new subnet? The MIPv6 requirement is that a router supporting Neighbor Discovery [11] must be operational on the subnet. This router sends out a “beacon” packet of 32 bytes called a Router Advertisement [11]. The recommended intervals are from 20 to 1500 milliseconds between advertisements. The first time an MN receives an advertisement, it can examine it to find out what the subnet prefix is and when to expect the next advertisement. If the “old” advertisement is overdue or an advertisement with a “new” subnet prefix shows up then MIPv6 has discovered that it has moved to a new subnet.

### C. Hierarchical MIPv6

Hierarchical mobility management for Mobile IPv6 is designed to reduce the amount of signaling between the MN, its CNs, and its HA [13]. HMIPv6 improves the handover management of basic MIPv6 by introducing a new

protocol agent called Mobility Anchor Points (MAP) [7]. MAP splits the management of the handover process into macro-mobility and micro-mobility and deals with them separately. In HMIPv6, MN assigns two addresses, regional care of address (RCoA) and link care of address (LCoA). These two addresses are very useful managing mobility (figure 1). A MN that enters a foreign network first configures its LCoA by the IPv6 address auto-configuration scheme. The MN then sends a local BU message to the MAP. This local BU message includes the MN's RCoA in the Home Address Option field and the LCoA is used as the source address of the BU message. This BU message binds the MN's RCoA to its LCoA. The MAP then performs duplicated address detection (DAD) procedure for the MN's RCoA on its link and returns a Binding Acknowledgement (BAck) message to the MN.

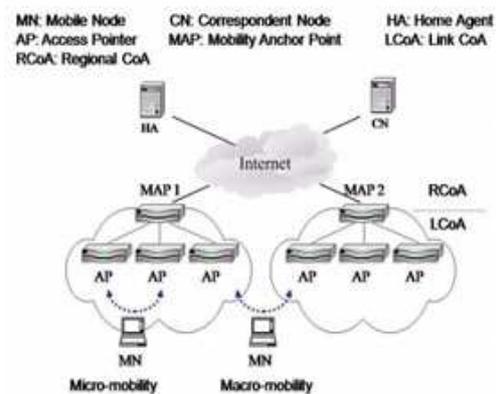


Figure 1. Hierarchical Mobile IPv6

## III. SHMIPv6: SEAMLESS HIERARCHICAL MOBILE IPv6

In this section, we will discuss our proposed scheme (SHMIPv6). Our solution *tries to* provide good *QoS* performance support for *multimedia applications*. The fast handover mechanism using HMIPv6 may guarantee seamless multimedia handover as long as the MN moves inter domain. But it is still not enough to support real time voice services. So we propose a fast buffering schema to reduce packet loss. The aim of SHMIPv6 can be described by the following three parameters:

### A. Reduce signaling overhead

- Using Hierarchical Mobile IPv6 (HMIPv6) reduce signaling overhead and support seamless handoff in IP-based wireless/mobile networks.
- Using several MAPs: Using one MAP keeps large number of packets waiting before it receives or sends them, and this causes long delay and large number of lost packets. The domain is composed of multiple Mobility Anchor Points; each MAP in the domain is attached with an Access Router (AR). The mechanism shares the traffic information among the MAPs in the domain to make decision of MAP reassignment. The MAPs at the domain give the same RCoA.

In HMIPv6 DAD procedure, it takes at least 1000 ms to detect that there is no duplicate address in the link. The regional registration procedure proposed by HMIPv6 offers low handoff delay, but it remains too high for real time applications which require handoff delay to avoid service degradation. So, to reduce total handover latency, SHMIPv6 propose two mechanisms: anticipated handover and Predictive Address Reservation mechanism.

*B. Low latency handover in SHMIPv6 : ( Predictive Address Reservation mechanism)*

With HMIPv6 allows not only reducing handoff delay but also signaling overhead. SHMIPv6 proposes to perform address allocation and registration procedures before the link layer handoff (L2) to reduce HMSIP handoff latency. This can be achieved by employing the movement detection scheme using link layer information (Figure 2). The base idea is to allocate a new IP address to the MN and allow him to re- REGISTER with its MAP (regional registration) using the link layer handoff triggering. In fact, address acquisition and HMIPv6 registration procedures are executed in parallel.

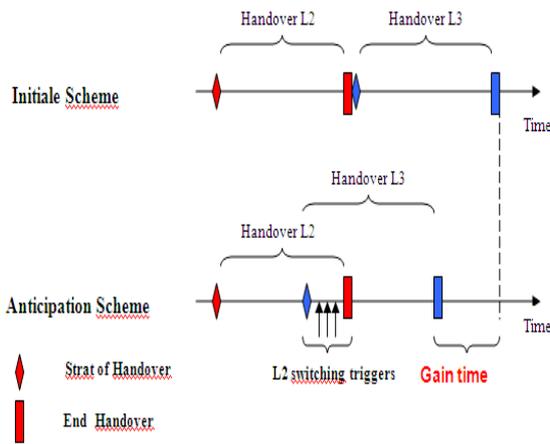


Figure 2. Anticipated Handover

*C. Loss free packet: ( Anticipated buffering process using security threshold)*

To minimize loss packet during handover, SHMIPv6 propose an anticipate buffering process with conditional tunneling. This process enabling the temporary storage of the tunneled before the handover packets, thus eliminating packet loss occurred during the link layer handover period. To do this we define the two following metrics:

1) *LT (Loss threshold)*: A packet may be considered as lost if it is received with signal strength (SS) less than LT (Loss threshold).

2) *ST(Security threshold)*: This threshold may be useful to synchronize with the start of Buffering networks Handover (Handover L3): related to the initial idea was to start buffering with a broadcast message « Handoff Initiate». However, there may be packets loss before sending this

message; we define the security threshold for anticipating the buffering before the signal level deteriorates. Soon as the mobile node receives the signal strength equal to the security threshold, it sends a message Application Control Buffer (RCB) to its old access router. This message acts as initiator of storing packets in the buffer. While the packets are being stored, the old access router sends a copy of these packets until there is a disconnection from the mobile node (reached the loss threshold). Note that the buffering continues until the connection with the new router is established and make a record with the new router.

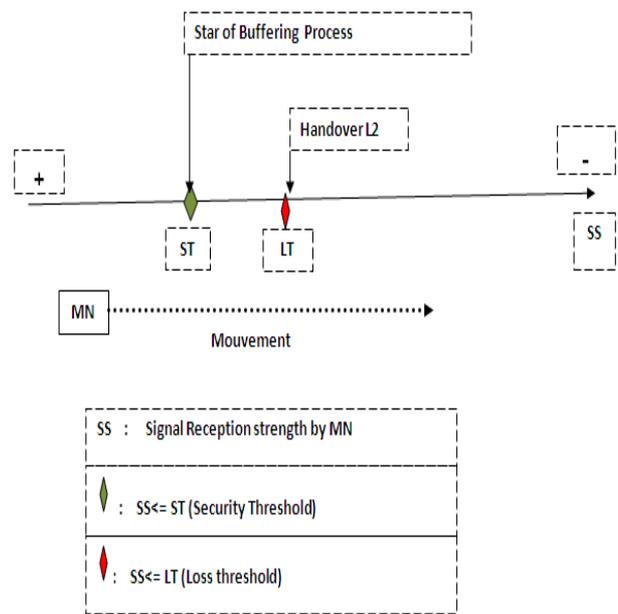


Figure 3. Anticipated Buffering Process

*D. Enhanced Scheme*

SHMIPv6 buffering scheme presented in figure4 can be summarized in five steps:

- *Step1*: Soon as the MN receives packets with:  $SS \leq ST$ , it sends a RCB message to old access router. This message acts as initiator of storing packets in the buffer.

1) *Step 2*: in this step three tasks are performed:

- a) *Handover L2 is initiate* ( $SS \leq LT$ )
- b) *Buffering process is executed in the PR*
- c) *Initialization of tunneling process.*

2) *Step 3*: in this step four tasks are performed:

- a) *Handover L2 is performed*
- b) *Handover L3 is initiate*
- c) *Running buffering process*
- d) *Tunneling parquets from PR to NR*

3) *step 4*: in this step, two tasks are executed

- a) *Handover L3 is performed*
- b) *Tunneling packets from PR to NR*

4) step 5: in this last step, scheduling process is executed for packets received via tunnel (sent from the PR) and packets received from the NR

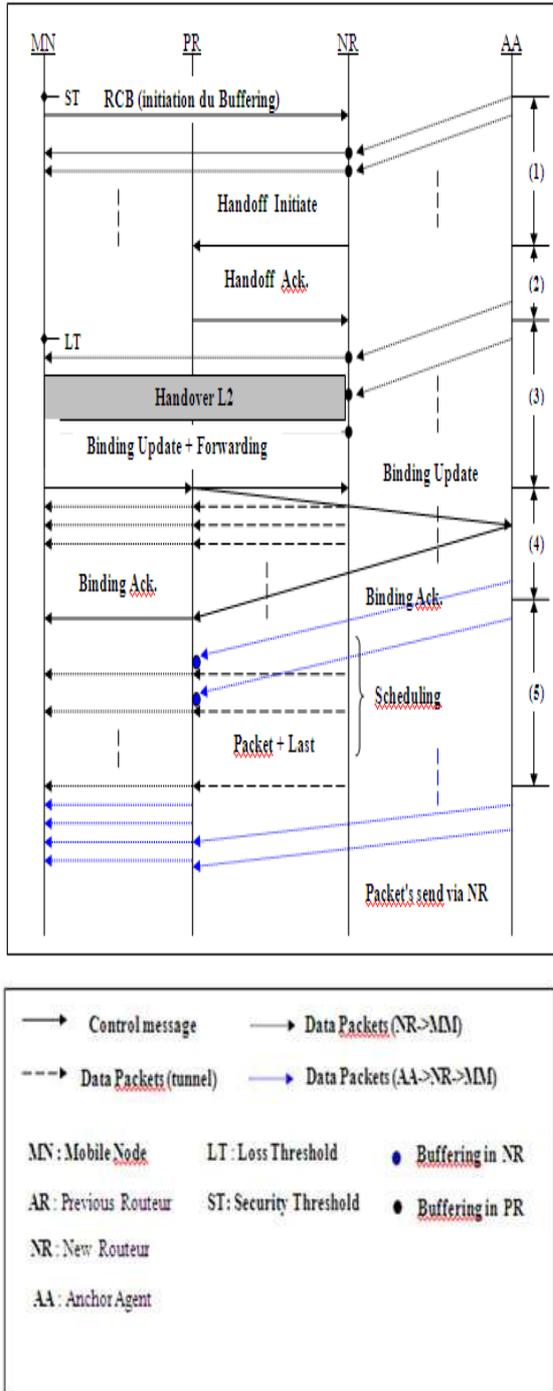


Figure 4. SHMIPv6

IV. SIMULATION RESULTS

Performance results will be provided in terms of handoff latency, packet loss, and jitter. We compare our algorithm against previous Mobile IPv6 propositions [5] [12] [14]

• Simulation Parameters

- Simulator: Ns-allinone-2.34
- Network: 600\*600 and 1000\*1000
- Modulation: DSSS
- Mobil node: 7 and 12
- Bandwidth: 2 mbps
- Traffic generator: Constant Bit Rate and Video
- Packet rate: 13 packet per second
- Packet size: 512 bytes
- Loss Threshold: 3.41828e-10 Watts

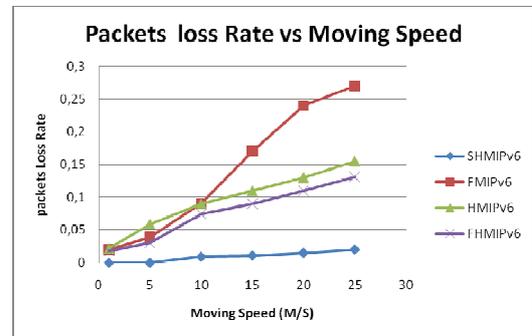


Figure 5. Packet loss rate vs. Moving Speed

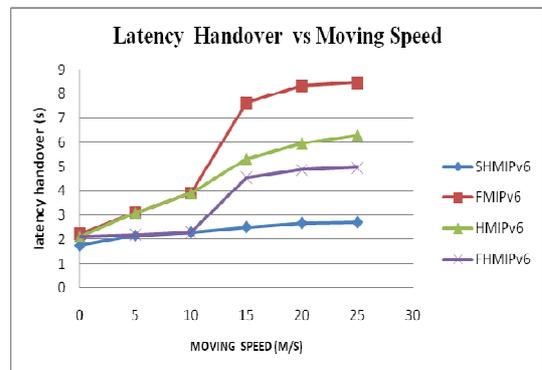


Figure 6. Latency Handover vs. Moving Speed

As we analyzed the performance of our proposed scheme we proved that SHMIPv6 transmits message faster and efficient than FMIPv6, FHMIPv6 and HMIPv6.

Figures 5, 6 show the increase in the handover latency and the packet loss due to an increase in moving speed of MNs. As can be seen, SHMIPv6 approach performs better than FHMIPv6, HMIPv6 and FMIPv6 in terms of the handover latency and packet loss. Although the SHMIPv6 (with the integration of thresholds: ST-LT, Using several MAPs, Predictive Address Reservation) is designed to minimize the packet loss and the latency during a handover, a worse performance is observed with respect to FMIPv6

and MIPv6. In contrast, SHMIPv6 provide a low latency handover (1.75s and 2.7s). This is due to the fundamental difference between handover registrations and anticipation Procedures in SHMIPv6 and other procedures.

Furthermore, the number of packets lost depends on the moving speed of MN. As seen in the figure 6, SHMIPv6 packets lost rate is between (0 and 0.02). This means that the packet loss can totally eliminated if we use an anticipated buffering scheme.

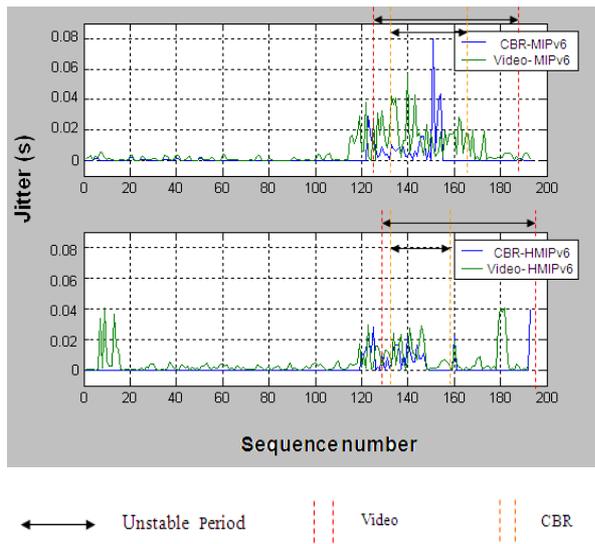


Figure 7. Jitter vs. Sequence number generated

Figure 7 shows the jitter comparison under CBR and video sequence generated. The mean jitter for CBR traffic is 0.08s and 0.038s in MIPv6 and SHMIPv6 correspondingly. In the other hand the mean jitter for video traffic is 0.058 and 0.04 s in MIPv6 and SHMIPv6 correspondingly. Furthermore, we can see that when the payload is light, there is no loss in both two kinds of protocols and for all type of traffic (CBR or video). We can see also, that there is a difference between MIPv6 and the SHMIPv6 under heavy background payload; Payload affects more to the performance of MIPv6 than SHMIPv6. When there is heavy background payload, the loss rate of MIPv6 is bigger than that of SHMIPv6.

## V. CONCLUSION

Both Hierarchical Mobile IPv6 and Predictive Address Reservation schemes have been proposed to reduce the handoff latency in their own ways. The Hierarchical Mobile IPv6 allows reducing handoff latency and overhead. On the other hand, the Predictive Address Reservation and anticipated handover use link layer information for earlier movement Detection and address configuration for the new point of Attachment so as to minimize the disruption of the services during the handoff process. The integration of

anticipated buffering reduces significantly the handoff packet losses during the handover process, but its integration with the HMIPv6 environment provides better handoff performance.

## REFERENCES

- [1] F. T. Harr, Realtime Routers for Wireless Networks, 2000, Ericsson <http://www.communiweb.net/ipv6/terhaar.pdf>, April 2005.
- [2] The Internet Engineering Task Force, <http://www.ietf.org>, March 2011.
- [3] R. Pazzi, Z. Zhang, and A. Boukerche, "Design and evaluation of a novel MAC layer handoff protocol for IEEE 802.11 wireless networks," *The Journal of Systems and Software*, Vol. 8, Issue 8, pp. 1364-1372, August 2010.
- [4] F. Belghoul, Y. Moret, and C. Bonnet, "Performance comparison and analysis on MIPv6, Fast MIPv6 Bi-casting and Eurecom IPv6 Soft Handover over IEEE802.11b," *WLANs Department of Mobile Communications, Institute Eurecom, IEEE Vehicular Technology Conference, IEEE 59thVTC*, Vol.5, pp. 2672- 2676, Milan 2004.
- [5] H. S. Yoo, R. Tolentino, B. Park, B. Y. Chang, and S. H. Kim, "ES-FHMIPv6: An Efficient Scheme for Fast Handover over HMIPv6 Networks," *International Journal of Future Generation Communication and Networking*, Vol. 2, Issue. 2, pp. 38-48, June 2009.
- [6] T. Park, "Adaptive handover control in IP-based Mobile Networks," Phd thesis, Institut for telecommunication Research, University of south Australia, 2003.
- [7] A. Vivaldi, M. Habaebi, B. Mohd Ali, and V. Prakash, "Fast Handover Algorithm for hierarchical mobile IPv6 Macro-mobility management," *Suranree J. Sci. Technol*, Vol. 42, Issue. 1, pp. 127-132, January 2004.
- [8] H. Yu, and M. Tao, "Fast Handover in Hierarchical Mobile IPv6 Based on Motion Pattern Detection of Mobile Node," *Wireless Pers Commun*, Vol. 19, Issue. 5, pp1338-1348, May 2010, DOI: 10.1007/s11277-010-0024-6.
- [9] M. Stemm, and R.H. Katz, "Vertical Handoffs in Wireless Overlay Networks," *ACM MONET*, Vol. 3, No. 4, pp.439-446, 1998.
- [10] C. Perkins, "IP Mobility Support for IPv4," RFC3344, IETF, August 2002, <http://www.ietf.org/rfc/rfc3344.txt?number=3344>, March 2011.
- [11] T. Narten, E. Nordmark, and W. Simpson. "Neighbor Discovery for IPv6," RFC2461, IETF, December 1998, <http://www.ietf.org/rfc/rfc2461.txt>, March 2011.
- [12] M. Alnas, I. Awan, and R.D.W. Holton, "Performance Evaluation of Fast Handover in Mobile IPv6 Based on Link-Layer Information," *The Journal of Systems and Software*, Vol. 83 Issue. 10, pp. 1774-1784, October 2010.
- [13] H. Soliman, C. Castelluccia, K. ElMalki, and L. Bellier, "Hierarchical Mobile IPv6 Mobility Management," RFC4140, August 2005, <https://www.ietf.org/rfc/rfc4140.txt>, March 2011.
- [14] M. Alrashdan, M. Ismail, and K. Jumari, "A Study on Effective Transfere Rate Over Smart Hierarchical Mobile IPv6 (SHMIPv6)," *IJCSNS International Journal of Computer Science and Network Security*, Vol.7, No.5, pp. 235-239, May 2007.
- [15] P. Macha'n, and J. zniak, "Simultaneous handover scheme for IEEE 802.11 WLANs with IEEE 802.21 triggers," *Telecommun Syst*, 2010, Vol. 43, No1-2, pp. 83-93, DOI :10.1007/s11235-009-9192-7.

# Heterogeneous Networks Handover Decision Triggering Algorithm Based on Measurements Messages Transfer using IP Option Header

Malak Z. Habeib<sup>1</sup>, Hussein A. Elsayed<sup>2</sup>, Salwa H. Elramly<sup>3</sup> and Magdy M. Ibrahim<sup>4</sup>

Electronics and Communication Engineering Dept.

Ain Shams University

Cairo, Egypt

<sup>1</sup>malakojc@yahoo.com, <sup>2</sup>helsayed2003@hotmail.com, <sup>3</sup>salwa\_elramly@eng.asu.edu.eg, <sup>4</sup>magdy\_ibrahim@yahoo.com

**Abstract**—The handover issue is one of the most important challenges in the next generation mobile networks. Traditional handover triggering conditions mainly based on signal strength, while the requirements for other triggering conditions such as user controlled handover or service based handover become an urgent need; especially when the handover is required among different networks. We propose a simple and easy mechanism for exchanging the handover metrics using proof of concept and logical visibility. This handover metrics represent the base in which the handover decision algorithms are constituted on. Metrics information can be transferred not only among different nodes that belong to the same radio access technology but also among different wireless access technologies as well. In this paper, we introduce new fields in the IP option header which are used for the handover metrics information exchange. We focus on some important metrics and how they can be read and written in the IP header. This allows any handover decision algorithm to openly use our scheme in a flexible way. The choice of the IP protocol comes from the trend of the next generation networks, which is based on IP networks.

**Keywords**—handover; vertical; horizontal; mobility; IP header; wireless; ping pong effect

## I. INTRODUCTION

Nowadays, the handover becomes a generic term because the Mobile Terminal (MT) not only needs to make the handover among different nodes inside the same wireless network, but also it needs to make the handover among different radio access networks; so it is called multi-interface or multi-mode mobile terminal [1]. If the MT requires the handover inside its current wireless access network it will be called Horizontal Handoff (HHO) or homogeneous handover, while it will be called Vertical Handoff (VHO) or heterogeneous handover if it requires making the handover among different technologies. Next generation wireless networks typically constitute different types of radio access technologies [2].

The handover terminology not only linked to the user mobility but also non-movable nodes can make the handover as well. For example, the user may request the handover from one wireless technology to another due to

the cost factor; which is independent on the mobility.

The generic handover process requires three phases: network discovery and measurements phase, taking the handover decision based on specific decision criteria phase and handover execution phase [3].

Jawad et al. [4] focused on the network selection, and how to choose network from a number of available networks in a heterogeneous system based on the Quality of Service (QoS). They achieved their target by proposing an architecture that combines QoS-Broker and network selection. For any network with sufficient QoS parameters, that matches the user request, the connection to that network will be triggered. While their mechanism is successful for guaranteeing the required QoS, it lacks of other metrics consideration, such as the vehicular speed moreover, the QoS-Broker is not an easy solution to implement because it requires additional cost.

In [5], George et al. introduce in their paper a network-based approach for access and interface selection in the context of resource management in heterogeneous wireless environments Universal Mobile for Telecommunications Service (UMTS), Wireless Local Area Network (WLAN) and Digital Video Broadcasting-Terrestrial (DVB-T). They interested in designing decision criteria by trying to optimize predefined cost function; however, they didn't consider the metrics and parameters criteria in their study.

Adiline and Anandha [6] proposed a user centric approach for controlling the handover between heterogeneous networks. In their approach, the mobility management is fully controlled by the terminal, and network selection is user-centric, power-saving, cost-aware, and performance-aware. However, they didn't address how the proposed handover metrics is practically achieved; moreover, they based the handover execution phase only on the Mobile IPv6. As well as, they protocol is not open for any wireless network.

A vertical handover using Media Independent Handover (MIH) layer is another protocol; this protocol is proposed by IEEE 802.21 working group [7]. MIH has many protocols which are developed for the heterogeneous networks handover. However, MIH is still limited to the handover preparation phase, as well as it

lacks of triggering and performance evaluation mechanisms. Dai et al. [8] address the triggering condition mechanisms but it is limited to only WLAN and Worldwide Interoperability for Microwave Access (WiMAX). Another work related to the vertical handover is presented in the literature [9]-[13]. Zhang et al. introduce an estimation of WLAN network conditions based on media access control network allocation vector occupancy [9], [10]. They used a Fast Fourier Transform (FFT) to detect the WLAN signal decay, but both of them didn't address how to estimate the WiMAX network conditions. In [11], Garg et al. introduce a handover criterion that combines the location using Global Positioning System (GPS) and IEEE 802.21 information elements; however, not all mobile terminals have GPS capabilities. Paper of [12] proposes handover rules based on a theoretically computed throughput, but without presenting any method to collect this information. In [13], Hassawa et al. proposed a generic vertical handover decision function, which provides handover decisions when roaming across heterogeneous wireless networks. However, it is very difficult to collect parameters such as cost, signal strength and vehicular speed then exchange them between MT and network nodes.

There are many handover protocols designed for VHO, but most of them focused only on the handover execution phase. For example, the following literature didn't show how network discovery and measurements collection phase or handover decision are addressed. For the IP supporting networks, the Mobile Internet Protocol (MIP) [14] is a typical mobility enabling protocol. It can be used for MIP version 4 (MIPv4) [15] and version 6 (MIPv6) as well [16]-[18]. Hierarchical Mobile IPv6 (HMIPv6) [19] facilitates local mobility management. The work in [20] introduced a Fast handover for MIP protocol that was used to reduce interruption time during handover. In [21], IPv6 protocol is used to enable a mobile node to configure a new Care of Address (CoA), when it changes its subnet. It was demonstrated in [22] that, Cellular IP (CIP) protocol can offer local mobility and handover support for moving nodes. CIP can co-operate with MIP to provide wide-area mobility support. In [23], a domain based approach for mobility support is proposed, which is called Handover Aware Wireless Access Internet Infrastructure (HAWAII). The handover issues that include horizontal and vertical handover using Hierarchical MIPv6 were discussed by Lee [24]. Maltz et al. [25] proposed Transport Control Protocol (TCP) connection that can divide the end to end connection into two connections: end to proxy and proxy to end. In [26], Multimedia Sockets (MSOCKS) uses Maltz technique for connection migration that can support multiple IP addresses. Seamless IP diversity based Generalized Mobility Architecture (SIGMA) [27] and Mobile Stream Control Transmission Protocol (mSCTP) [28] support soft handoff using IP diversity. Another technique is used in [29] to freeze the current TCP

connection, till making the handover by advertising a zero window size to the core network, then unfreezes the connection after finishing the handover. Nowadays, numerous studies are focused on the mobility supported Session Initiation Protocol (SIP) at the application layer level [30]. The details of how the SIP protocol can provide terminal and service mobility is discussed in [31].

Some of the above mentioned work only focused on the handover execution phase, while the rest concentrated on the triggering conditions of the handover decision phase. However, there is no specific study focused on how to exchange non-traditional triggering handover information such as user forced handover or speed based handover especially in VHO. The purpose of this paper is to introduce a new concept for exchanging the handover information using the IP option header.

In this paper, we consider the following hypothesis: any user terminal is equipped with more than one wireless interface. Moreover, we suppose that, more than one wireless access technology is always available for any user access. We will not expose to how the wireless network is discovered or how the handover protocol is executed.

The rest of the paper is organized as follows. Section II describes the reason for why we choose the IP protocol as a transport layer for handover information exchange. The details of events and measurements metrics; which can be transferred through the IP option header and used to take the handover decision, will be discussed in Section III. In Section IV, we will introduce a handover decision phase algorithm, which based on our proposed new handover metrics. The proposed algorithm is called as a ping pong and vehicular speed handover decision-based algorithm. Finally, Section V presents both paper conclusion and future work.

## II. THE IMPORTANCE OF USING THE IP PROTOCOL AS A BASE FOR OUR SCHEME

The IP is a dominate protocol that is used for the Next Generation Wireless Network (NGWN). The term All IP Networks comes from converging all network services that are based on the IP protocol; irrespective of whether the access is wired or wireless. The NGWN is one network that transports all media sources by encapsulating them into packets, such as we have on the Internet.

All IP based network has most important advantages over its predecessors. First of all, the IP protocol not only becomes integrated in all networks but it is also independent of the used radio access technologies as well. This means that, a core IP network can support different wireless access technologies such as cellular, WiMAX and WLAN. Now the core IP network can be evolved independently from the access network; this is the key advantages of using all IP. Also implementation and investment of the IP network is very easy with low costs. Moreover, using the IP protocol gives any user the

availability to have just one terminal that can support any type of service with low cost. All the above mentioned advantages of using the IP protocol emphasize the right choice of the IP protocol as a base for our scheme.

### III. IP ADDED HANDOVER OPTION HEADER FIELDS

In this section, we introduce new added fields in the IP option header for handover purposes.

#### A. Normal IPv4 Packet Header Overview

The IPv4 header consists of 20 byte as a mandatory header and 40 byte as a variable option header [32]. The IP option header is not normally used; it may be used according to the need. For example, the security option header may be required in all IP packets. The IP option header has a variable length according to its type. There are two types of the IP option header. Type-1 consists of one option byte; and type-2 consists of type, length and value fields.

#### B. The Handover Fields

The option type '11111' is not assigned for any purpose, so that we select this option field to exchange the handover related information that is carried directly in the header of the IP packet. We can transfer the handover information whenever it is needed, without interrupting any user session. Moreover, this information can be inserted directly whenever it is required. This means that, a very low delay is taken for the handover information exchange from one point to another among different network nodes.

#### C. Vertical or Horizontal HandoverFlag

This is '1' bit flag that is used to determine whether the associated handover information, which is carried in the IP header belongs to HHO or VHO. A vertical handover scenario is required in case the value of this flag equals to one, otherwise a horizontal handover information exchange is required. Using this flag means that, our mechanism can be useful not only for HHO scenarios but also for VHO.

#### D. Handover Counter

This portion of IP option header consists of four bits, which are used for exchanging the number of handover counts among network elements. Four bits means we have maximum up to '16' horizontal or vertical handover attempts. This field together with vertical or horizontal handover flag can be used to decide either it is needed to make VHO or HHO. By this way, we are able to count the number of vertical or horizontal handover events for a specific MT. This information is one of the key factors that are used to take a handover decision, especially in the VHO.

#### E. Forced Handover Flag

The handover decision not only controlled by the network itself, but also it may be based on the user choice. This is considered as one of the most important

requirements of the next generation mobile networks. In other words, we need to support user controlled handover as well as network controlled handover [33]. The purpose of this flag is to make it easy for the MT to request from the network or force it to trigger the handover process. In this case, the handover decision is taken by the user or MT and transferred to the network side in the IP header to proceed in handover execution. If the value of this flag equals zero, the user will leave the handover decision to the network hand.

#### F. Vehicular Speed Based Handover Flag

This flag is used to give a chance of triggering the handover decision based on the vehicular speed. Moreover, another field is required to show the vehicular speed level; so that the handover decision can be triggered vertically from one system to another accordingly. The levels of the vehicular speed can be handled by a speed based handover field, which will be discussed later in the next paragraph. The value of this flag controls the presence of optional speed based handover field. If we set this flag by one this means that, we have vehicular speed based handover field, otherwise there is no existence to this field at all.

#### G. Speed Based Handover Field

The length of speed based handover header consists of '3' bits that gives '8' permutations. According to the vehicular speed level, we can judge which appropriate target network suitable for the handover. Table I states the ranges of these '8' vehicular speed levels. We suppose that, three bits are enough to differentiate among '8' vehicular speed levels.

This field is optional; its presence depends on a vehicular speed based handover flag. If the value of this flag equals '1' it will report the presence of the vehicular speed based handover.

There are many systems that have different coverage cell size such as Public Land Mobile Network (PLMN), WiMAX, WLAN or Personal Area Network (PAN). So that more precise vehicular speed levels are required for accurate actions. This explains why we have many levels that represent these mobility classes.

We prefer transmitting the vehicular speed range indicator to exchange the absolute vehicular speed value. This is because it will save the header length as long as the range is enough to take accurate action. We need to confirm the difference between both vehicular speed-based handover and handover counter fields. The former not only depends on handover counter, but also it depends on other key factors. The handover decision can be taken not only based on the user mobility factor, but also it can be taken based on other factors such as the network cost, network load, user choice and many other metrics that can trigger the handover. Many of previous factors can trigger the handover while they are independent on the mobility. So that the handover counter can be used for other purposes rather than the mobility in the handover decision.

TABLE I. MOBILITY CLASSES FIELD

Value	Description of mobility classes in Kilo meter per hour
000	Vehicular speed 0 km/h
001	Vehicular speed > 0 km/h to 1Km/h
010	Vehicular speed > 1 km/h to 5 km/h
011	Vehicular speed > 5 km/h to 10 km/h
100	Vehicular speed from 10 to 60 km/h
101	Vehicular speed from 60 to 120 km/h
110	Vehicular speed from 120 to 250 km/h
111	Vehicular speed from 250 to 350 km/h

H. Service Priority Levels

Priority level field consists of ‘4’ bits this means that, we have ‘16’ permutations of the service levels. The MT can carry service information based on the user setting at the terminal side. We can use service priority levels field to transfer this information. By this way, the handover decision can be taken based on the service type. We suppose that, there is a table that contains a mapping between the service and its corresponding suitable network. Moreover, we can transfer the service priority information which can trigger the change in the current serving network. This enhances the availability of taking the handover decision based on the service type. Table II shows all service priority levels. The term Not Applicable (N/A) means that, the corresponding network doesn’t support the service right now; however, it may be available in the future.

I. Green Field indicator

This field is used to book a room in the IP header, which is particularly dedicated for battery level indication. We assign two bits to represent four battery levels. Table III shows the proposed battery levels from the strongest to the weakest level.

TABLE II. SERVICE PRIORITY LEVELS

Priority ID	Priority level
0	N/A (Not Applicable)
1	Level1 (Higher priority)
2	Level2
3	Level3
4	Level4
5	Level5
6	Level6
7	Level7
8	Level8
9	Level9
10	Level10
11	Level11
12	Level12
13	Level13
14	Level14
15	Level15 (Lower priority)

TABLE III. GREEN FIELD BATTERY INDICATOR

Value	Battery level
00	Strong bateery level
01	Level 2
10	Level 3
11	Ver low battery level

J. Uplink /Downlink Flag

This field is ‘1’ bit flag which is used to check the direction of exchanged information. From one side, the MT can exchange the handover information to the network; however, the network from the other side can transfer the handover information to the MT. We can use this information together with the source and destination IP addresses to check the exact path of handover information transfer.

This information is useful for different scenarios. For example, specific information can be transferred from one network node to another different network node. This can give the MT information about the availability of its surrounding wireless networks. This helps the MT during its handover discovery phase. Normally, the network discovery phase is done by the MT itself not by the network, however; we can seek the help of this network information in the discovery phase by using this proposed flag. Of course, this saves the mobile station battery life and guides the MT to make the smooth handover. The study of how the network discovery phase is done by the network side is out of this research scope.

IV. PING PONG AND VEHICULAR SPEED HANDOVER DECISION-BASED ALGORITHM

This algorithm is mainly based on both the vertical or horizontal flag and the handover counter field. We introduce this algorithm to avoid the ping pong phenomena; by frequently transferring from serving wireless network node to another. In other words, we can transfer from vertical to horizontal handover to avoid vertical ping pong effect; or transfer from horizontal to vertical handover to avoid horizontal ping pong phenomena. In this paper, we will focus in avoiding the horizontal ping pong effect. This algorithm uses the handover counter to count the handover attempts in a specific time. We will refer to this time by timer ‘T’; the handover decision may be triggered, when the wireless access node detects the value of handover counter exceeds certain threshold within the time ‘T’. The handover type can be detected by checking the vertical or horizontal handover flag. Fig. 1 depicts the signaling flow diagram according to the proposed handover decision algorithm.

We assume the threshold value of horizontal handover counter equals ‘10’. If the handover attempts reach this threshold within a certain time ‘T’, the vertical handover decision will be triggered. Fig. 2 shows the flowchart of the proposed algorithm.

The ping pong effect may be caused by the serving wireless technology coverage instability. In this case, it is recommended to search for another available wireless network to serve this user. Moreover, the proposed algorithm is also feasible in case we think about certain user with multi-interface terminal navigate with WLAN, while its vehicular speed increasing rapidly see Fig. 3. If available, our algorithm recommends the bigger coverage

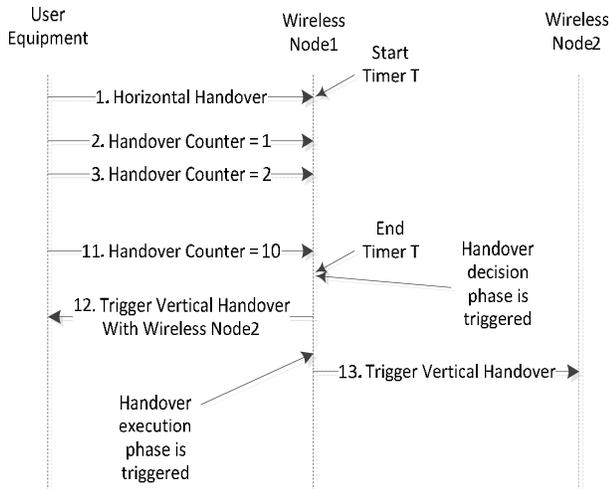


Figure 1. Signaling flow of horizontal ping pong and vehicular speed handover decision-based algorithm.

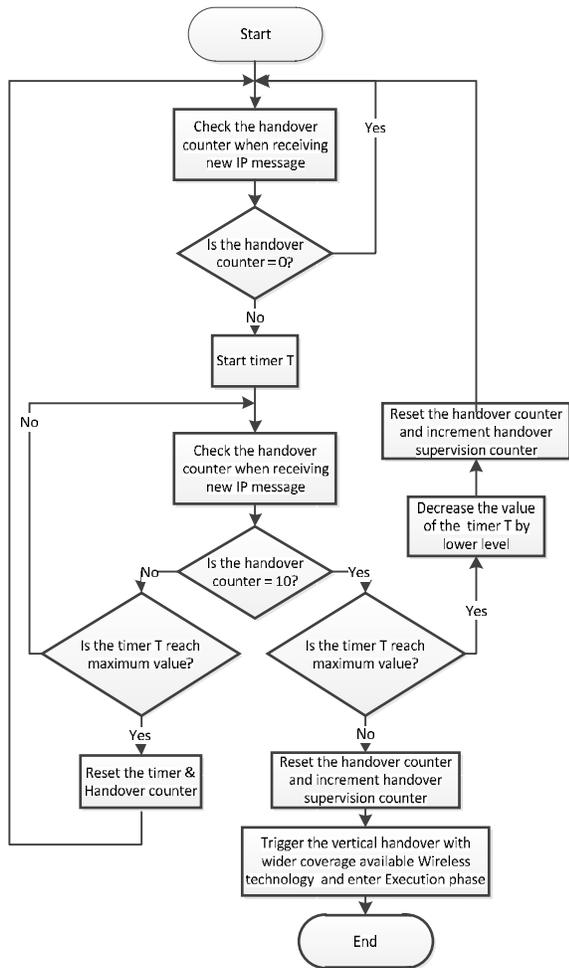


Figure 2. Flowchart of the horizontal ping pong and vehicular speed handover decision-based algorithm.

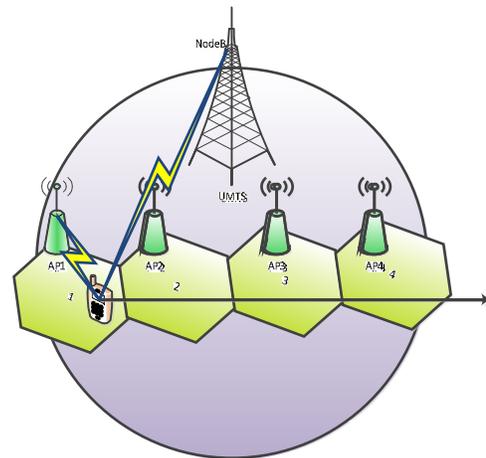


Figure 3. Triggering vertical handover condition based on horizontal ping pong and vehicular speed handover decision-based algorithm between WLAN and UMTS networks.

area supporting technology to deal with either high vehicular speed or ping pong handover scenarios.

If the handover counter value is less than or equals to '10', another counter called supervising counter will be incremented; while resetting the value of both handover counter and timer 'T' by zero. Note that, the supervising counter is a software module counter, and it doesn't exchange in the IP option header. Assume that, after checking the handover counter we found its value equals '10' in the same time, the vertical or horizontal handover flag equals '0'. This means that, ten horizontal handover attempts are performed.

Now, the network node can take a decision of a vertical handover. By the same way, the horizontal handover decision is taken in case there are a lot of vertical handover attempts are detected. By adding this field, the handover performance will be improved very much, and it will be easy to have an optimum handover decision.

We have many handover scenarios may be happened. Different handover software modules may use these metrics, which are introduced in our proposed fields; from different perspective. This guarantees the flexibility of using our scheme.

## V. CONCLUSION

To sum up, we introduce in this paper a mechanism that let us use the IP option header for handover information exchange. We introduce new IP option header fields, which are dedicated for handover metric information exchange.

We put extra '25' bit in the IP header, which is normally has at least '160' bit. These extra bits are used to introduce new eight handover metrics. Our scheme increases the IP signaling overhead by 13.51%; however, it gives more varieties for the handover decision algorithms to use our proposed metrics.

We also introduce handover decision algorithm that is based on our new metrics for both vertical or horizontal

flag and handover counter as well. This handover decision algorithm gives the flexibility to change the handover from horizontal to vertical and vice versa; to avoid the ping pong effect and address vehicular speed-based handover.

The choice of IP protocol for handover information transfer complies with the trend of the next generation wireless network, which is based on IP protocol. Our scheme gives the availability to take the handover decision not only based on traditional handover measurements such as signal strength but also based on new metrics. We introduce new added handover information such as vertical or horizontal handover flag, forced handover flag ;which is based on the user desire, handover counter, vehicular speed based handover flag, speed based handover field, service priority levels, green field that indicates to the user equipment battery level and finally the uplink /downlink flag. The use of forced handover flag guarantees the availability of using both network controlled and user controlled handover. This is smoothly done by easy and simple notification way.

Our proposed mechanism flexibly works for both horizontal and vertical handover. Any horizontal or vertical handover protocol can use our proposed methodology to enhance its performance. Different decision algorithms can also use our scheme to enhance the handover performance. Moreover, the implementation of our proof of concept scheme will be left as a future work; in which we can move from conceptual to realistic level.

#### REFERENCES

- [1] S. McCann, W. Groting, A.Pandolfi and E. Hepworth, "Next generation multimode terminals," Fifth IEEE International Conference on 3G Mobile Communication Technologies, IEEE, pp. 143–147, 2004.
- [2] F. Siddiqui and S. Zeadally, "Mobility Management across Hybrid Wireless Networks: Trends and Challenges," Computer Communications, vol. 29, no. 9, pp. 1363–1385, May 2006.
- [3] Z. Lei, "End to end architecture and mechanisms for mobile and wireless communications in the Internet," Ph.D. dissertation, Dept. Computer and Network, Toulouse Univ., Toulouse, 2009.
- [4] M. Jawad, W. Ismael, and M. Singh, "Optimizing network selection to support end-user QoS requirements for next generation networks," IJCSNS, vol. 8 no. 6, pp.113-117, June 2008.
- [5] S. Lee et al., "Vertical handoff decision algorithms for providing optimized performance in heterogeneous wireless networks," IEEE Trans. Vehic. Tech., vol. 58, no. 2, pp. 865–881, Feb. 2009.
- [6] Adiline Macriga, Anandha Kumar, "Mobility management for seamless information flow in heterogeneous networks using hybrid handover," IJCSNS, vol.10 no.2, pp.113-117, Feb. 2010.
- [7] V.Gupta et al., "IEEE P802.21 Tutorial," July 2006.
- [8] Z. Dai, R. Fracchia, J. Gosteau, P. Pellati, and G. Vivier, "Vertical handover criteria and algorithm in IEEE 802.11 and 802.16 hybrid networks," IEEE International Conference on Communications, pp. 2480–2484, May 2008.
- [9] Q.Zhang et al., "Efficient mobility management for vertical handoff between WWAN and WLAN," IEEE Communications Magazine, vol. 41 no. 11, Nov. 2003.
- [10] Q.Zhang et al., "A seamless and proactive end-to-end mobility solution for roaming across heterogeneous wireless networks," IEEE Journal on Selected Areas in Communications, vol. 22, no. 5, pp. 834-848, June 2004.
- [11] A.Garg and K.C.Yow, "Determining the best network to handover among various IEEE 802.11 and IEEE 802.16 networks by a mobile device," 2nd Int. Conf. on Mobile Technology, pp. 1–6, Nov. 2005.
- [12] Y.Choi and S.Choi, "Criteria for vertical handoff between IEEE 802.16e and 802.11 systems," Joint Conference on Communications and Information, pp. 1–10, Apr. 2006.
- [13] A. Hassawa, N. Nasser and H. Hassanein, "Generic vertical handoff decision function for heterogeneous wireless networks," IFIP Conference on Wireless and Optical Communications, pp. 239-243, Mar 2005.
- [14] C. Perkins, IP Mobility Support for IPv4, RFC 3220, Jan. 2002.
- [15] C. Perkins, IP Mobility Support for Ipv4, RFC 3344, August 2002.
- [16] D. Johnson, Mobile Support in IPv6, RFC 3775, June 2004
- [17] N. Montavont, T. Noe'l, "Handover management for mobile nodes in IPv6 networks," IEEE Commun. Magn. vol. 40, pp. 8, pp. 38–43, Aug. 2002
- [18] S. Vaughan, "Mobile IPv6 and the future of wireless internet access," Computer, vol. 36, no. 2, pp. 18–20, Feb. 2003.
- [19] A. Salkintzis, "Interworking between WLANs and third-generation cellular data networks," Proceedings of the 57th IEEE Semiannual Vehicular Technology Conference, 3, pp. 1802–1806, 2003.
- [20] P. McCann, Mobile IPv6 Fast Handovers for 802.11 Networks. IETF Mobile IP Group, 2004.
- [21] P. Vidales, L. Patanapongpibul, G. Mapp and A. Hopper, "Experiences with heterogeneous wireless networks, unveiling the challenges," 2<sup>nd</sup> International Working Conference on Performance Modeling and Evaluation of Heterogeneous \_Networks (HET-ETs), pp. 26 – 28, July 2004.
- [22] A. Campbell et al., "Design, implementation, and evaluation of cellular IP," IEEE Pers. Commun., pp. 42-49, Aug. 2000.
- [23] R. Ramjee et al., "HAWAII: A domain-based approach for supporting mobility in wide-area wireless networks," IEEE/ACM Trans. Net., vol. 10, no. 3, pp. 396–410, June 2002.
- [24] C. Lee, L. Chen, M. Chen and Y. Sun, "A framework of handoffs in wireless overlay networks based on mobile IPv6," IEEE J. Sel. Areas Commun. vol. 23, no. 11, pp. 2118–2128, Nov. 2005.
- [25] D. Maltz and P. Bhagwat, "TCP splicing for application layer proxy performance," IBM Research Report RC 21139, IBM T.J. Watson Research Center, March 1998.
- [26] D. Maltz and P. Bhagwat, "MSOCKS: An architecture for transport layer mobility," IEEE INFOCOM, San Francisco, California, USA, Vol. 3, pp. 1037–1045, March 29 - April 2, 1998.
- [27] S. Fu, L. Ma, M. Atiquzzaman, and Y. Lee, "Architecture and performance of SIGMA: A seamless handover scheme for data networks," IEEE ICC, Seoul, pp. 3249–3253, May 2005.
- [28] T. Go\_, J. Moronski, D. S. Phatak, and V. Gupta, "Freeze-TCP: a true end-to-end TCP enhancement mechanism for mobile environments," IEEE INFOCOM, Tel Aviv, Israel, pp. 1537–1545, March 26-30, 2000.
- [29] S. Koh, M. Chang, and M. Lee, "mSCTP for soft handover in transport layer," IEEE Communications Letters, vol. 8, no. 3, pp. 189–191, March 2004.
- [30] RFC3261. SIP: Session Initiation Protocol, June 2002.
- [31] H. Schulzrinne and E. Wedlund, "Application-layer mobility using SIP," SIGMOBILE Mob. Comput. Commun. Rev., vol. 4 no. 3, pp. 47-57, Jul. 2000.
- [32] Internet Protocol DARPA Internet program protocol specification, IETF Standard RFC791-1981. Available: <www.ietf.org/rfc/rfc791.txt> 26.4.2011
- [33] W. Ma, Y. Fang and P. Lin, "Mobility management strategy based on user mobility patterns in wireless networks," IEEE Transactions on Vehicular Technology, vol. 56, no. 1, pp. 322–330, Jan. 2007.

# The Impacts of Dynamic Broadcast Schemes on the Performance of Routing Protocols in MANETs

Dimitrios Liarokapis  
 School of Engineering and Computing  
 Glasgow Caledonian University  
 Glasgow, UK  
 e-mail: Dimitrios.Liarokapis@gcu.ac.uk

Ali Shahrabi  
 School of Engineering and Computing  
 Glasgow Caledonian University  
 Glasgow, UK  
 e-mail: A.Shahrabi@gcu.ac.uk

**Abstract**—In this paper, a new implementation of AODV, called AODV-ProbA, is presented that substitutes SF with the Probability-based Adaptive (ProbA) broadcast algorithm. AODV-ProbA is compared against normal AODV and another proposed enhancement for AODV, called DP-AODV. An algorithmic comparison between AODV-ProbA and DP-AODV shows minor but critically important differences in the use of local density knowledge and adjustment of the probability threshold. The simulation results also confirm our hypothesis that AODV-ProbA performs considerably better than AODV and DP-AODV in highly mobile dense networks under moderate to heavy traffic load.

**Keywords**—dynamic; adaptive; probability; broadcasting; MANETs

## I. INTRODUCTION

Over the past few years many studies have been conducted to develop broadcast mechanisms to alleviate the effects of SF [1], [2]. The focus of the early works was on the schemes where the mobile nodes make the rebroadcast decision based on fixed and preconfigured thresholds. Despite the fact that these schemes have been shown to considerably improve the overall performance of the network, they have been found to highly depend on the combination of threshold selected, density and load. The degree of dependency is such that in certain network topologies even SF performs better than these schemes [3].

Adaptive schemes have consequently been proposed to alleviate these dependencies. In such schemes the threshold used for the broadcast operation changes according to the local density of the network, within the transmission range of the sender (number of one hop neighbors) or within an expanded neighborhood area (number of two hops neighbors). To determine the density of the network locally, most of these schemes either exchange HELLO packets [4], [5] or use a positioning system such as GPS [6]. These schemes either introduce more overhead traffic to the network or demand the existence of expensive, and in many cases not very reliable, positioning systems. There are also adaptive schemes that decide upon the local density of the network based on duplicate receptions of a packet for the duration of a random or fixed period of time [7], [8].

A wide variety of broadcasting algorithms are being proposed under different assumptions. However, the credibility of simulation results and conclusions made when the network is only under broadcast traffic is very questionable as such scenarios are highly unrealistic. This approach ignores the dynamic interactions between

broadcasting algorithms and other components of networks. Hence, it is critical that when proposing a new algorithm, we evaluate it with accurate modelling of the underlying routing protocols and communication mechanisms. Clearly, after using such models a comprehensible understanding of the factors that affect the performance of a network emerges.

In this paper, we evaluate the performance of SF, ProbA [7] and DP [8] as broadcast mechanisms that take part in the route discovery process of the AODV routing protocol. In order to assess network performance, three performance metrics, namely packet delivery ratio, end to end delay and throughput, are used. This is attributed to the fact that the level of network performance visible to the end user is more important than that of internal network components.

The rest of this paper is organized as follows. Section 2 presents the related work for broadcasting in MANETs including a brief description of ProbA and DP. The algorithmic comparison of ProbA and DP is presented in Section 3. The results of extensive simulation study are presented in Section 4. Finally, Section 5 concludes this paper.

## II. RELATED WORK

Over the past years some probability-based (PB) algorithms have been proposed for broadcasting in MANETs. Probability-based algorithms are those which decide upon relaying a broadcast message using a probability value. ProbA and DP are two probability-based (PB) algorithms which have been proposed recently.

### A. Probability-based Adaptive (ProbA) Scheme

This approach introduces an extra step in the PB algorithm. According to PB, the receiving node applies the fixed probability threshold for the broadcast decision exactly after the packet is received. In ProbA, the mobile node falls into a listening mode for a random number of time slots upon reception of a new broadcast message.

ProbA takes advantage of this listening period and calculates the number of duplicate packets received, using a simple counter which is initialized to a value of 1 when a broadcast packet is received for the first time. The number of identical packets arriving at the mobile node is closely connected to the number of neighboring nodes. Each time the value of the counter increases, the probability threshold is tuned according to a pattern that is introduced administratively. This pattern is a scaled “if” statement, where the probability threshold changes its value depending on the current counter value. The number of possible values

**Algorithm: ProbA**
**Input:** broadcast message (*msg*)

**Output:** decides whether to rebroadcast *msg* or not

```

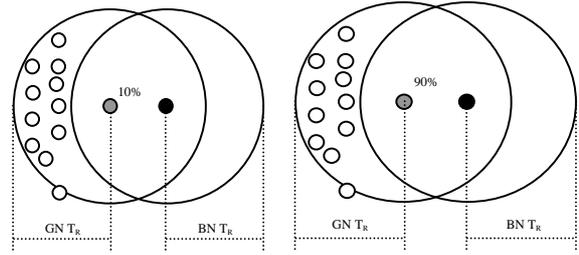
1: if ( msg is heard for the first time ) {
2:   count = 1;
3:   Generate a random number rand between 1 and 100;
4:   while (wait for a random number of slots){
5:     if (msg is heard again)
6:       count++;}
7:   if (count < c1)
8:     P = P1;
9:   else if (count < c2)
10:    P = P2;
11:   else if (...
...:   ... ..
n+1:   else if (count > cn)
n+2:    P = Pn;
n+3:   if (P > rand)
n+4:    exit;
n+5:   else { submit msg for transmission;
n+6:    exit;}}
    
```

**Fig. 1: ProbA algorithm.**

for the probability threshold is a parameter that is set by default but needs to follow an exponential and not a linear trend [9]. The value of the probability threshold could change multiple times during the listening period and every time a duplicate broadcast packet is received. The details of the ProbA algorithm are presented in Fig. 1 where  $P$  is the probability threshold,  $count$  is the counter described above,  $P_1, P_2, \dots, P_n$  are the pre-determined probability threshold values and  $c_1, c_2, \dots, c_n$  are the pre-determined counter values.

ProbA's primary goal is not to accurately calculate the number of neighboring nodes, but to decide upon the active density level of the network locally inside the transmission radius. This feature gives an extra advantage to ProbA in comparison to other adaptive schemes. An algorithm which is based on HELLO packets or a GPS system cannot properly estimate the number of active nodes retransmitting a broadcast message. For instance, grey node (GN) in Fig. 2 calculates the exact number of nodes inside the transmission radius. Either using GPS or HELLO packets, the end result of the calculation will be very close to 12, the total of all white nodes (WN) and black nodes (BN). As a result, GN will decide that the network is very dense locally and tune the probability threshold to be low (for example 10%), in order to avoid rebroadcasting that may cause collisions and contention. As the threshold is now very low, it is very likely that GN will not rebroadcast. Thus, none of the WNs will receive the broadcast packet.

In ProbA, GN will wait for a random period of time counting duplicate packets, Fig. 3. As GN has received a message once the counter will be set to 1 requiring a very high threshold value (for example 90%). It is highly possible at this point, as the threshold is very high, that GN will rebroadcast the packet and all WNs will receive it.


**Fig. 2: HELLO-GPS      Fig. 3: ProbA**

### B. Dynamic Probabilistic (DP) Broadcasting

DP also uses a packet counter in order to estimate the density of the network locally. A counter is maintained in every node for every broadcast packet received. The counter increases by 1 every time a duplicate packet is received. A high counter value implies high local density and, on the contrary, a low counter value represents a low level of local density. The probability threshold is increased in case that the counter is very low and decreased if it is high.

The decision of a node to increase or decrease the probability threshold and consequently to rebroadcast the packet or not has an effect on the neighboring nodes counter, as a rebroadcast will in turn increase their counters. According to the [8], “this kind of adaptation causes a dynamic equilibrium between rebroadcast probabilities and packet counter values among neighboring nodes”. This equilibrium state should lead to optimal results, although it is hard to reach that state as the mobile nodes may be constantly moving. For that reason, the probability threshold needs to be adjusted as quickly as possible. In addition, according to DP drastic changes in the probability threshold should be avoided.

DP dictates that a node should rebroadcast a packet depending on the current probability  $P$  if the packet is received for the first  $N_C$  times, where  $N_C$  is the threshold value to indicate whether enough duplicate packets were received or not. The probability  $P$  is decreased by a small constant  $d$  when an additional copy beyond  $N_C$  of an existing packet is received. The probability  $P$  is increased by another small constant  $d_1$  if a node has not heard anything within a time interval  $t$ . An upper  $P_u$  and a lower  $P_l$  bounds are also set. The algorithm is presented in Fig. 4.

Setting the value of  $t$  and the initial value of  $P$  is critical. If  $t$  is set too low, the counter may be checked too often and the packet counter may remain low. In this case, the probability value could remain the same. If  $t$  is set too high the counter may be checked too less and the packet counter may exceed the threshold often and the probability could be set too low.

The value of  $N_{neighbour}$  is calculated using the formula:

$$N_{neighbour} = (N-1) \frac{\pi R^2}{A}. \quad (1)$$

The initial value of  $P$  is set as follows.

```

Algorithm: DP
Input: broadcast message (msg)
Output: decides whether to rebroadcast msg or not

1: if (msg is received){
2:   if (msg is in the received message list)
3:     if (count > NC){
4:       P = P - d;
5:       if (P < Pl)
6:         P = Pl;
7:     }
8:     count = count + 1;
9:   else {
10:    count = 1;
11:    Add msg ID to the received packet list with an
    expiration time;
12:    Submit msg for transmission with probability P;
13:    #####
14:    for (every time interval t)
15:      if (no msg is received within t)
16:        if (count < NC){
17:          P = P + d1;
18:          if (P > Pu)
19:            P = Pu;
20:          Remove msg ID from received message list;}

```

Fig. 4: DP algorithm.

$$P = \begin{cases} 1, & \text{where } \frac{6}{N_{\text{neighbour}}} \geq 1 \\ \frac{6}{N_{\text{neighbour}}}, & \text{where } \frac{6}{N_{\text{neighbour}}} < 1 \\ 0, & \text{where } \frac{6}{N_{\text{neighbour}}} \leq 0 \end{cases} \quad (2)$$

III. PROBA VS DP

The core of both ProbA and DP algorithms is to adaptively make decision to rebroadcast a packet or not depending on the local density of the network. A node starts a timer to enter a listening (or learning) mode upon reception of a new broadcast packet. During listening period, duplicate packets are counted. A high value reflects high number of active neighbouring nodes requiring a lower value for probability threshold. On the contrary, if the counter value is low, the probability threshold value needs to have a high value as a rebroadcast will provide a significantly large extra coverage area and thus the broadcast operation will most likely not die out. It should be remembered that the local density of the network is not calculated accurately. The counter value is just an indication of the number of active neighbouring nodes. This is not an algorithmic weakness as the counter value is proportional to the node density in the

surrounding area for a given rebroadcast probability distribution among neighbouring nodes. Despite the fact that both algorithms are based on the same logic of implementing adaptivity, they differ in two critical points; how the listening period is accommodated inside the broadcast algorithm and what is followed when adjusting the probability threshold.

A. Listening Period

According to the DP algorithm, a node does not wait for the timer to expire before making the rebroadcast decision. It immediately decides whether to rebroadcast or not. The timer is used to accumulate knowledge for future decisions. The counter increases when a duplicate packet is received and in turn the probability threshold is decreased. If nothing is heard during the listening period the counter is constantly set to 1 and the probability threshold is increased. Whereas, the ProbA algorithm dictates that the node can only decide upon the rebroadcast of the packet after the timer has expired, thus it makes its decision based on freshly obtained knowledge upon the local density of the network. DP would perform poorly when the network topology changes fast, as the node would base its decision on some stale knowledge. This laziness may lead to poor performance in highly mobile networks.

DP may look faster than other timer-based algorithms, as pointed out in [8]. However, this could only be true if the timer value is greater than 5-10% of the total end to end delay of the entire process. For example, a waiting time of 30ms does not have a significant effect on a process that could last 300ms or even longer in case of high traffic load in the network where delay can exceed the value of 1sec. Furthermore, a small waiting time could also aid to avoid further collisions and consequently a rather lengthy back-off process.

B. Probability Threshold Adjustment

The increase or decrease of the probability threshold is closely related to the potential additional coverage area that could be achieved when the broadcast packet is transmitted. If a large extra area is predicted to be covered by rebroadcasting of a packet, the probability threshold should be set to a high value. That is the case when the counter value is low. On the contrary, if the predicted coverage area is small, the probability threshold should be adjusted to a low value. This is also the case when the counter value is high. It is obvious that counter value, probability threshold and extra coverage area greatly affect one another in that order.

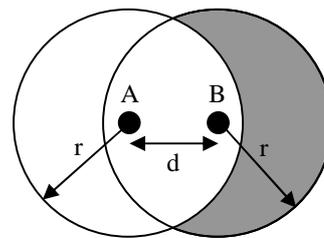


Fig. 5: Extra area analysis

The DP algorithm uses a linear pattern for the adjustment of the probability threshold. For every increase of the counter value, the probability threshold is decreased by a small constant [8]. Furthermore, for every time interval that there are no duplicate broadcast packets received, the probability increases by another small constant. The ProbA algorithm makes use of a scaled *if* statement for the adjustment of the probability [7]. This should lead to an exponential decrease of the probability depending on the counter value. An example of the scaled *if* statement is as follows:

---

```

if(count = 1)    P = 90%;
else if(count < 4) P = 50%;
else            P = 10%;
    
```

---

In order to conclude which of the two patterns is more suitable, we need to take into consideration the redundant broadcast analysis performed in [9]. Consider the scenario in Fig. 5. Node A sends a broadcast packet and node B decides to rebroadcast it. Let  $S_A$  and  $S_B$  denote the circle areas covered by the transmission ranges of nodes A and B respectively. The gray area represents the additional area that will be covered by B's rebroadcast named  $S_{B-A}$ . We can derive that:

$$|S_{B-A}| = \pi r^2 - INTC(d), \quad (3)$$

where  $INTC(d)$  is the intersection area of the two circles centered at two points distanced by  $d$ .

$$INTC(d) = 4 \int_{d/2}^r \sqrt{r^2 - x^2} dx. \quad (4)$$

The extra coverage area gets the maximum value when  $r = d$  and is equal to:

$$\pi r^2 - INTC(r) = r^2 \left( \frac{\pi}{3} + \frac{\sqrt{3}}{2} \right) \approx 0.61\pi r^2. \quad (5)$$

Thus, B's rebroadcast can cover an extra area of 61% of the area covered by the previous transmission. The average extra coverage area can be obtained by integrating the above value over the circle of radius  $x$  centered at A for  $x$  in  $[0, r]$ :

$$\int_0^r \frac{2\pi x \cdot [\pi r^2 - INTC(x)]}{\pi r^2} dx \approx 0.41\pi r^2. \quad (6)$$

A rebroadcast can cover an additional of 41% area in

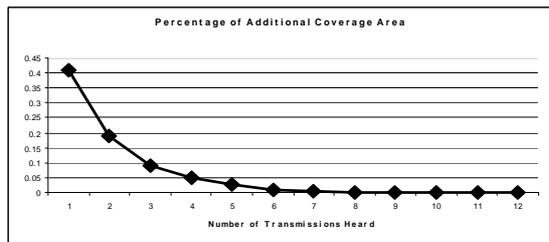


Fig. 6: Analysis of Redundancy

average. Following the same pattern, the extra area covered can be calculated depending on the number of transmissions heard for the broadcast packet. The result is shown in the graph of Fig. 6.

This analysis confirms ProbA's hypothesis that the probability threshold adjustment should follow an exponential and not a linear decrease pattern proving the superiority of ProbA over DP.

#### IV. PERFORMANCE ANALYSIS

In this section, we present the simulation results of our performance comparison study that will confirm the findings of the algorithmic comparison performed in the previous section. The algorithms of normal AODV with AODV-ProbA and DP-AODV are compared. The performance metrics for comparison include the packet delivery ratio (PDR), end to end delay, and throughput.

##### A. Simulation Setup

The simulator used for the experiments is NS-2. All experiments are grouped into 3 different categories depending on node density, traffic load (number of connections) and mobility. For the first group of scenarios node density increases from 20 nodes up to 200 nodes with a constant step of 20. The second group includes the results against traffic load starting from 20 TCP connections and reaches the maximum of 60 connections, again with a constant increase step of 10 connections. The last group evaluates the performance of the three algorithms against mobility. The starting point for the node speed is 10m/s and increases by 5m/s until it reaches the maximum value of 30m/s.

All nodes are placed randomly within a network topology of 1000x1000 square meters. Transmission range for all nodes is set to 250m and channel capacity is 2Mbps. Each simulation run is executed for 800sec of simulation time. Nodes move inside the network with a maximum speed of 20m/sec for the first two groups and 0 pause time for all. Node movement is generated using the setdest command provided by NS-2, following the random waypoint model. Every scenario is run 3 times with different random movement of nodes, in order to avoid any extremes that could compromise the reliability of our results. Final results are calculated as the average of the 3 repetitions. Default AODV parameters are used for all protocols, as our implementations only dealt with the broadcast mechanism used. The type of traffic used in our experiments is Constant Bit Rate (CBR). The number of connections per scenario was kept the same with a value of 50 connections for the first and third group of experiments. Packet generation rate is set to 1.0 packet per second. Data payload is 512 bytes.

AODV-ProbA uses 3 different probability thresholds depending again on the density of the network locally. Values for the probability thresholds used are, 10% for counter value of 1, 50% for counter values of 2 and 3 and 90% for counter values of 4 and higher. Probabilities are set following an exponential pattern as described in III.B. The algorithm of DP-AODV requires the set of 5 additional parameters. They are summarized in the table below:

TABLE I. DP-AODV PARAMETERS

Probability Decrease Constant $d$	1%
Probability Increase Constant $d_1$	2%
Time interval $t$	35msec
Upper Probability Bound	90%
Lower Probability Bound	10%

The initial probability threshold  $P$  is calculated using the formula presented in Section II.B. The justification for the values of  $d$  and  $d_1$  is that the authors of DP argue in favor of non-dramatic changes in the probability. Low values of 1% and 2% are chosen respectively. The value of  $d_1$  is double the value of  $d$  in order to reach an equilibrium state. The probability  $P$  decreases more times in average as in our dense scenarios it is more likely for a node to receive the broadcast packet more than once during the simulation time. For reasons of fairness the time interval  $t$  for DP is set to 35msec. That is the average listening time for AODV-ProbA as well. Upper and lower probability bounds are set to the same values for both algorithms.

The following performance parameters are considered for our simulation experiments. It is noted that all metrics are concerned with both TCP and AODV traffic.

**Packet Delivery Ratio (PDR)** – The percentage of successful packet deliveries throughout the simulation time.

**End to End Delay** – The amount of time elapsed from the

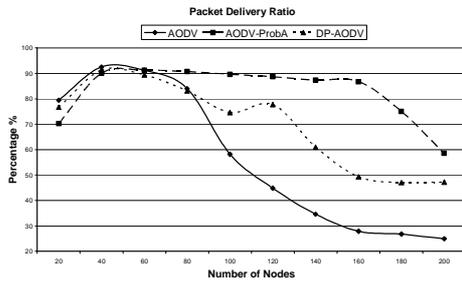


Fig. 7: PDR vs Density

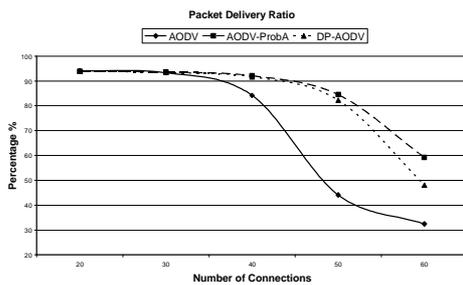


Fig. 8: PDR vs Tr. Load

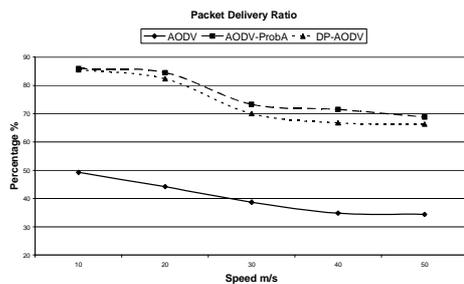


Fig. 9: PDR vs Mobility

time a packet was originated from the source node until the time it is successfully delivered to the destination node.

**Throughput** – The average rate of successful data delivery in the network measured in kbps.

**B. Packet Delivery Ratio**

A mobile node will miss a packet if all of its neighbors decide to suppress rebroadcast in case of an AODV packet or if there is a collision and the TCP packet never reaches its destination.

Fig. 7 shows the packet delivery ratio for a network against node density. All three algorithms perform in a similar way for sparse topologies of up to 60 nodes. For medium and high node density topologies, AODV-ProbA performs better than both AODV and DP-AODV. Fig. 8 shows the PDR percentage for a network against traffic load. Once again, for low traffic density of 20 and 30 connections, all algorithms produce the same results. In case of more connections, AODV-ProbA and DP-AODV perform better than normal AODV, with the latter being slightly outperformed. The PDR level of a network when node speed is increased is presented in Fig. 9. AODV-ProbA produces higher PDR for all average node speeds. Normal AODV performs poorly reaching down to 35% of PDR.

**C. End to End Delay**

In general, the metric of end to end delay is found to produce very similar trends with the results for PDR or reachability. Especially in scenarios with high levels of node and traffic load densities, back-off and medium detection mechanisms may delay the transmission of packets in addition to the high probability of the packet never reaching its destination.

As shown in Fig. 10, when AODV-ProbA is used, a packet travels from source to destination with high speed. The performance of DP-AODV is average when delay is measured and normal AODV produces unacceptable delay for medium and high node density networks. Fig. 11 shows the end to end delay for a network against traffic load. Both DP-AODV and AODV-ProbA perform better than normal AODV with the latter producing slightly lower average delay for all scenarios. End to end delay against mobility is presented in Fig. 12. Normal AODV with simple flooding causes delay to be on average 150% higher than the other two algorithms. Once again, AODV-ProbA slightly outperforms the Dynamic Probabilistic algorithm.

**D. Throughput**

Throughput is an important metric that represents a network's ability to transmit data. It is a very popular metric in QoS performance comparison studies in MANETs and is defined as the number of bits or kbits transmitted per time unit.

In Fig. 13 we compare the network throughput for different node densities. Normal AODV is outperformed for medium and high density levels. Despite the fact that performance for DP-AODV is slightly higher for sparse and average node density networks when compared against AODV-ProbA, when the network becomes extremely dense

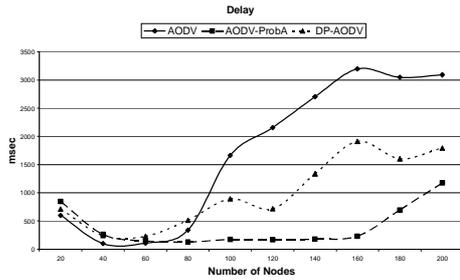


Fig. 10: Delay vs Density

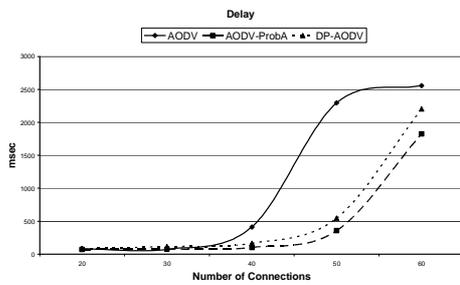


Fig. 11: Delay vs Tr. Load

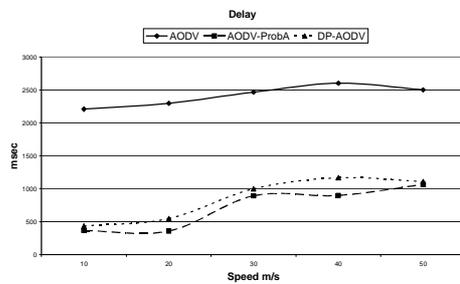


Fig. 12: Delay vs Mobility

its performance rapidly decreases. The performance of a network in terms of throughput when different traffic loads are configured is shown in Fig. 14. All algorithms perform almost identical for 20, 30 and 40 connections. When the number of connections increases, throughput for normal AODV begins to decrease sharply. AODV-ProbA performs better than DP-AODV for very high traffic loads once again. Fig. 15 confirms the superiority of AODV-ProbA for scenarios with different node speeds, as it constantly produces higher throughput than AODV and DP-AODV.

V. CONCLUSION

In this paper, a comparison between the effects of two probabilistic adaptive schemes, ProbA and DP in AODV routing protocol has been presented. Two key differences between these two algorithms have been highlighted; the role of the listening period as part of the adaptivity mechanism and the way this is implemented in addition to the different mathematical pattern followed when adjusting the probability threshold. The result of this comparison has led us to the fact that ProbA should outperform DP in dense networks with highly mobile nodes. The experimental results presented in this work have confirmed the superiority of ProbA against DP in almost all simulation scenarios used in terms of packet delivery ratio, end to end delay, and throughput under various network sizes and traffic loads as well as different node speeds.

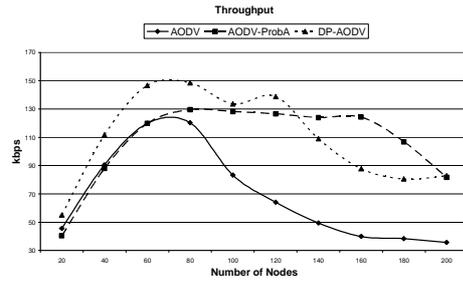


Fig. 13: Throughput vs Density

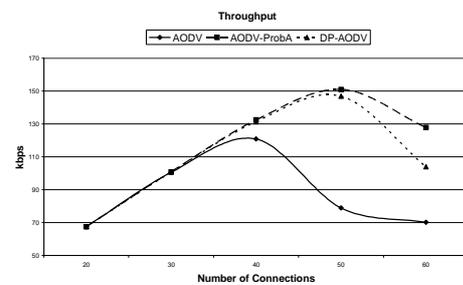


Fig. 14: Throughput vs Tr. Load

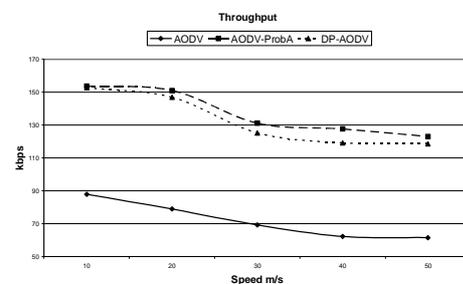


Fig. 15: Throughput vs Mobility

REFERENCES

- [1] S. Leng, L. Zhang, L. W. Yu, and C. H. Tan, "An efficient broadcast relay scheme for MANETs," *Com. Com.* 28, 5, pp. 567-476, 2005.
- [2] R. Purtoosi, H. Taheri, A. Mohammadi and F. Foroozan, "Improving broadcast performance by traffic isolation in wireless ad hoc networks," *Res. Art. Int. J. Com. Syst.* 19, 9, pp. 1029-1043, 2006.
- [3] B. Williams, and T. Camp, "Comparison of broadcast techniques for mobile ad hoc networks," In *Proceedings of Mobile Ad Hoc Networking and Computing*, MOBIHOC, pp. 194-205, 2002.
- [4] M. D. Colagrosso, "Intelligent Broadcasting in Mobile Ad Hoc Networks: Three Classes of Adaptive Protocols," In *Proceedings of EURASIP*, pp. 25-25, 2007.
- [5] Y. Tseng, S. Ni, and E. Shih, "Adaptive Approaches to Relieving Broadcast Storms in a Wireless Multihop Mobile Ad Hoc Network," In *IEEE Transactions on Computers*, pp. 545-557, 2003.
- [6] D. Young-Ching, S. Ching-Chi, and L. Ferng-Ching, "An Adaptive Medium Access Control Protocol for Reliable Broadcast and Unicast in Ad Hoc Networks," In *Proceedings of IEICE - Tran. on Inf. & Sys.*, pp. 527-535, 2006.
- [7] D. Liarokapis and A. Shahrabi, "A probability-based adaptive scheme for broadcasting in MANETs," In *Proceedings of the 6th international Conference on Mobile Technology, Application & Systems*, ACM Mobility, Nice, France, 2009.
- [8] Q. Zhang and D. P. Agrawal, "Dynamic probabilistic broadcasting in MANETs," In *J. Par. Distrib. Comput.* 65, 2, pp. 220-233, Feb. 2005.
- [9] S. Ni, Y. Tseng, Y. Chen and J. Sheu, "The broadcast storm problem in a mobile ad hoc network," In *Proceedings of the 5th Annual ACM/IEEE int. Conf. on Mobile Comp. and Networking*, MobiCom '99, pp. 151-162, Seattle, Washington, United States, 1999.

# Cyclostationary Detection in Spectrum Pooling System of Undefined Secondary Users

Nazar Radhi<sup>1</sup>, Kahtan Aziz<sup>2</sup>, Rafed Sabbar Abbas<sup>3</sup>, Hamed AL-Raweshidy<sup>4</sup>

<sup>1,3,4</sup> *Wireless Network & Communication Centre, School of Engineering & Design  
Brunel University, London, UK*

<sup>2</sup> *College of Engineering and Applied Sciences, Al-Ghurair University, Dubai, UAE*

[Nazar.Radhi@brunel.ac.uk](mailto:Nazar.Radhi@brunel.ac.uk)<sup>1</sup>, [Kahtan@agu.ac.ae](mailto:Kahtan@agu.ac.ae)<sup>2</sup>

[RafedSabbar.Abbas@brunel.ac.uk](mailto:RafedSabbar.Abbas@brunel.ac.uk)<sup>3</sup>, [Hamed.Al-raweshidy@brunel.ac.uk](mailto:Hamed.Al-raweshidy@brunel.ac.uk)<sup>4</sup>

**Abstract**— Spectrum sensing continuing to emerge as an essential topic for the cognitive networks where two kinds of users primary and secondary will share the band. This paper proposes a method for real-time detection of secondary users at the base stations. Cognitive Radios may hide themselves in between the primary users and Rental secondary users based on spectrum pooling system to avoid being charged for using spectrum. To deal with such scenario, a cyclostationary Fast Fourier Transform FFT Accumulation Method (FAM) has been used to develop a new scheme for channels users recognition. Channels users are tracked according to the changes in their signal parameters for instance modulation techniques. The Matlab simulation runs three signals transmitted on spectrum pooling system channels. Obtained spectral correlation density functions show successful Differentiation between signals.

**Keywords**— *Cognitive Radio; Sensing; Cyclostationary Detection; Undefined.*

## I. INTRODUCTION

Nowadays, wireless systems are based on fixed spectrum allocations, allocated fixed spectral bandwidth to licensed user at any time. Which lead to a useless of scarce and expensivespectral resources and result in inefficiency utilizing spectrum. Dynamic spectrum access techniques secure greater spectral-usage efficiency and enhanced access to frequency spectrum based on spectrum pooling [1]. of Spectrum pooling is a resource sharing strategy which allows the licensed owner to share portion of his licensed spectrum with a rental secondary users [2], cognitive radio users absent until he needs it himself. The goal of the spectrum pooling is to improve spectral efficiency by overlaying new wireless radio systems on a licensed one (the primary users) without interfering to the primary users, and without changing its operations. In order to keep existing and without harmful interference with rental users, cognitive radios technology must have the ability to detect unused spectrum, it is a very important process in spectrum pooling system.

One of the most significant issues in the cognitive radio technology is spectrum sensing, because in cognitive radio system, the systems distinguish the radio environment by the spectrum sensing. If the spectrum

sensing does not work accurately, the cognitive radio system will have incorrect information about the radio environment, and the system will try to use the spectrum which a primary user uses and does not use the spectrum which the primary user does not use. It results in the several performance degradation of: the cognitive radio system and the primary user [3].

Traditionally, there are three schemes which are used for spectrum sensing, such as: Matched filter detection, Energy detection and Feature detection [4].

Matched filter detection: When the parameter of the primary user signal is known to the CR user, the optimal detector in stationary Gaussian noise is the matched filter, it maximizes received signal-to-noise ratio. While the matched filter requires a priori characteristics knowledge of the primary user signal, e.g. modulation type and order, pulse shaping, packet format.

Energy detection: If the receiver cannot gather sufficient in formation about the primary user signal, the optimal detector is an energy detector. However, the energy detector measures energy in each narrowband channel and determines the presence of a primary user if the energy detected in a narrowband channel is higher than a certain threshold. However, to achieve high receiver sensitivity, a low threshold has to be used. In some cases, the threshold has to be lower than the noise floor, in which case the detection fails. The problem is even more complex due to the fact that the noise is most likely non-Gaussian because of the presence of CR user's interference.

Feature detection: Most of the signals encountered in wireless communications are cyclostationary inherent, whereas the noise is stationary [5]. The wireless communication signals loaded with sinusoidal carriers, pulse trains, repeating codes, hopping sequences, cyclic prefixes, and signals are cyclostationary because their mean value and autocorrelation function exhibit periodicity. This periodicity trend is used to perform various signal processing tasks that includes detection, recognition and estimation of the received signals [6]. As a result, the cyclostationarity of the primary signals can be used to detect their presence. The cyclostationarity of a

signal is not reflected in the power spectral density (PSD); however, it is reflected in the spectral correlation density (SCD) function which is obtained by taking the Fourier transform of the cyclic autocorrelation function. Therefore, spectral correlation analysis of the received data can be used to identify the signal. Higher order spectral statistics have also been used to identify weak users [5].

The paper has been organized as follow: the proposed model to be used in this investigation is discussed in Section II. Section III introduces detection method. In Section IV, reports on simulated implementations and analyses of the proposed model. Section V presents concluding remarks.

## II. PROPOSED MODEL

The cognitive radios (CRs) are designed to work in a crowded wireless environment. Thus, scarce spectrum may lead selfish cognitive networks to use illegally the spectrum. High reconfigurability specifications of the cognitive radios make them capable to adapt their signal parameters according to their needs and the channels they are working in. Although, CRs designing have passed a long way of development to allow these transceivers to be obtainable in the near future, it is scarcely visible how these services will be monitored. To verify the cognitive network, a supporting sensing network is designed to examination the spectrum. Observed holes suitable to transmit are reported whenever a request to transmit and the occasion are available. We think that: duties for this sensing network should be extended to include the CRs recognition. A wider network may be created by adding awareness abilities to the sensing network to produce a novel robust monitoring system.

The new designed scheme will be capable of noticing the white holes in the spectrum, and also identify each channel user. Such a development requires the amalgamation of the optional monitoring system and information resources for occurrence the Spectrum Broker. The observed data are then being sent instantaneously to the decision makers in the main wireless providers for additional processing.

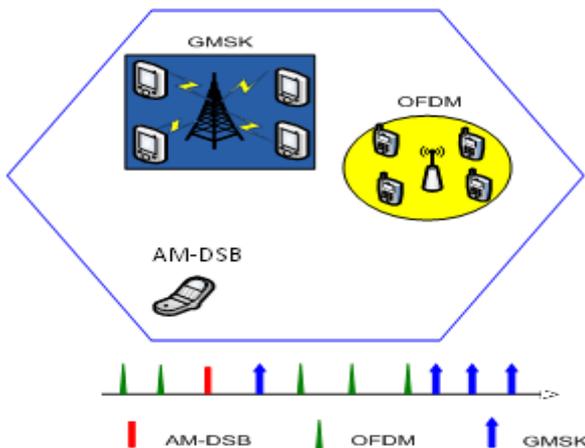


Fig .1 spectrum pooling system

The opportunity that a certain CR may transmit illegally falls outside the current definitions of cognitive networks. These bluffer cognitive radios can use their technical advances to adapt their carrier frequencies to transmit on a certain channel when the primary user is off. However, they still need to transmit using different signal parameters to maintain broadcast dedicated to their end users. This exploit may occur at any time and can be happen rarely or even constantly. The recommended observation scheme will use the FFT Accumulation Method FAM to detect deceptive CRs behaviour instantaneously. FAM is presented as the algorithm for cyclic spectrum analysing. This method is derivative from the cyclostationary technique which is widely acceptable as the most effective sensing procedure for the cognitive radios. This work is a natural expansion of our work in [7] by incorporating the recently suggested spectrum pooling system and different CR users.

## III. DETECTION METHOD

The cyclostationary processing theory is proposed here as the algorithm for the developed identification scheme. Most of the communication signals can be modelled as cyclostationary random process. Let say, a zeros-mean discrete time signal  $x(n)$  is cyclostationary with periodic  $T$  if its autocorrelation function  $R_x(n, k)$  is also periodic in  $T$ , as shown in Equation 1 [7].

$$R_x(n, k) = R_x(n + T, k + T) \tag{1}$$

To gain an intuition into how cyclostationary based detection works, it is beneficial to define the Cyclic Autocorrelation Function (CAF) [7].

$$R_x^\alpha(n, k) = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{n=-N}^N x(n) x^*(n - k) e^{-i2\pi\alpha n} e^{i\pi\alpha k} \tag{2}$$

where  $R_x^\alpha(k)$  is the CAF of discrete time signal  $x(n)$  and  $\alpha$  is the cyclic frequency. The CAF can also be interpreted as the measured amount of correlation between different frequency-shifted versions of a given signal. For different signal, the CAF exhibits different features, which are generally used for detecting the presence of the signal. However, this feature is not easily seen by only observing its CAF, in time domain. For this reason, it is important to determine in the frequency domain the amount of correlation between frequency-shift versions of the signal. By applying the Fourier transform to the CAF,

Giving the cyclic spectral correlation function as shown in equation 3 [7].

$$S_x^\alpha = \sum_{k=-\infty}^{\infty} R_x^\alpha(k) e^{i2\pi f k} \tag{3}$$

where  $S_x^\alpha(f)$  is called cyclic spectral correlation function or cyclic spectrum of discrete time signal  $x(n)$ . Notably, if  $\alpha=0$ , the CAF and the Cyclic Spectrum (CS) reduce to the conventional autocorrelation function and power spectral density function, respectively. FAM method is based on modifications of time smoothed cyclic cross period gram which is defined as [7]:

$$S_{xyT}^\alpha(n, f) = \lim_{N \rightarrow \infty} \frac{1}{2N+1} \sum_{n=-N}^N \frac{1}{T} X_T(n, f) + \frac{\alpha}{2} Y_T^*(n, f - \frac{\alpha}{2}) \quad (4)$$

where  $X_T(n, f + \frac{\alpha}{2})$  and  $Y_T(n, f - \frac{\alpha}{2})$  are the complex the complex envelopes of narrow band, band pass components of the signals  $X(n)$  and  $Y(n)$ , respectively. These complex envelopes are computed in the following way [7]:

$$X_T(n, f) = \sum_{k=-N'/2}^{+N'/2} \alpha(k) x(-k) e^{-i2\pi f(n-k)Ts} \quad (5)$$

$$Y_T(n, f) = \sum_{k=-N'/2}^{+N'/2} \alpha(k) y(n-k) e^{-i2\pi f(n-k)Ts} \quad (6)$$

where  $\alpha(k)$  is a data narrowed window of length  $T = N'Ts$  seconds and the sampling period  $Ts$ . The FAM method is validated by estimating the cyclic spectrum. Fig. 2 illustrates the implementation model of FFT Accumulation Method.

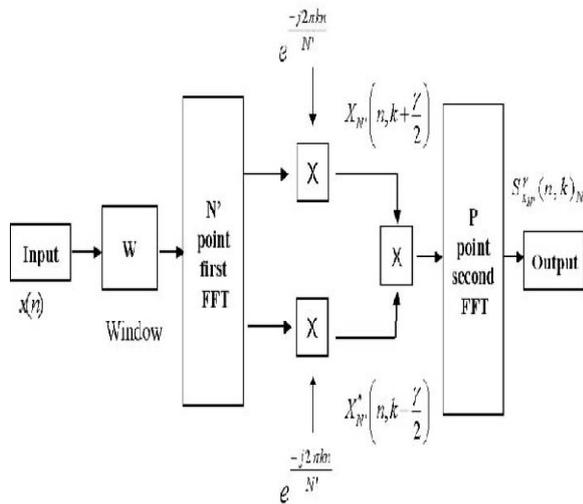


Fig. 2 FFT Accumulation Method

The FAM method works as follow:

- The complex envelopes are estimated efficiently by means of a sliding  $N'$  point FFT, followed by a downshift in frequency to baseband.
- In order to allow for an even more efficiency estimation, the  $N'$  point FFT is hoped over the data in blocks of  $L$  samples which means that  $L$  data point are skipped between computations of the  $N'$  point FFT.

- After the complex envelopes are computed and the product sequences between each one of them and the complex conjugate of the other  $s$  are formed, the time smoothing is accomplished by means of a  $P$  point FFT.

The value of  $L$  was chosen to be equal to  $N'/4$ . The value of  $N'$  is determined according to the desired resolution in frequency ( $\Delta f$ ) used in the algorithm, and is given by [7]:

$$N' = \frac{f_s}{\Delta f} \quad (7)$$

The value of  $P$  is determined according to the desired resolution in cyclic frequency ( $\Delta\alpha$ ), and is given by [7]:

$$P = \frac{f_s}{L\Delta\alpha} \quad (8)$$

#### IV. SIMULATION

A Matlab simulation code was created to generate FAM as the sensing mode for our model. We demonstrate a spectrum pooling example which has a GSM network as primary users and cognitive radio network as rental secondary users and cognitive node as hiding secondary user, respectively based on modulation of GMSK, OFDM and AM-DSB, as illustrated in Fig. 1 three signals were fed into the code and simulation was run at frequency of 805 MHz.

Fig. 3 shows the spectral correlation function for the GMSK signal. The modulation used here to generate this signal is quadrate phase shift keyed modulation. It is easy to decide the modulation sort from the signals profile.

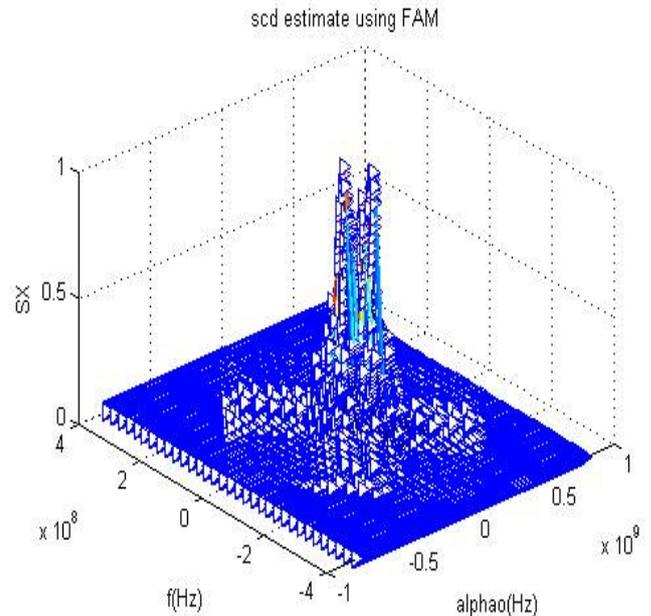


Fig. 3 Cyclic spectrum of the GMSK signal

Fig. 3 shows an overview for the received primary signal

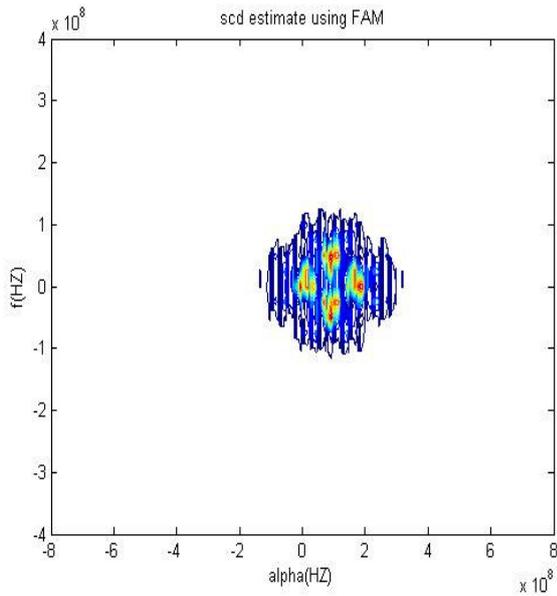


Fig. 4 Contour figure of the GSM signal

Subsequently, we examine the rental secondary user signal using the same technique. Assuming that the rental CR network transmitted signal in OFDM, the detected waveform is shown in Fig. 4.

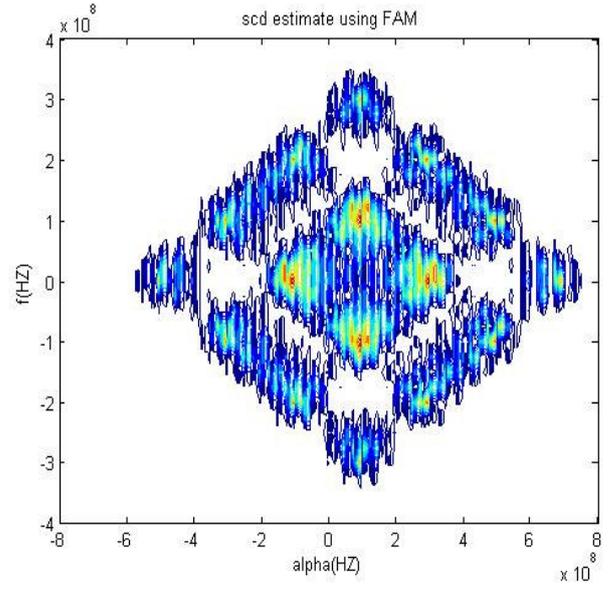


Fig. 6 Contour figure of the OFDM signal

Meanwhile, we examine the hiding secondary user signal using the same FAM technique. Assuming that the hiding CR node transmitted signal in AM signal. The modulation used here is Double side-band Large Carrier DSB-LC, the detect waveform is shown in Fig. 6.

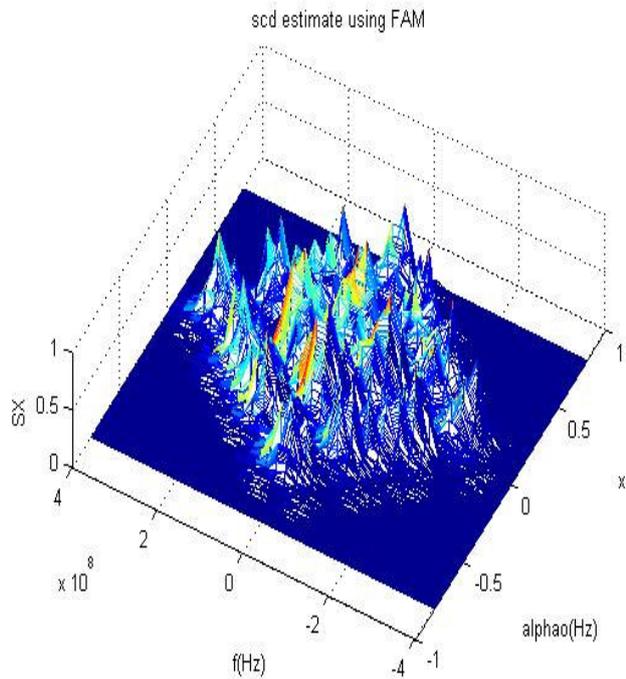


Fig. 5 Cyclic Spectrum of the OFDM signal

It is significant to look at the received signals from diverse sides. This will avoid any ambiguity in the signals modulation source category. Hence, the other outlook for the rental secondary User spectral function is show in Fig. 5.

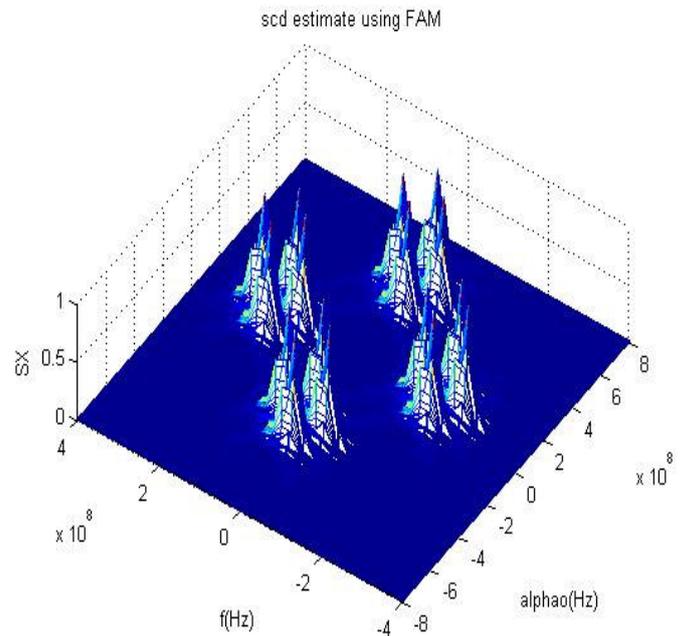


Fig. 7 Cyclic spectrum of the AM-DSB signal

The AM-DSB signal can be reobtainable also in an additional elevation sight to assure the modulation type. Fig. 7 shows overview for the received hiding secondary user signal.

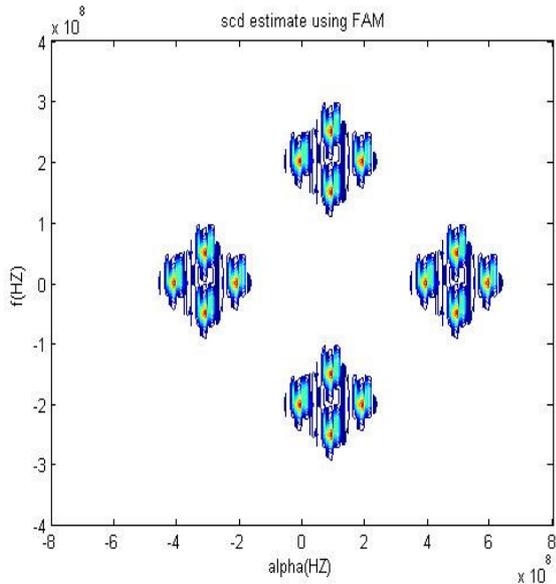


Fig. 8 Contour figure of the AM-DSB signal

It is substantially noticeable the differences between the primary, rental secondary and hiding secondary simulated signals using the FFT accumulative method. These results make this method a preferable choice for this category of signal discovery. Additionally, cyclic spectrum enables accurate examination for the signal periodic changes. Thus, minor functions resulted from wireless environmental changes and interference can be estimated, compared, and identified. The obtained results show the FAM correlation function. Detected signals in time domain are transferred to the frequency domain. Subsequently, other estimations are implemented to define each signal type. Although, this will ensure the accuracy of obtained result, it is not known the effects of the time spend in this process on the detection speed. Reasonable,  $\alpha$  changes speed affects on CAF and CS will be critical factor on the results reliability advisability. The simulation was performed assuming perfect transmission conditions without any consideration for the environmental and systematic noise.

Up to our knowledge, no efforts tried to use the FAM characters to make a distinction between different channels signals as this papers presents. We argue that a designed monitoring system can be developed using FAM and enhanced by self learning to recognize any CR prohibited transmission and deal with it at once. The learning features will let the anticipated system more efficient in introducing to various signals. Future work will be undertaken to detect hiding signals with noise by Cognitive Radio users using FAM method.

## V. CONCLUSION

A new strategy was presented for detecting the simulation transmission from cognitive radio. Cognitive radios hiding themselves by using primary channels in spectrum pooling system or by using same primary and rental secondary frequencies were detect in this paper. Our matlab simulation process three signals: primary,

rental secondary and hiding secondary transmitting on the pooling system channels. The simulated signals were analysed corresponding to the changes in their modulation techniques. FAM used as the analysing algorithm for the scenario studied. Detected signals were recognized in accordance with their frequency domain representations performed by the used method. Primary and two secondary generated signals were examined and detected clearly. The presented modification will aware managing entities for the future networks on any uncharted transmission whenever it occurs. The suggested model will be required to guarantee the reliability of the charging system.

## REFERENCES

- [1] M. Oner and F. Jondral, "Extracting the Channel Allocation Information in a Spectrum Pooling Systems exploiting system cyclostationary", 15th IEEE international Symposium on personal, Indoor and Mobile Radio Communications, 2004. PIMRC 2004, vol. 1, pp. 551-555, 5-8 Sep. 2004.
- [2] P. Si, H. Ji, F. Richard and C. M. Leung, "Optimal Cooperative Internetwork Spectrum Sharing for Cognitive Radio Systems With Spectrum Pooling," IEEE Transaction on Vehicular Technology, vol. 59, no. 4, pp. 1760-1768, May 2010.
- [3] OFCOM, "Digital dividend: cognitive access", Consultation on license-exempting cognitive device using interleaved spectrum, Publication date: 16 February 2009.
- [4] T. Yucek, H. Arslan "A Survey of Spectrum Sensing Algorithms for Cognitive Radio Applications", IEEE Communication Surveys & Tutorials, vol. 11, no. 1, First Quarter 2009. pp: 113-130.
- [5] G. Bansal, Md. Hossain, P. Kaligineedi, H. Merceir, C. Nicola, U. Phuyal, Md. Rashid, K.C. Wavegedara, Z. Hasan, M. Khabbazian, and V.K. Bhargava, "Some Research Issues in Cognitive Radio Networks", pp. 1-7 AFRICON conference proceedings 2007.
- [6] J. Chen, A. Gibson, and J. Zafara, "Cyclostationary Spectrum Detection in Cognitive Radios" IET seminar on Cognitive Radio and Software Dified Radio, pp. 1-5, September 2008. (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [7] A. Al-Dulaimi, N. Radhi and H. S. Al-Raweshidy, "Cyclostationary Detection of Undefined Secondary Users," 3<sup>rd</sup> International Conference on Next Generation Mobile Applications, Services and Technologies (NGMAST 09), pp. 1-4, September 2009.

# Restore Partitioning in MANETs with Dynamically Deployed Helping Hosts — Message Cabs (MCabs)

Ting Wang, Chor Ping Low

School of Electrical and Electronic Engineering,  
Nanyang Technological University, Singapore,  
wang0235@e.ntu.edu.sg, icplow@ntu.edu.sg

**Abstract**—Helping hosts are intensively used in various schemes to restore partitioned Mobile Ad Hoc Networks (MANETs). Most of the existing schemes offers only deterministic helping host deployment, and are thus not able to deal with fluctuating network traffic, which is a practical condition in many MANET applications. In this paper, we argue that dynamic helping hosts, with deployment that response to the changes in the traffic demand of the network, may overcome this drawback and reduce the message delay in the networks. To demonstrate the effectiveness of this observation, we propose a new helping host scheme namely the *Message Cab (MCab)* scheme for partition restoration in MANETs, and validate the performance through simulations.

**Index Terms**—Mobile Ad Hoc Networks (MANETs), Helping Host, Message Cab, Adaptive Route Design, Dynamic Deployment

## I. INTRODUCTION

A *Mobile Ad Hoc Network* (MANET), as described in [1], is a kind of mobile wireless networks which is comprised of a collection of mobile hosts connected through wireless channels. The direct connections between the hosts are referred to as *links*. While the mobility of hosts enables the network to span over a large area, it also causes a highly dynamic topology, which is a major challenge to the applications of MANETs. When a host moves out of another's communication range, the link between them breaks, and the entire message routing path may be destroyed by this broken link. This possibility of link breakage may split hosts into different parts between which there is no possible path. This in turn may result in packets not being able to reach their destinations. We refer to this phenomenon as *network partitioning*. Each isolated part of a network is referred to as a *partition* of the network. The partitioning problem makes a critical strike on ad hoc routing because most protocols typically assume that the network is always connected. To enhance the reliability and conserve energy, partitioning should be restored in MANETs.

Various approaches have been proposed for MANETs survivability and restoration. In particular, *helping hosts* are deployed to reconnect the network partitions, and we refer to such action as *partition restoration*. The helping hosts are able to reconnect the network connectivity by moving from one partition to another in a MANET despite the fact that the *normal hosts* (i.e. non-helping hosts) may be partitioned. This idea is also extensively discussed in *Delay Tolerant Networks (DTN)* and *Wireless Sensor Networks (WSN)* in order to collect

data from disconnected parts of the networks.

The helping hosts are addressed by different names in various schemes. In [2], [3], [4], [5], they are referred to as *ferries*, while in [6], [7], [8], they are called *helping nodes* or *forwarding nodes*. *Data MULEs* in [9], [10], [11] are also known as a kind of helping hosts, and in the DakNet project [12] buses are used as helping nodes to connect broadband network to rural villages.

In most of these existing schemes, such as [2], [8], [11], [12], the helping hosts are different from the normal hosts and are assigned as helping hosts prior to the commencement of network operations. The deployment is thus *static*, or *deterministic*. Since the number of helping hosts is fixed, it may turns out that there is too many, or too few helping hosts to meet the traffic demand in the network. Thus static deployment may not be adaptive to network traffic demand. To overcome this drawback, *dynamic deployment* is used in [6], where normal hosts with suitable moving direction and speed are chosen as helping hosts (*a.k.a helping nodes*). However, it is a centralized scheme and thus not scalable with network size. More importantly, under the assumption that the network is partitioned, it is infeasible to make all the hosts' movement information available to a central server to perform the selection. Similar problem can also be observed in [7].

Therefore, we propose a localized algorithm, namely the *Dynamic Cab Deployment (DCD)* algorithm to deploy helping hosts — *Message Cabs (MCabs)* — in this paper. The DCD algorithm allows the number of helping hosts to change dynamically with the volume of traffic, and thereby reduces the weighted average delay of messages in the network, and enhances the scalability of the deployment process. We will demonstrate that the MCab scheme overcomes the drawbacks of existing schemes by incurring lower average message delay and the results are validated through simulations.

In the following parts of this paper, we introduce the backgrounds of our work in Section II, and the model together with our objectives in Section III. The main part of the *Message Cab (MCab)* scheme, namely the *Dynamic Cab Deployment (DCD)* algorithm is presented in Section IV. Simulation and results are discussed in Section V, while Section VI concludes this paper.

## II. ASSUMPTIONS AND NOTATIONS

While many existing schemes (e.g. [2], [5], [9]) focus on stationary hosts with known locations and predictable network

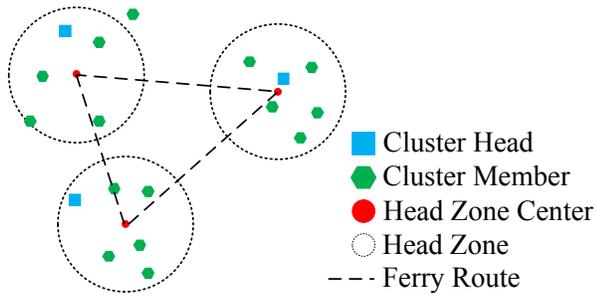


Fig. 1: Network Topology

traffic, we try to deal with a more practical scenario, where:

- the hosts are mobile;
- the traffic in the network is randomly variant and is thus not predictable.

As discussed in [13], movements of hosts in MANETs can be considered as *Levy Flights*, in which the hosts tend to stay in a certain area for a long time and occasionally make long distance flights to places far away. We could thus assume that a host stays in the same cluster for a certain period of time, and will also sometimes decides to move to another cluster. Practically, this could be due to the assignment of tasks to the host, such as in a wireless sensor network, a sensor may need to reallocate itself to new positions to collect new data.

Let's assume that the MANET of our interest has  $n_h$  mobile hosts. We assume the network is partitioned to  $n_c$  components, each of which forms a *cluster* with a chosen *cluster head*. The hosts in the same cluster are always connected. In order to simplify the problem, we restrict the movement of a cluster head to be inside a circular area, which is referred to as the *head zone*. The radius of head zone is equal to the communication range of the cluster head, denoted as  $r$ , as shown in Fig. 1. The center of head zone of cluster  $s$  is denoted as point  $C_s$ ,  $1 \leq s \leq n_c$ . We use  $\epsilon_{sd}$  to denote the time taken by a host to travel from  $C_s$  to  $C_d$ ,  $1 \leq s, d \leq n_c$ . We note that  $\epsilon_{sd}$  is a positive real number. We say cluster  $d$  is a *neighbor* of cluster  $s$  if a host is able to move directly to cluster  $d$  from cluster  $s$  in one hop.

Using proper clustering schemes, such as [14], the cluster head will always be aware of the changes in its cluster members, such as the movements, arrivals of new members and departures of existing members. It also knows the locations of other clusters, so when a host decides to move to a new place, the cluster head knows which is the next cluster the host is going to join. In addition, the cluster head works as a proxy between other normal hosts in the cluster and the helping host as well. If a helping host stops at  $C_s$ , it is able to communicate with the head of cluster  $s$  and deliver the messages to other hosts in cluster  $s$  via the cluster head. The route of a helping host will be a sequence of head zone centers ( $C_s$ 's).

### III. SYSTEM MODEL AND OBJECTIVES

To demonstrate our methodology of the M Cab scheme, we describe an analogy between the MANET and a simple

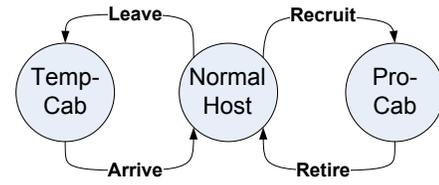


Fig. 2: States of a Host

transportation system which serves several small towns. It will also explain the reason why we name our helping hosts as *cabs* and how they are different from *ferries*.

The towns represent the clusters, each of which is distant away from the others. The cars are like the hosts — they usually move within a town, and when it is necessary, they are also able to move from one town to another. The passengers are akin to messages, which by themselves cannot move from town to town.

*Buses* are provided as an existing solution to the inter-town transportation problem. They are similar to the other statically deployed helping hosts (such as message ferries) that we have discussed in the previous section. The static deployment and fixed route design of buses perform poorly with varying volume of passengers. Hence, it would be preferable if we have a flexible solution.

Besides taking buses, the passengers could consider hiring a *cab* to travel to another town. The route of a cab will depend on the passengers' demands, and is thus more flexible than the bus routes. Practically, it is also more convenient in terms of saving time to hire a cab. The number of cabs is determined by how many passengers there are. When there are more passengers, more cabs can be recruited from the private cars (normal hosts) available; when the number of passengers drops, some cabs could retire and become private cars again.

Moreover, since it is possible that some cars will move to another town by their drivers' own decisions, a passenger may also take a ride from a private car which is also moving to the his/her destination town. This will utilize the mobility of private cars in the transportation system to serve the passengers' needs and save their traveling time. A car offering ride to passengers is like a one-time temporary cab that only works for a single trip to a particular town. We refer to such cars as *temp-cabs*. To distinguish from them, we refer to those cabs that works for multiple trips in a longer period of time as professional cabs, or *pro-cabs*.

We imitate the method of hiring pro-cabs and taking rides from temp-cabs in the MANETs and propose the *Message Cab (MCab)* scheme. As shown in Fig. 2, the state of a normal host, which is not a cab nor a cluster head, can transform to a *pro-cab* or a *temp-cab* under the control of the *Dynamic Cab Deployment (DCD)* algorithm.

The key difference between message cabs and the other types of helping hosts in the existing works (such as ferries, helping nodes *etc.*) is that they are not selected before the network starts. Therefore the number of cabs and cab routes can dynamically change with the traffic. Although this kind of deployment does not allow the helping hosts to have higher

capability in moving speed or storage space (since message cabs are just selected normal hosts), we found that it can still effectively reduce the message delay in the network.

We note that since hosts in the same cluster are connected, the delay incurred by the message transmission within a cluster (*intra-cluster communication*) is much smaller than the delivery between different clusters (*inter-cluster communication*), and is less relevant to the cab deployment route design. As a consequence, intra-cluster communication will be omitted in our following discussion.

We should note that before a message is collected by a cab, it needs to wait at some cluster head for a certain time duration. We refer to this amount of time spent on waiting as the *waiting delay* ( $\omega$ ) of the message. After it is collected by a cab, the cab will travel from its source cluster to its destination cluster, and thus incurs a *traveling delay* ( $\tau$ ) for the message. Therefore, the *overall delay*  $\delta$  of a message is  $\delta = \omega + \tau$ .

Assume within time duration  $(0, t]$ , there are  $n_m$  messages that have been transmitted by the cabs. The size of message  $i$  ( $1 \leq i \leq n_m$ ) is  $\mu_i$ . The waiting, traveling and overall delay of message  $i$  are denoted as  $\omega_i$ ,  $\tau_i$  and  $\delta_i$  respectively. Similar to the objectives in [2], we are interested in reducing the *weighted average overall delay* ( $\Delta$ ) of the messages:

$$\Delta = \frac{\sum_{i=1}^{n_m} \mu_i \delta_i}{\sum_{i=1}^{n_m} \mu_i} = \frac{\sum_{i=1}^{n_m} \mu_i (\omega_i + \tau_i)}{\sum_{i=1}^{n_m} \mu_i} = \Delta_\omega + \Delta_\tau, \quad (1)$$

where the *weighted average waiting delay* ( $\Delta_\omega$ ) and the *weighted average traveling delay* ( $\Delta_\tau$ ) of the messages are given by

$$\Delta_\omega = \frac{\sum_{i=1}^{n_m} \mu_i \omega_i}{\sum_{i=1}^{n_m} \mu_i}, \quad \text{and} \quad \Delta_\tau = \frac{\sum_{i=1}^{n_m} \mu_i \tau_i}{\sum_{i=1}^{n_m} \mu_i}. \quad (2)$$

We can see that if there are more cabs in the network, each cluster could be visited more frequently, and the waiting delay of messages reduces. It shows that the waiting delay ( $\Delta_\omega$ ) is closely related to the cab deployment plan. On the other hand, optimization of cab routes results in a shorter traveling delay  $\Delta_\tau$ .

#### IV. DYNAMIC DEPLOYMENT OF MESSAGE CABS

In the ideal scenario, once a inter-cluster message is received by the cluster head, there is always a cab ready to carry it towards its destination, and  $\Delta_\omega$  will be 0. However, in practise, with unpredictable and fluctuating traffic, it is impossible to have a cab ready for message delivery as soon as a message arrives at the cluster head. It would also be inefficient if a cab carries only one message and does not fully utilize its storage space. To solve this problem, we try to make use of the normal hosts mobility as a *temp-cab* to give a ride to the messages as well as to recruit *pro-cabs* from normal hosts in the DCD algorithm. We show that by doing so, we are able to bound the weighted waiting delay of

messages from above by a predetermined value, denoted as  $\Omega$  seconds.

Since the change of cluster membership is handled by the cluster head, a host (say host  $a$ ) needs to report to the head of its current cluster, say cluster  $s$  (as source) before it leaves for another cluster (say cluster  $d$ , as destination). It becomes possible that the head of cluster  $s$  lets host  $a$  to carry some messages which have cluster  $d$  as their destinations. Host  $a$  will then deliver these messages to the cluster head  $d$  when it arrives there and registers itself as a new cluster member. Therefore, when a host is leaving a cluster, its state changes from normal host to temp-cab, and it is used to deliver as many messages as possible across the partitioned clusters. Upon its arrival, it delivers the messages to the head of the new cluster and its state changes back to that of a normal host. This procedure of deploying a temp-cab is depicted in Fig. 3a.

On the other hand, a timer is used to control the time when a *pro-cab* should be selected. Assuming there are  $n'_m$  inter-cluster messages stored in cluster head  $s$ , of each the size is  $\mu_i$ . The *current waiting delay* of a message is the length of the time duration since the message is received by the cluster head, and is denoted as  $\omega'_i$ . The timer is set to

$$t = \Omega - \frac{\sum_{i=1}^{n'_m} \mu_i \omega'_i}{\sum_{i=1}^{n'_m} \mu_i},$$

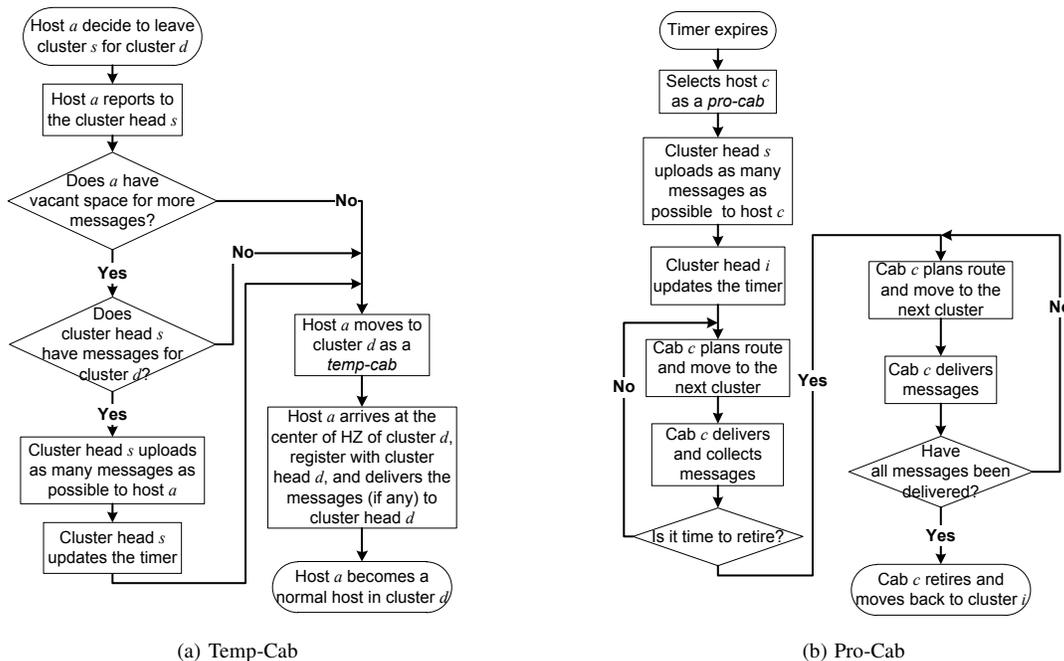
so that if a pro-cab is recruited right before the timer expires, the average weighted delay of the messages stored in cluster head  $s$  will be less than  $\Omega$ , which is the required upper bound. We note that the value of  $t$  needs to be updated every time there is a change in the number of inter-cluster messages which are stored in cluster head  $s$ , such as when a new inter-cluster message is received and stored, or when some messages have been uploaded to a cab (either temp or pro) by cluster head  $s$ .

We let the time duration between the recruitment and retirement of a cab be denoted as  $\Theta$ . After having served for  $\Theta$  seconds as a pro-cab, it retires. Upon its retirement, a pro-cab does not stop immediately. It continues moving among the clusters to deliver the remaining messages it has already collected, but without collecting new messages. When all the messages are delivered, it moves back to the cluster where it has been recruited, and register with the head as a normal host again.

Fig. 3 depicts the flowcharts for the DCD algorithm. To avoid going into too much details, we make some simplifications to the algorithm:

- The pro-cabs are randomly selected from the normal hosts in the cluster;
- The messages are dropped according to the *hot-potato* rule.

By the hot-potato rule we mean that we deal with the elements that have heaviest impact to the system by either processing them or dropping them first. By dropping those messages with


 Fig. 3: Flowcharts of the *Dynamic Cab Deployment (DCD)* Algorithm

heavier weighted delay, the average weighted delay of the successfully delivered messages can be reduced.

With these simplifications, the DCD algorithm is a fully distributed scheme which does not require any global information. Moreover, it is not difficult to see that the worst case complexity of the DCD algorithm is  $O(n_m)$ , where  $n_m$  is the number of messages that have been generated during the period of network operation.

## V. SIMULATION RESULTS

To validate the effectiveness of the M Cab scheme, extensive simulations are carried out. We randomly generate the network topology and traffic in C++ programs and compare the results with existing schemes.

We model a MANET constructed in a 500m by 500m area. The number of hosts in each cluster is set to 20 at the beginning of the simulations. The communication range of the hosts is 5m, and the storage size is 10Mb. The hosts move at a speed of 10m/s.

The traffic in the MANET can be specified by the number of messages per unit time, but as the size of messages also varies, it is more convenient to describe it as the size of data transmitted per unit time, *i.e.* kilobits per second (kb/s). We refer it as the *volume* of the traffic.

The number of pro-cabs may vary in the simulation. We define the *effective number of pro-cabs* as the average number of pro-cabs over time to describe how many pro-cabs are deployed in the MANET.

### A. Number of Pro-Cabs

We start with our study on the performance of the DCD algorithm. The main objective of the DCD algorithm is to

TABLE I: Effective Number of Pro-Cabs

	$\rho = 5000s$		$\rho = 200s$	
	$\Theta = 200s$	$\Theta = 2000s$	$\Theta = 200s$	$\Theta = 2000s$
$\Omega = 100s$	2.81	3.78	1.13	1.35
$\Omega = 500s$	0.98	1.49	0.15	0.44

dynamically recruit cabs in the MANET to deliver inter-cluster messages. The algorithm is controlled by two parameters, namely  $\Omega$  (the upper bound of the weighted waiting delay) and  $\Theta$  (the time duration that a host serves as a pro-cab). To deploy temp-cabs, the performance of the DCD algorithm is also closely related to the frequency in which the hosts move from one cluster to another. A variable  $\rho$  is defined as the average length of period (in seconds) that a host stays in a cluster before it moves to another one.

Fig. 4 displays how the number of recruited cabs changes with the volume of inter-cluster traffic. The left vertical axis corresponds to the traffic volume while the number of pro-cabs are indicated on the right axis. To study the impact of  $\Omega$ ,  $\Theta$  and  $\rho$  to the performance of the DCD algorithm, we choose two (high and low) values for each of these variables. Each of the figures corresponds to a particular combination of the values of  $\Omega$  and  $\rho$ , and plots the two groups of results which correspond to the two values of  $\Theta$ . As shown in Table I, the values of  $\Omega$  are 500s and 100s; the values of  $\Theta$  are 2000s and 200s; the values of  $\rho$  are 5000s and 200s. The effective number of pro-cabs are also shown in Table I.

1) *The value of  $\rho$* : The value of  $\rho$  significantly changes the effective number of pro-cabs.

With a high value of  $\rho$  ( $\rho = 5000s$ ), the hosts tend to stay in the same cluster for a long time, and thus there are

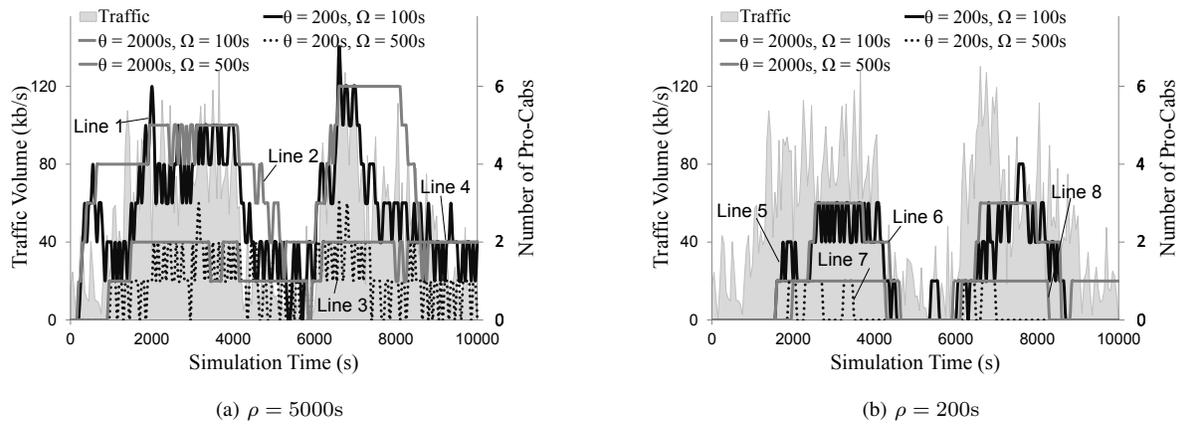


Fig. 4: Performance of the DCD Algorithm

less chances that messages could be delivered by a temp-cab. Therefore, more pro-cabs have to be recruited to keep the waiting delay below  $\Omega$ .

If the value of  $\rho$  is low ( $\rho = 200s$ ), it implies that the hosts are frequently recruited and retired, and lines 1, 3, 5 and 7 appear to be more “spiky” than line 2, 4, 6 and 8 respectively in Fig. 4.

3) *The value of  $\Omega$ :* The value of  $\Omega$  influences the number of cabs more directly. When the value of  $\Omega$  is high ( $\Omega = 500s$ ), the cluster heads are able to tolerate larger weighted waiting delay before recruiting new pro-cabs, and less pro-cabs will be hired in the MANETs. This is the reason why in Table I, the effective numbers of pro-cabs are much smaller when  $\Omega = 500s$ . On the other hand, if the value  $\Omega$  is low ( $\Omega = 100s$ ), the cluster heads have to frequently hire new pro-cabs to keep the waiting delay of the messages low. From Fig. 4 it is easy to observe that the number of pro-cabs significantly increases when the value of  $\Omega$  is low.

2) *The value of  $\Theta$ :*  $\Theta$  controls the length of duration for which a host serves as a pro-cab. It controls the frequency for triggering the pro-cab recruiting procedure, and thus affects how fast the system could respond to the change in traffic volume.

When the value of  $\Theta$  is high ( $\Theta = 2000s$ ), the pro-cabs have a long service time and a slow retirement. Comparing line 1 with line 2 in Fig. 4a on the time interval  $[7000, 8000]$ , we can see that the number of cabs stays as high as 6 when  $\Theta = 2000s$  (as depicted by line 2) even though the traffic volume has already dropped to a lower level, where only about 3 pro-cabs will be deployed if  $\Theta = 200s$  (as depicted by line 1), meaning that several cabs may be unnecessary for the purpose of keeping the weighted waiting delay low. Similar phenomena can also be observed in the time intervals  $[4000, 5000]$  and  $[9000, 10000]$  in Fig. 4a, and  $[8500, 10000]$  in Fig. 4b, resulting in the effective numbers of cabs being much larger when  $\Theta = 2000s$  (as shown in Table I).

A low value of  $\Theta$  ( $\Theta = 200s$ ) causes the pro-cabs to only serve for a short period of time, and change back to the state of normal host sooner. However, messages are still

being generated and stored in the cluster heads, causing new pro-cabs being recruited. The change in the cab number is thus more rapid and frequent than when  $\Theta = 2000s$ . Since cabs are frequently recruited and retired, and lines 1, 3, 5 and 7 appear to be more “spiky” than line 2, 4, 6 and 8 respectively in Fig. 4.

3) *The value of  $\Omega$ :* The value of  $\Omega$  influences the number of cabs more directly.

When the value of  $\Omega$  is high ( $\Omega = 500s$ ), the cluster heads are able to tolerate larger weighted waiting delay before recruiting new pro-cabs, and less pro-cabs will be hired in the MANETs. This is the reason why in Table I, the effective numbers of pro-cabs are much smaller when  $\Omega = 500s$ .

On the other hand, if the value  $\Omega$  is low ( $\Omega = 100s$ ), the cluster heads have to frequently hire new pro-cabs to keep the waiting delay of the messages low. From Fig. 4 it is easy to observe that the number of pro-cabs significantly increases when the value of  $\Omega$  is low.

## B. Delay of Messages

One of the advantages of the M Cab scheme is that it can be used with different types of routes, since route design and helping host deployment are often solved as two separate problem in the existing schemes, although their performance can be inter-related, as we will show in this section. Two route design algorithms for helping host are considered, namely the TSP route (used in [2]) and the *Adaptive Message Ferry Route (aMFR)* (proposed in [15]). We compare the values of message delay caused by these types of route with and without the DCD algorithm. A single helping host is deployed in these two cases.

We demonstrate this result in Fig. 5 by showing how the M Cab scheme perform when there are different number of clusters ( $n_c$ ). In each figure, two values of  $\rho$  (5000s and 200s) are used in the simulation. Two different implementations of the DCD algorithm —  $\Omega = 100s, \Theta = 2000s$  (denoted as DCD-a and  $\Omega = 500s, \Theta = 200s$  (denoted as DCD-b) — with  $\rho = 5000s$  are demonstrated. The results for the other two cases ( $\Omega = 500s, \Theta = 2000s$  and  $\Omega = 100s, \Theta = 200s$ )

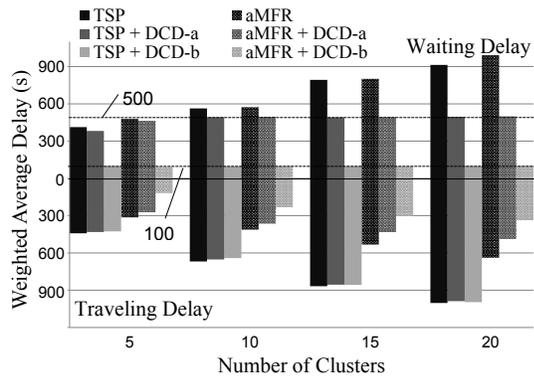


Fig. 5: Delay of Messages

and other values of  $\rho$  exhibit similar results, and are thus omitted in this section.

In Fig. 5 the simulation results are shown as a group of columns. The upper part (above 0) of each column shows the duration of average waiting delay  $\Delta_w$ , and the lower part (below 0) shows the duration of average traveling delay  $\Delta_r$ . As a result, the entire bar length shows the overall average delay  $\Delta$  of the corresponding scheme.

It can be seen from Fig. 5 that the total delay is much lower with the DCD algorithm for both routes. For the TSP routes, since every helping host follow exactly the same route, the traveling delay of the messages delivered by the pro-cabs will be the same, independent to the number of pro-cabs. However, temp-cabs deliver messages in a single hop, and therefore causes less delay than the pro-cabs that follows the TSP route, and thus we can still observe from the lower parts of the solid columns that the message traveling delay is also slightly reduced by the DCD algorithm. The waiting delay of the messages is reduced by recruiting more pro-cabs when the current average waiting delay is about to exceed the threshold  $\Omega$ . As shown in the graph, the values of waiting delay with DCD-a and DCD-b are bounded from above by the values of 100s and 500s, respectively.

The reduced message waiting delay can also be observed from the results for aMFR routes. Adaptive routes are used in aMFR, where a helping host's route depends on the messages carried by it. When more helping hosts are deployed, each of them will carry less messages on average. Thus a message can be delivered with less hops, and the traveling delay can also be reduced as shown in [15]. This is also shown by the dotted columns in Fig. 5.

## VI. CONCLUSION

In this paper, we propose a scheme with flexible helping hosts, namely *Message Cab (MCab)* for message delivery in partitioned MANETs. It uses the *Dynamic Cab Deployment (DCD)* algorithm to select helping hosts (MCabs) according to the traffic volume in the network. Comparing with the existing schemes with helping hosts, we have shown that the MCab scheme effectively shorten the delay of the messages with

the simulations and is adaptive to the varying traffic in the network, which has not been discussed in the existing works.

## REFERENCES

- [1] S. Corson and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations," United States, 1999.
- [2] W. Zhao and M. Ammar, "Message Ferrying: Proactive Routing in Highly-Partitioned Wireless Ad Hoc Networks," in *Proc. of the Ninth IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS'03)*, Washington, DC, USA, 2003, pp. 308–314.
- [3] M. M. B. Tariq, M. Ammar, and E. Zegura, "Message Ferry Route Design for Sparse Ad Hoc Networks with Mobile Nodes," in *Proceedings of the 7th ACM international symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '06)*. New York, NY, USA: ACM, 2006, pp. 37–48.
- [4] M. Ye, X. Tang, and D. L. Lee, "Fair Delay Tolerant Mobile Data Ferrying," in *Proce. 10th International Conference on Mobile Data Management: Systems, Services and Middleware (MDM'09)*. Washington, DC, USA: IEEE Computer Society, 2009, pp. 182–191.
- [5] M. H. Ammar, D. Chakrabarty, A. D. Sarma, S. Kalyanasundaram, and R. J. Lipton, "Algorithms for Message Ferrying on Mobile Ad Hoc Networks," in *Proc. the IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'09)*, 2009, pp. 13–24.
- [6] Q. Li and D. Rus, "Communication in Disconnected Ad Hoc Networks Using Message Relay," *J. Parallel Distrib. Comput.*, vol. 63, no. 1, pp. 75–86, 2003.
- [7] S.-H. Chung, K.-F. Su, C.-H. Chou, and H. C. Jiau, "Improving Data Transmission with Helping Nodes for Geographical Ad Hoc Routing," *Computer Networks*, vol. 51, pp. 4997 – 5010, 2007.
- [8] C.-H. Ou, K.-F. Su, and H. C. Jiau, "Connecting Network Partitions with Location-Assisted Forwarding Nodes in Mobile Ad Hoc Environments," in *Proc. 10th IEEE Pacific Rim International Symposium on Dependable Computing (PRDC'04)*, Washington, DC, USA, 2004, pp. 239–247.
- [9] D. Borsetti, C. Casetti, C.-F. Chiasserini, M. Fiore, and J. M. Barceló-Ordinas, "Virtual Data Mules for Data Collection in Road-Side Sensor Networks," in *Proc. 2nd International Workshop on Mobile Opportunistic Networking (MobiOpp'10)*. New York, NY, USA: ACM, 2010, pp. 32–40.
- [10] R. C. Shah, S. Roy, S. Jain, and W. Brunett, "Data MULEs: Modeling a Three-tier Architecture for Sparse Sensor Networks," Intel Research, Tech. Rep. IRS-TR-03-001, 2003.
- [11] D. Jea, A. Somasundara, and M. Srivastava, "Multiple Controlled Mobile Elements (Data Mules) for Data Collection in Sensor Networks," *Lecture Notes in Computer Science*, vol. 3560, pp. 244–257, 2005.
- [12] A. A. Hasson, R. Fletcher, and A. Pentland, "DakNet: A Road to Universal Broadband Connectivity," First Mile Solutions, Tech. Rep., 2003.
- [13] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong, "On the Levy-walk Nature of Human Mobility: Do Humans Walk like Monkeys?" in *Proc. of the 27th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '08)*, 2008.
- [14] Y. Zhang, C. P. Low, J. M. Ng, and T. Wang, "An Efficient Group Partition Prediction Scheme for MANETs," in *Proceedings of the 2009 IEEE Wireless Communications and Networking Conference (WCNC '09)*, 2009, pp. 1–6.
- [15] T. Wang and C. P. Low, "Adaptive Message Ferry Route (aMFR) for Partitioned MANETs," in *Proc. of the 2nd International Conference on Mobile Lightweight Wireless Systems (MOBILIGHT '10)*, 2010, p. TBA.

# Increasing the Speed of a Wireless Network by Processing Indoor

Alexey Lagunov

Department Computer Science and Electronic Devices  
Pomorsky State University of name M.V. Lomonosov  
Archangelsk, Russian Federation  
lagunov@pomorsu.ru

**Abstract**— Usage of technology MIMO in standard 802.11n allows to increase data transmission to 300 Mbit/s. This network successful operation is necessary to send signals from transmitting antenna to receiving antenna in different ways, that is there is multibeam reflection an indoor. This research is suggested to obtain the optimal interfection time by special indoor treatment. The radio wave propagation statistical theory is consider by Author an indoor, the time Interfection T defenition formula is derived. According to author conclusions the time T can be selected by indoor special treatment: disposing of surface with high reflection coefficient indoors. Time T increases from 210 ns (before indoor treatment) to 390 ns(after room treatment). Increasing time T more then 400 ns author doesn't recommended, as «guard time» may be exceeded and required result does not achive. Practical researches with standard 802.11n equipment using confirm the theoretical conclusions and show, that data transmission rate after treatment increases to 10-15% upon the average.

**Keywords** - *RadioEthernet; wireless, interference; disturbance; Wi-Fi, reflectance.*

## I. INTRODUCTION

In the modern communication system intercomputer channel bandwidth is need to increase. The bandwidth may be increased by frequency band expansion or radiating power rise. Nevertheless such methods have disadvantages because of human ecological safety and electromagnetic compatibility frequency band expansion or radiating power rise is limited. So, if the required data rate isn't provided by possible frequency band expansion or radiating power rise, one of the effective problem solution may be usage of adaptive antenna arrays with a weakly correlated antenna elements. Communication systems, which used such antenna, is named MIMO system (Multiple Input Multiple Output). Such system uses at making of wireless network of IEEE 802.11n standard, which was approved in September 11, 2009.

Standard 802.11n the data rate increases in almost four times compared to the 802.11g standard devices( which maximal rate is equal to 54 Mbps) on conditions that it uses in regime 802.11n with other with device 802.11n.

High date rate is provided by:

1. Large quantity of bearing usage: 802.11g with 48

bearing against 802.11n with 52 – increases bandwidth from 54 Mbps to 58.5 Mbps.

2. FEC (Forward Error Correction): 802.11g has maximal value FEC - 3/4, 802.11n — 5/6. Increased bandwidth from 58.5 Mbps to 65 Mbps.

3. Guard Interval: 802.11a guard interval between transmission is 800 ns. In 802.11n it is decreased to 400 ns, so bandwidth increases from Mbps to 72.2 Mbps.

4. MIMO: spatial multiplexing can support a linear increase in bandwidth depending on the number of antennas on each side. The standard maximum permitted number of antennas is four. This allows to increase the total bandwidth up to 288.9 Mbps the subject to 72.2 Mbps channel.

5. Channels 40 MHz: all previous versions of 802.11 have the bandwidth 20MHz. 802.11n has a mode for the bearing channel width of 40 MHz. Due to the increasing amount of data bandwidth per channel is increased from 52 to 108. This increases the overall throughput to 150 Mbps. Taking into account the 4 channels of MIMO technology can theoretically get the speed of 600 Mbps.

802.11n devices work in the range 2.4 - 2.5 or 5.0 GHz. In this case, you can use one of three modes:

1. HT Greenfield mode - means that the next device does not use the old standards on the same frequency. If they try to do this, they will not be able to connect at 802.11n.

2. Non-HT (Legacy) mode - 802.11n AP in a Non-HT mode sends all packets in the old format 802.11a/g, so all devices can understand it.

3. HT Mixed mode - as soon as the majority points 802.11n will work in device support 802.11b/g and 802.11a. HT Enhancements will be used at the same time with the mechanism of HT Protection, which will enable devices to work with old cards.

During the ten-year period of work the following problems is identified and partially solved:

- 1.The unpredictability of coverage. In previous versions of wireless coverage was close to spherical. In networks with 802.11n coverage area has a complex structure and difficult to calculation and planning.

2. In the range of 2.4 GHz can only use one channel of 40 MHz. The most perspective usage of such a regime is in the range of 5GHz.

3. Network performance 802.11n in the range of 2.4 - 2.5

GHz decreases due to the small value of signal to noise ratio (SNR) due to the presence of a large number of devices operating in the same range: Bluetooth, Microwave Ovens, Cordless phones, Wireless Cameras, Game controllers, Baby Monitors, Digital video devices and many more. It is recommended to use a range of 5.0 GHz, in which the noise is much less.

4. When the bandwidth of a cable connection wireless access points to the wired network is more than 100 Mbps, the trouble in the way of traffic becomes a wireless controller, a speed which is recommended to increase up to 1 Gbps.

5. Wireless network security: a large coverage area implies an increase in risk of unauthorized access, illegitimate passive devices can listen to the broadcast, unnoticed, Fluke Networks products can detect the device (including passive) connected to the network 802.11n; 40MHz channel scan takes twice as much time than scanning 20MHz channel mechanism of block confirmation of 802.11n is the vulnerability to DoS attacks.

6. One access point can connect up to 254 users. The bandwidth of a single half-duplex radio channel is divided into all simultaneously transmitting users. Given the numerous repetitions of data transmission because of conflicts with their simultaneous transmission of real bandwidth per user goes down to very small quantities. Therefore, the number of concurrent clients in terms of a single point of access must be in the range of 12 to 50. For example, this problem manifested itself during the Mobile World Congress [1].

MIMO enables omni-directional transmission on a lot of antenna pairs. This ensures they get the signal receiver and is called spatial streams. Data packet can be divided into several packets and sent on different streams. The receiver collects them back into a single data packet. 802.11n supports up to 4 threads. But more often considered the standard antenna configuration circuit for transmitting and receiving information,  $3 \times 3$  or  $2 \times 3$ .

Many researchers engages, eg [2,3], to raise productivity wireless IEEE 802.11n standard. These works are devoted to the advancement of the antennas and radio devices. A distinctive feature of our work is that we are engaged in the advancement of the transmission medium, since the success of MIMO operation mode is necessary to establish such a regime under which the multipath propagation of radio waves. We limit the wireless network work in one room. This regime occurs when installing access points in places of public access: restaurants, libraries, hotel lobby, etc. The aim of this work - forming the best possible conditions of multibeam propagation in a room and experimental verification of the results.

## II. THE STATISTICAL THEORY OF RADIOWAVES DISTRIBUTION IN INDOOR

With the standard 802.11 POP in an indoor the signal doesn't only reach to subscriber P on the right line, but also

on the ways repeatedly broken through the signal reflections from walls, a ceiling and a floor indoors. At each signal's reflection from surfaces, limiting an indoor the some part of signal's energy is absorbed, in spite of all this the reflected signal level at the subscriber will be quite enough for creating a handicap, named interference, at the expense of a phase difference.

Let's examine a definition problem of time, during which the interferential handicaps level will reduced to the acceptable signal-to-noise ratio value. Let's name the time T as interreflection time.

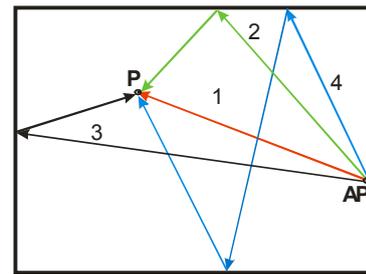


Figure 1. Repeated reflexion of radio-waves indoors

Through each point in indoor volume the great number of the reflected waves passes moving on various directions at a time. Field strength and vibrational speed vector in each point are determined by result of an interference of all these waves. In this case to calculate a signal level in each point is represented with difficulty sufficiently. However, the great number of interference waves suggests the problem to be solved by statistical treatment with usage of the mathematical probability theory basis at least in some cases. Assuming, that orientations, amplitudes and phases of waves superimposed on each other are distributed more or less chaotically, we can examine these waves as not coherent and consider the energy density in each point to be the sum of the energy density connected to each these waves. If wave movement in an indoor really has such indigested character without presence of oscillatory movement prevailing directions and symmetry in amplitudes distribution, research statistical methods are quite right and lead to the important practical results [4].

Statistical research of electromagnetic fields in the closed indoors requires prior average time definition between two consecutive signal reflections from surfaces, limiting a indoors, and also an reflections average undergone by radio-waves in unit of time. Let's imagine the surface element  $dS$ , limiting volume  $V$  of an indoor and a radio-wave, moving at an angle of  $\theta$  to an element  $dS$  normal (Fig.2). What probability that this wave will fall to platform  $dS$  per unit time from  $t$  to  $t+dt$ ? We take notice, that during  $dt$  on platform  $dS$  only the electromagnetic wave energy, being contained in volume of the cylinder with basis  $dS$  and height  $c_0 dt \cos\theta$  (where  $c_0$  - speed of distribution of a radio-wave), can fall at an angle of  $\theta$ .

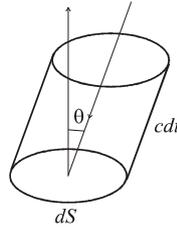


Figure 2. Cylinder

The cylinder volume is  $dV = c_0 dt \cos \theta dS$

The required probability is determined by the relation volume dV to the indoor volume V:

$$W_1 = \frac{1}{2} \frac{dV}{V} = \frac{1}{2} \frac{c_0 dt \cos \theta dS}{V} \quad (1)$$

The multiplier 1/2 explains that a wave movement can be in the platform direction, or in an opposite direction with equal probability.

It is necessary to take into account, what probability will be that some wave chosen randomly pass at an angle of  $\theta$  to dS. For probability definition we will surround a platform dS with a hemisphere of any radius r and allocate band on a hemisphere with the area  $2\pi r \sin \theta \cdot rd\theta$ , on which the radius - vector r forms with a normal dS at an angle of  $\theta$  (Fig.3).

The probability of falling wave at an angle between  $\theta$  and  $\theta+d\theta$  is the correlation of the area band to a surface to a hemisphere  $2\pi r^2$ :

$$W_2 = \frac{2\pi r \sin \theta \cdot rd\theta}{2\pi r^2} = \sin \theta d\theta \quad (2)$$

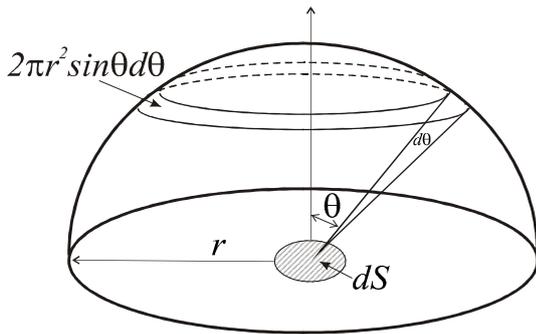


Figure 3. Hemisphere

Now it is possible noting, what probability of signal's falling on a platform dS per unit time dt at an angle between  $\theta$  and  $\theta+d\theta$  is:

$$W_3 = W_1 W_2 = \frac{1}{2} \frac{c_0 dt dS}{V} \cos \theta \sin \theta d\theta \quad (3)$$

According to additive rule the probability of a signal falling on dS at an angle in the range  $[0, \pi]$ , related to the same time interval dt, results by integration of this result on  $\theta$ :

$$W_4 = \frac{1}{2} \frac{c_0 dt dS}{V} \int_0^{\pi/2} \cos \theta \sin \theta d\theta = \frac{c_0 dt dS}{4V} \quad (4)$$

The probability of a signal falling per unit time dt at an angle and on any element of a surface S, limiting volume V of an indoor is:

$$W = \frac{c_0 dt}{4V} \int_S dS = \frac{c_0 S}{4V} dt \quad (5)$$

The found probability is proportional to any time interval dt. Increasing this interval, eventually, we will receive such time interval  $\tau$ , within range of signal falling on any surface S point should be exactly. As the authentic event probability is equal to one, required interval  $\tau$ , between two consecutive signal reflexions, will be defined from condition

$$\frac{c_0 S}{4V} \tau = 1 \quad \tau = \frac{4V}{c_0 S} \quad (6)$$

The average of reflections per unit time is

$$\nu = \frac{1}{\tau} = \frac{c_0 S}{4V} \quad (7)$$

The radiowave mean free path is

$$\lambda = c_0 \tau = \frac{4V}{S} \quad (8)$$

As mentioned above, some part of energy is absorbed at each signal reflection from surfaces limiting an indoor. It is necessary to specify an absorptivity quantitative measure of these surfaces for calculation of signal reflexions duration obviously. Let's designate symbol  $\alpha$  as an average value of energy relative reduction at each separate absorption of a signal

$$\alpha = \frac{\Delta E}{E} \quad (9)$$

There is E - the general energy store of indoors before the absorption act,  $\Delta E$  - average size of the energy absorbed by an indoor surface per the one act. It is clear, that the size, defined by the formula (9), has only statistical meaning: as separate signal's absorption acts occur place at falling of

waves under different angles on the surfaces possessing, generally speaking, various absorbed ability, the energy  $\Delta E$  absorbed per the one act, can actually have the various values being subject to statistical averaging at invariable size  $E$ . Last condition is satisfied, obviously, in a stationary conditions when the energy transmitted by a signal source, continuously fills up the losses, caused by signal absorption indoors. We consider the average energy to be absorbed per the one act is proportional to the general energy store indoors:  $\Delta E = \alpha E$ .

As absorption act occurs  $\nu$  per unit time the average value of the signal absorbed energy per unit time, is

$$\delta E = \nu \Delta E = \alpha \nu E = \alpha \frac{c_0 S}{4V} E \tag{10}$$

The formula (9) defines  $\alpha$  as average relative reduction of signal energy indoors in the one absorption act. But, in addition to that, it is possible to give to coefficient  $\alpha$  other interpretation.

Energy  $\Delta E$  falling at small time  $\Delta t$  on surface  $S$ , limiting volume  $V$  of an indoor, it is possible to define, multiplying an energy stock  $E$  on the probability of its falling calculated on time  $\Delta t$ . According to (5), this probability is

equal  $\frac{c_0 S}{4V} \cdot \Delta t$ . Therefore, let's write  $\Delta' E = \frac{c_0 S}{4V} \cdot \Delta t \cdot E$ .  
Energy  $\delta' E$ , falling on surface  $S$  per unit time, is equal

$$\delta' E = \frac{\Delta' E}{\Delta t} = \frac{c_0 S}{4V} E \tag{11}$$

Comparing this result with (10), we have:

$$\alpha = \frac{\delta E}{\delta' E} \tag{12}$$

If one see, that the indoor is limited by a homogeneous for the physical properties surface (12) defines factor  $\alpha$  as the relation of the energy absorbed by this surface, to energy, diffusely falling on it. The term «diffusive falling» [5] concerns to ergodic process when all directions of carrying over energy signal are equiprobable. It is possible to define, therefore,  $\alpha$  as signal absorption factor (in conditions diffusive fields) such homogeneous surface which, limiting an indoor, causes the same losses of signal energy, as well as actually cash non-uniform surface. So size  $\alpha$  we name average absorption factor, and its product on surface  $S$  we name the general absorption of a premise

$$A = \alpha S \tag{13}$$

If the indoor is limited by surfaces  $S_1, S_2, \dots$  with absorption factors  $\alpha_1, \alpha_2, \dots$  (measured in conditions

diffusive an electromagnetic field) the general absorption is defined by the sum of absorption separate surface sites:

$$A = \sum_i \alpha_i S_i \tag{14}$$

Considering that average radiowave absorption factor value

$$\alpha = \frac{A}{S} = \frac{1}{S} \sum_i \alpha_i S_i, \tag{15}$$

where  $\alpha_i$ - absorption factor of a homogeneous surfaces  $S_i$ , limiting an indoor.

The reasons developed above, give the chance of the a radio-waves distribution phenomena mode establishment statistical calculation in the closed indoor. Let at the moment of time  $t=0$  in an indoor the signal source starts to do with capacity  $P$  per unit time  $\tau$ , equal to average radiowave mean free path, the signal source gives energy  $P\tau$  indoor. As a result of the first reflection, that is by the moment  $2\tau$ , from this energy there will be a part  $P\tau(1-\alpha)$  according to 9. But a signal source again gives to an indoor energy  $P\tau$  per an time interval from  $\tau$  up to  $2\tau$ . Thus, by the moment  $t=2\tau$  indoor there will be an energy  $P\tau + P\tau(1-\alpha) = P\tau[1 + (1-\alpha)]$

Developing the idea further, we consider, that energy will be equal  $P\tau + P\tau[1 + (1-\alpha)](1-\alpha) = P\tau[1 + (1-\alpha) + (1-\alpha)^2]$  per the time interval  $t=3\tau$  indoors.

Act according to an induction, we will receive by the moment  $t=n\tau$ :

$$E = P\tau[1 + (1-\alpha) + (1-\alpha)^2 + \dots + (1-\alpha)^{n-1}] = P\tau \frac{1 - (1-\alpha)^n}{1 - (1-\alpha)} = \frac{P\tau}{\alpha} [1 - (1-\alpha)^n] \tag{16}$$

Supposing  $n=t/\tau$  and substituting value  $\tau$  from (6), we will receive:

$$E = \frac{4P}{c_0 \alpha S} \left[ 1 - (1-\alpha)^{\frac{c_0 S}{4V} t} \right] \tag{17}$$

Using identity  $(1-\alpha)^x = e^{x \ln(1-\alpha)}$ , we will receive the formula of energy increase indoor after signal inclusion

$$E(t) = \frac{4P}{c_0 \alpha S} V \left[ 1 - e^{\frac{c_0 S \ln(1-\alpha)}{4V} t} \right] \tag{18\alpha}$$

For signal energy average density increase we receive

$$\varpi(t) = \frac{E(t)}{V} = \frac{4P}{c_0\alpha S} \left[ 1 - e^{-\frac{c_0 S \ln(1-\alpha)}{4V} t} \right] \quad (18\beta)$$

As  $\alpha < 1$ , that follows from definition of this size both as by the formula (9) and the formula (12) –  $\ln(1-\alpha)$ . So, an exponent is negative.

$$\lim_{t \rightarrow \infty} \left[ e^{-\frac{c_0 S \ln(1-\alpha)}{4V} t} \right] = 0$$

Thus,

Therefore, for a stationary mode when energy losses due to absorption completely compensate the energy transmitted by a signal source, from (18) formulas follow

$$E_0 = \frac{4P}{c_0\alpha S} V \quad \varpi_0 = \frac{4P}{c_0\alpha S} \quad (19\alpha)$$

Now let's proceed to signal reflection process analysis after turning off signal's source. Let per the time interval  $t=0$  when energy containing indoors has some size  $E_0$ , the signal source is switched off. Through time  $\tau$  indoors there will be energy  $E_0(1-\alpha)$ , by the time interval  $2\tau$  the energy stock will decrease to size  $E_0(1-\alpha)^2$  etc.

By the time interval  $t=n\tau$  energy will be equal  $E(t) = E_0(1-\alpha)^n = E_0(1-\alpha)^{\frac{t}{\tau}}$ .

Replaced indicative function exponential and substituting average time value of free path  $\tau$  according to (6), we will find reflection formulas indoors after source turning off:

$$E(t) = E_0 e^{-\frac{c_0 S \ln(1-\alpha)}{4V} t} \quad (20\alpha) \quad \varpi(t) = \varpi_0 e^{-\frac{c_0 S \ln(1-\alpha)}{4V} t} \quad (20\beta)$$

Diagram signal change processes indoors, constructed on the equations (20), and are presented on Fig.4. Level of falling off process value to be presented in the form of a descending straight line, it is clear from picture. The this straight line dip to an abscissas axis defines level of falling off signal's speed in dB/sec. The formula for falling rate is easy receiving, noticed, that on the equations (20).

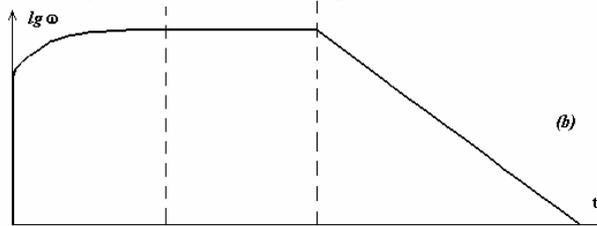


Figure 4. Process of change of a signal indoors at inclusion and source deenergizing

$$N = 10 \lg \frac{\varpi}{\varpi_0} = -\frac{c_0 S}{4V} \ln(1-\alpha) \cdot t (10 \lg e) = \left( \frac{10 c_0 S}{4V} \lg \frac{1}{1-\alpha} \right) t$$

From here it is clear, that level of falling rate is equal after turning off a signal's source

$$\beta = \frac{N}{t} = \frac{10 c_0 S}{4V} \lg \frac{1}{1-\alpha} \quad (21)$$

Let's find time necessary for reduction a signal level to noise level

$$T = \frac{N_{s/N}}{\beta} = \frac{N_{s/N} \cdot 4V}{10 c_0 S \lg \frac{1}{1-\alpha}} = \frac{0,4 N_{s/N} V}{c_0 S \lg \frac{1}{1-\alpha}} \quad (22)$$

In indoor with small average absorption factor absorption ( $\alpha < 0.2$ ) it is possible to time T by the simplified formula, which turns out as follows.

Expanding logarithm into a series, we write

$$\lg \frac{1}{1-\alpha} = -\lg(1-\alpha) = -0,434 \ln(1-\alpha) = 0,434 \left( \alpha + \frac{\alpha^2}{2} + \frac{\alpha^3}{3} + \dots \right)$$

At small  $\alpha$  it is possible to be limited to the first member of the series, which substitution in (22) leads to result:

$$T = \frac{N_{s/N} \cdot 4V}{4,34 c_0 S \alpha} \quad (23)$$

The average coefficient of radio wave indoor absorption is determined by the following formula:

$$\alpha = \frac{1}{S} \sum_i \alpha_i S_i, \quad (24)$$

where  $\alpha_i$  – absorption coefficient of one of the conspecific indoor surface  $S_i$ .

The problem is that up to date the data on absorption coefficient of radio waves of range 2.4 and 5 for materials used in modern buildings are quite limited.

To determine radio wave absorption coefficient we suggest applying the method described in [6], which refers to radio wave of range 2.4 GHz reflection coefficient for three-layered environment in this work.

### III. THE METHOD OF MEASURING DIELECTRIC CAPACITIVITY

To explore the dielectric properties of the materials the waveguide method of short-circuit contact and free running [6] is the optimal one. The method is based on defining the standing wave ratio (SWR) and the phase of a microwave

signal passing through a sample. First the microwave path is calibrated by choosing the reference plane while measuring wave phase. To be able to do that we should define the standing wave short-circuited line minimum position. After that the shorted waveguide section without sample is connected to the line end. The shorted plane shifts a bit, as the section length isn't equal to the integer number of radiation half-waves in the waveguide.

After calibration the waveguide section with a sample is set on the microwave path which results in shifting the standing wave minimum as the minimum depends on dielectric in question properties. The standing wave minimum / dielectric properties ratio is calculated on the basis of the respective electrodynamic problem which is reduced to a complex transcendental equation. The usage of waveguide method of short and free running [7] allows to avoid the solution of this equation. This method provides reliable results and consists in SWR measurement and shifting standing wave minimum with regard to the chosen reference plane for a given sample, at the end of which the shorted and free running conditions are created. Skipping the derivation, we only give the final formula to calculate the imaginary and real part of complex dielectric capacitance  $\epsilon'$  and  $\epsilon''$ :

$$\epsilon' = \frac{AC + BD}{A^2 + B^2} \cdot \left\{ 1 - \left( \frac{\lambda_0}{2a} \right)^2 \right\} + \left( \frac{\lambda_0}{2a} \right)^2 \quad (33)$$

$$\epsilon'' = \frac{BC - AD}{A^2 + B^2} \cdot \left\{ 1 - \left( \frac{\lambda_0}{2a} \right)^2 \right\}, \quad (34)$$

where:

$$A = 1 - S_1 S_2 \operatorname{tg}(\beta \Delta x_1) \operatorname{tg}(\beta \Delta x_2),$$

$$B = S_1 \operatorname{tg}(\beta \Delta x_1) + S_2 \operatorname{tg}(\beta \Delta x_2),$$

$$C = S_1 S_2 - \operatorname{tg}(\beta \Delta x_1) \operatorname{tg}(\beta \Delta x_2),$$

$$D = S_1 \operatorname{tg}(\beta \Delta x_2) \cdot S_2 \operatorname{tg}(\beta \Delta x_1),$$

$S_1, S_2$  – SWR under short and free running conditions respectively;

$$\beta = \frac{2\pi}{\lambda_w} - \text{propagation constant in air-filled waveguide}$$

( $\lambda_w$  — waveguide wavelength);

$\Delta x_1, \Delta x_2$  – standing wave minimum shift with regard to the chosen reference plane at short and free running respectively; in case of “half-infinite” layer  $\Delta x_1 = \Delta x_2$  and  $S_1 = S_2$ ;

$\lambda_0$  – wavelength in free space;

$2a = \lambda_{cr}$  – critical wave length, where  $a$  – the waveguide width (the bigger cross-section side).

The block-diagram of proposed measuring stand is shown on Figure 6. The stand includes signal generator 1 (G4-79 (1,78-2,56 GHz) or G4-81 (4-5,6 GHz)), ferrite isolator 2, providing isolation  $\sim 20$  dB, attenuator 3, waveguide measuring line 4, selective voltmeter 5 and waveguide section with the sample under analysis 6, at the end of which conditions of short and free running can be created with the help of piston 7 in turn. The voltage of SWR of such short is not worse than 30.

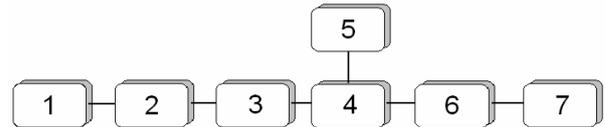


Figure 5. Block-diagram of the measuring stand for complex capacitance research

Absolute measurement error  $\epsilon'$  is not more than 0.5.

The above-given measurements allow to obtain  $\epsilon'$  and  $\epsilon''$  value for different materials used indoors. In turn this allows to define the average absorption coefficient  $\alpha$  value for such materials applying formulas (28-33). The resultant value allows to calculate time T (time Interflexion) for the indoor processes.

#### IV. APPLICATION OF THE STATISTICAL THEORY AT CONSTRUCTION OF WIRELESS NETWORKS

The network using technology MIMO is required to optimize the magnitude T. Time T is calculated for given indoor by (22). Having received theoretical calculation results the wireless network practical research in indoor is carried out. We used the wireless network Radio Ethernet, making with standard equipment IEEE 802.11n usage. Router Linksys WRT610N, Netgear WNDR3700 and TRENDnet TEW-671BR are used as POP.

The research was carried out with method [8] and rate was measured by IxChariot (<http://www.ixiacom.com>).

TCP-traffic (with max size package mainly) is generated by the program and different situation as receiving, transmission and both synchronous (direction to adapter in PC) is modeling.

POP (Depending from model no all point was available) is set to operate with 802.11n range on channel 1(5) in regime «40 MHz», previous generation network security regime was switched off, ciphering WPA2-PSK with AES algorithm was switched on. Others setting were standard.

That network works sufficiently stable should take into account, as data transmission rate negligibly changed during all test. After first examination cycle we did indoor treatment, placed supplementary elements with high reflection property. These elements placed behind POP and along radio wave diffusion trace. To avoid the standing wave reflection element planes placed by different angle to each other. Time T increased from 210 ns to 390 ns according to calculation results. Increase time T more 400

ns isn't recommended, as "guard time" can be overdraw and require result isn't archived.

Test results are shown on Figures 6, 7, 8, 9.

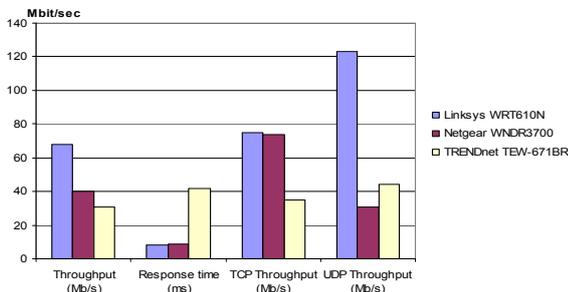


Figure 6. 2.4 GGz before processing

In the range 2.4GGz we received max rate in transmission regime UDP (about 123 Mbit/s) for POP Linksys WRT610N. Throughput is a twice less – on the order of 68 Mbit/s.

The second indicator in the POP Netgear WNDR3700. Worst performance in terms of in the POP TRENDnet TEW-671BR.

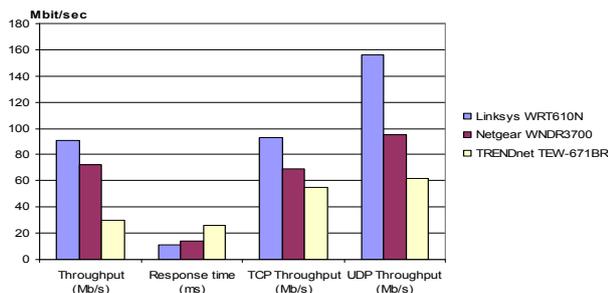


Figure 7. 5 GGz before processing

In the range 5GGz we received max rate (about 156 Mbit/s) for POP Linksys WRT610N. Response time of less is better, 11 ms for POP Linksys WRT610N.

After processing premises speed increased by 10-15%.

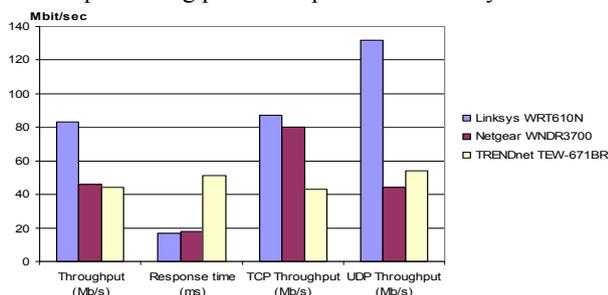


Figure 8. 2.4 GGz after processing

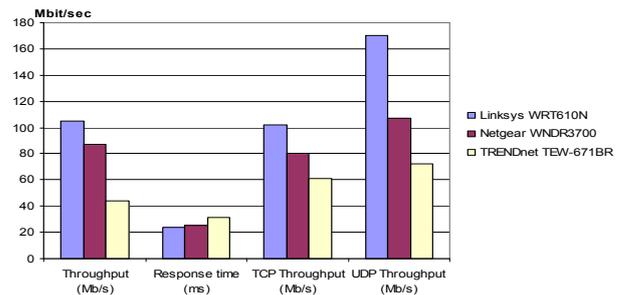


Figure 9. 5 GGz after processing

CONCLUSIONS

In section 2 we developed a statistical theory in result of the mathematical analysis of radio wave propagation in a room. Statistical theory of propagation in a room to determine interflektion time during which the level of interference noise is reduced to the permissible ratio of NSR values. In Section 3 we proposed a method for finding the dielectric constant materials. Using the theory [6] we were able to find the various materials radio waves absorption coefficients used in buildings. Knowing these values by (23,24), we calculated the interflektion time for a particular room. . a special treatment, which increases the value with the value 210 ns to a value of 390 ns, is held to ensure the regime of radio waves multiple reflection from the surfaces of the room. Treatment included the location of the additional elements with high reflective properties. Such elements are located beyond the access point and along the propagation. In order not to set up standing waves, the plane of the reflecting elements have not been parallel, they was at various angles to each other. Increased time interflektion to 400 ns, we do not recommend, since it can lead to problems with the guard interval.

With this program IxChariot we simulated different situations: reception, transfer, simultaneous transmission and reception (direction - with respect to the adapter to a PC) with the packages mainly the maximum size. The comparison was performed on 4 basic parameters: Throughput (Mbs), Response time (ms), TCP Throughput (Mbs), UDP Throughput (Mbs). Speed was estimated by the maximum value, Response time -to the minimum value. The maximum speed of 170 Mbs was reached at a frequency of 5GHz mode UDP Throughput with using a router Linksys WRT610N. In the 5GHz range devices creating noise is less than in the 2.4 GHz range, so a top speed is managed to achieve only in 123 Mbs in this range. In general, the increase in the rate was 10-15% compared with the rate in the room, not being processed.

The experimental results are consistent with the developed theory. The theory may be used to improve the performance of a wireless network for 802.11n in the room.

Further, our work will focus on establishing a program to perform calculations according to advanced theories and experimental works on the transfer of video and audio wireless client.

The statistical theory of radio-waves distribution in indoor allows to define time interfection T achievements of demanded parity SNR. The Indoor after-treatment processing is carried out by selection in large area extent indoors with small factor of absorption for demanded technology MIMO interference noise level achievement. These surfaces are placed to increase T in wireless clients place.

Mentioned above method usage allows to increase data transmission rate in the indoor at wireless network design.

#### REFERENCES

- [1] Why Wi-Fi sucked at Mobile World Congress. [http://money.cnn.com/2011/02/18/technology/mwc\\_wifi/http://money.cnn.com/2011/02/18/technology/mwc\\_wifi/](http://money.cnn.com/2011/02/18/technology/mwc_wifi/http://money.cnn.com/2011/02/18/technology/mwc_wifi/). Last access date: 30.03.2011.
- [2] C. Zhuo, and H. Suzuki. Performance of 802.11n WLAN with transmit antenna selection in measured indoor channels. Proc. Communications Theory Workshop (AusCTW 2008), IEEE Press 2008, pp.139-143, doi: 10.1109/AUSCTW.2008.4460836.
- [3] Yu-Tzu Huang, Jia-Shi Lin, and Kai-Ten Feng. Performance analysis for aggregated selective repeat ARQ scheme in IEEE 802.11n networks. Proc. Personal, Indoor and Mobile Radio Communications, IEEE 20th International Symposium, IEEE Press, 2009, pp.37-41, doi: 10.1109/PIMRC.2009.5449921.
- [4] M. Schroeder, Statistical parameters of the frequency response curves of large indoors. Publisher: JAES, 1987, vol.34 – pp.299-305.
- [5] L. Brehovskih. Waves in sandwich mediums (M.: Pub. AS USSR, 1956G).
- [6] A. Lagunov. Research Interference Handicapes to Radio Networks in a Range of 2.4 GHz. Proc. Third International Conference on Wireless and Mobile Communications (ICWMC'07), IEEE Press, March 2007, P.32, doi: 10.1109/ICWMC.2007.74.
- [7] A. Brandt. Dielectric study at microwave frequencies (M: Publishing House of Physical-Mathematical Literature, 1963).
- [8] E. Zajtsev. The Technique of testing of routers. <http://www.ixbt.com/comm/route>. Last access date: 30.03.2011.

# Thresholds Determination for new Backoff Algorithm in MANETs

Muneer Bani Yassein, Wail Mardini, Zainab AbuTaye'

Jordan University of Science and Technology  
masadeh@just.edu.jo mardini@just.edu.jo, zoabutaye08@cit.just.edu.jo

**Abstract** - Backoff algorithms are one class of collision resolution algorithms used in the medium access control protocols in mobile ad hoc networks. When there are different nodes competing to access a shared channel at the same time, the possibility of collision is highly probable, especially in high traffic load networks. This study aims to study the possibility of using different values for increments and decrements in the contention windows in order to improve the performance.

**Keywords** - Medium Access Control protocols; Backoff Algorithm; Ad hoc network; network performance.

## I. INTRODUCTION

Mobile Ad hoc Networks (MANETs) [1, 2] work without requiring any preexisting communication infrastructure. These types of networks gain a high importance and attract attention due to the need of rapid deployment in emergency cases such as military operations, search and rescue operations and disaster recovery that do not have enough time to build an infrastructure. A MANET is an autonomous system of wireless nodes connected by wireless links. Each node not only acts as a sender or receiver but also as a router in order to convey the packet via intermediate nodes until it reaches the desired destination (multi-hop technique). These nodes have mobility characteristic that allow forming a dynamic network topology which is highly changeable and random.

Collision is considered as the major problem in wireless networks, so the backoff mechanism should be applied in order to decrease the collision and to achieve an efficient use of the shared channel. IEEE 802.11 DCF, the most widely used standard in MAC protocol, uses a Binary Exponential Backoff (BEB) algorithm which uniformly chooses the backoff value from the Contention Window (CW).

The main idea of using the backoff algorithm is to reduce collision and resolve contention among different nodes. Once the collision has occurred, the collided nodes are needed to defer for a period time known as retransmission delay (or backoff) which is usually selected randomly from bounded contention window that has a predetermined lower and upper values. These values are based on the number of active nodes and traffic load in the network. As an example CW<sub>min</sub> and CW<sub>max</sub> are usually set

to 31 and 1023 respectively in IEEE 802.11 DCF, and set to 2 and 1024 respectively in Ethernet [3, 4, 5, 6].

Researchers have proposed many backoff algorithms in order to utilize the performance for IEEE 802.11 DCF MAC protocol. Some of these algorithms will be mentioned in the following section.

## II. BACKGROUND AND RELATED WORK

Many backoff algorithms have been proposed in order to decrease the collision and to achieve an efficient use of the shared channel.

In [7], Bharghavan, et al. have proposed Medium Access with Collision Avoidance-Wireless (MACAW) protocol which used a Multiplicative Increase Linear Decrease (MILD) Backoff algorithm. In the MILD algorithm, the nodes increase their contention window multiplicatively upon collision or failure in transmission and decrease their contention window linearly upon success. This algorithm introduced to address unfairness problem in Binary Exponential Backoff (BEB) algorithm.

Manaseer and Masadeh [8] have proposed a Pessimistic Linear Exponential Backoff (PLEB) algorithm. This algorithm based on assumption that the failure in transmission process is caused by the congestion in network. It is considered as a result of combination between linear and exponential increment methods. Using these two increment methods will help to achieve the aim of this algorithm in improving the performance of a MANET in terms of network throughput and average packet delay. By using the linear increment, this algorithm will improve the performance by reducing network delay. On the other hand, using the exponential increment will improve network throughput.

Manaseer, et al. [9,10] have proposed Fibonacci Increment Backoff (FIB) algorithm and Logarithmic (LOG) Backoff algorithm, respectively. The former algorithm uses a famous math series called Fibonacci Series, which aims to reduce the differences between successive contention window sizes, this algorithm achieves a higher throughput when compared with (BEB) algorithm, the later algorithm uses Logarithmic increments in order to utilize the distribution of random numbers. It achieves a higher throughput and less packet loss. It also achieves stability of network throughput over various speeds of nodes.

In [4], Haas and Deng have proposed the Sensing Backoff Algorithm (SBA) in order to utilize the network

throughput and fairness issues. This algorithm based on sensing mechanism (overhearing the channel to get the needed information). So, each node changes its backoff interval based on the results of the sensed channel status.

Deng, et al. [11] have proposed linear Multiplicative Increase Linear Decrease (LMILD) Backoff algorithm; in this algorithm, the collided nodes multiplicatively increase their contention windows, while other nodes overhearing the collision increase their contention window in linear way. Upon a success, all nodes decrease their contention windows in a linear way.

In [12], Exponential Increase Exponential Decrease (EIED) backoff algorithm was proposed to improve the performance of the IEEE 802.11 DCF. Upon a collision or Failure, nodes exponentially increase their contention window and upon a success all nodes exponentially decrease their contention windows. This algorithm surpasses BEB in terms of throughput and delay.

Choi, et al. [13] have proposed Predictive DCF (P-DCF) Backoff algorithm to be used in IEEE 802.11 DCF. This algorithm enables nodes to choose their next backoff times by listening to the channel continuously. It reduces the collision probability and outperforms the BEB algorithm in terms of throughput and delay.

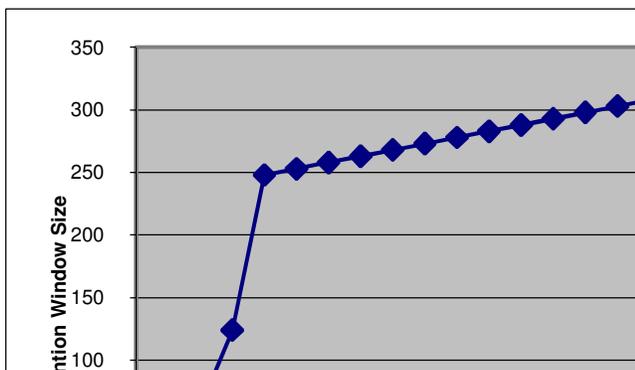


Figure 1. Exponential and linear increase of CW in PLEB.

Figure 1 shows the increment behavior of PLEB. In this paper, we study the impact of using different values of this exponential and linear increment. We also study the impact of having repeating this process in multiple phases.

When the backoff mechanism is applied, the contention window size needs to be increased as a response to a failure or collision. As the first response to a failure, we study increasing the CW size in an exponential way. Increasing by using exponential will improve network throughput by producing enough length of backoff times, so the new Contention Window (CW) is calculated using the following formula:

$$CW_{new} = CW * K \quad (1)$$

where K is an exponential increase factor.

Then, after a number of exponential increments, we study the impact of linearly increases the CW in order to avoid the rapid growth of CW size that causes a high increase of the backoff value and thus results in reducing network delay, so the new contention window is changed using the following formula:

$$CW_{new} = CW + T \quad (2)$$

where T is a linear increase factor.

Upon a success in transmission, we want to decrease the CW in a linear way instead of resetting the CW to its minimum value as in BEB. We use a linear decrease mechanism to solve the fairness problem (avoid channel domination of only one node) due to that resetting mechanism in BEB takes only one successful transmission to reach  $CW_{min}$  which causes a huge variation of the contention window size and degrades the performance in heavy loaded network since each new packet starts with the minimum contention window value which is considered as a small value in heavy loaded network, so in our algorithm the new contention window is changed using the following formula:

$$CW_{new} = CW - Y \quad (3)$$

where Y is a linear decrease factor.

As justified later in the following section, based on experimental results, we choose the value of Y to be 2.

### III. SIMULATION RESULTS AND ANALYSIS

We use Glomosim (version 2.03) simulator [14, 15] to study the impact of different values in order to choose a suitable values and compare them with Pessimistic Linear/Exponential Backoff (PLEB) Algorithm.

The network consists of 100 nodes randomly placed in a 2200m × 600m rectangular field. We use the Random Waypoint as the mobility model. Constant bit rate (CBR) with 512 byte data packets is used. The MAC layer protocol is 802.11. The main parameters used in the simulations are summarized in Table 1.

TABLE 1: SIMULATION PARAMETERS

Parameter	Value
Network area	2200m × 600 m
Transmission	250 m
Number of nodes	100 nodes
Nodes speed	1, 4, 10 m/s
Pause time	0 s
Bandwidth	2 MHz
Traffic type	CBR
Packet rate	4
Packet size	512 B
Simulation time	500 s

The metrics used to compare the performance are the following:

1. **Throughput:** the total amount of data packets successfully received at the destination.
2. **Packet Delivery Ratio (PDR):** the ratio of total amount of data packets received by the destination to the total amount of data packets sent by the CBR sources.
3. **Average End-to-End delay (EED):** is the average delay taken for a data packet that is originated at the source to be received at the destination. All possible delays caused by buffering during route discovery latency, retransmission delays at the MAC, queuing at the interface queue and propagation delay are included.

Many experiments are conducted to study the effect of our proposed algorithm in terms of throughput, packet delivery ratio and end to end delay in order to justify our selected values. Figures 2-4 show the impact of number of increments of linear and exponential on throughput, packet delivery ratio and end to end delay respectively with keeping the resetting mechanism as in BEB and PLEB.

Experiment 1 to experiment 4 display 3 increments of exponential in the two phases within increase factors  $k$  and  $u$  equal 1.5 and varying the number of linear increments in phase 1 to 5, 10, 15, 20 respectively within increase factors  $t$  and  $v$  equal 5.

Experiment 5 to experiment 8 display 2 increments of exponential in the two phases within increase factors  $k$  and  $u$  equal 1.5 and varying the number of linear increments in phase 1 to 5, 10, 15, 20 respectively within increase factors  $t$  and  $v$  equal 5.

Results show that experiment 1 has a better performance in terms of throughput and packet delivery ratio as compared to PLEB with no improvement in delay metric.

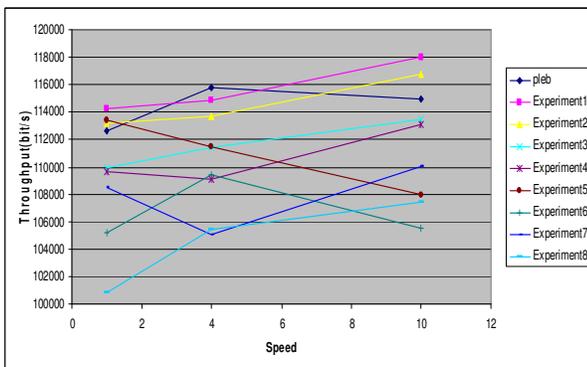


Figure 2. Impact of number of increments of linear and exponential on throughput for  $K= 1.5, T= 5, U= 1.5, V= 5$ .

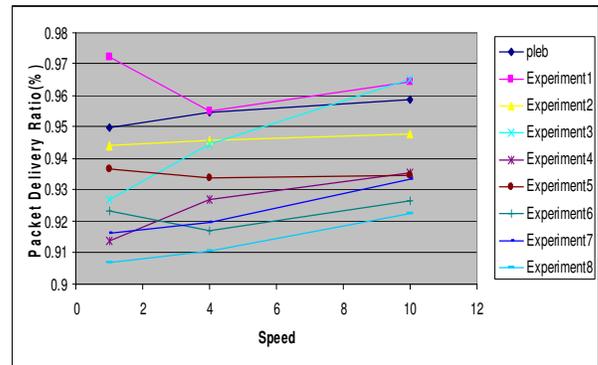


Figure 3. Impact of number of increments of linear and exponential on PDF for  $K= 1.5, T= 5, U= 1.5, V= 5$ .

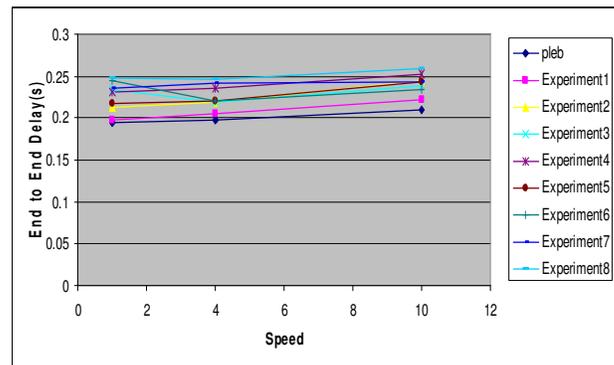


Figure 4 Impact of number of increments of linear and exponential on delay for  $K= 1.5, T= 5, U= 1.5, V= 5$ .

Depending on the results of previous figures, another type of experiments were held related to experiment 1. Figure 5, Figure 6 and Figure 7 show the impact of number of increments of linear and exponential on throughput, packet delivery ratio and end to end delay respectively.

Experiment 1 to experiment 3 display 3 increments of exponential with  $k=u=1.5$  but varying the decrease process of CW upon successful transmission as exponentially decrease with  $y=1.5$ , resetting to  $CW_{min}$  and linearly decrease with  $y= 2$  respectively.

Experiment 4 to experiment 6 display 3 increments of exponential at phase 1 and 2 increments at phase 2 with  $k=u= 1.8$  but varying the decrease process of CW upon successful transmission as exponentially decrease with  $y=1.8$ , resetting to  $CW_{min}$  and linearly decrease with  $y= 2$  respectively.

Experiment 7 to experiment 9 display 3 increments of exponential at phase 1 and 1 increment at phase 2 with  $k=u= 2$  but varying the decrease process of CW upon successful transmission as exponentially decrease with  $y=2$ , resetting to  $CW_{min}$  and linearly decrease with  $y= 2$  respectively.

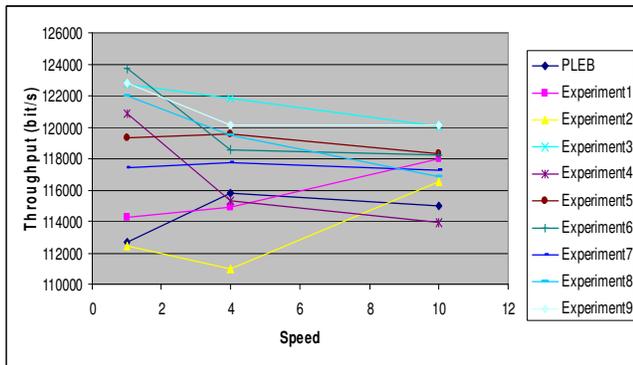


Figure 5. Impact of varying exponential increase factor and resetting mechanism of linear on throughput.

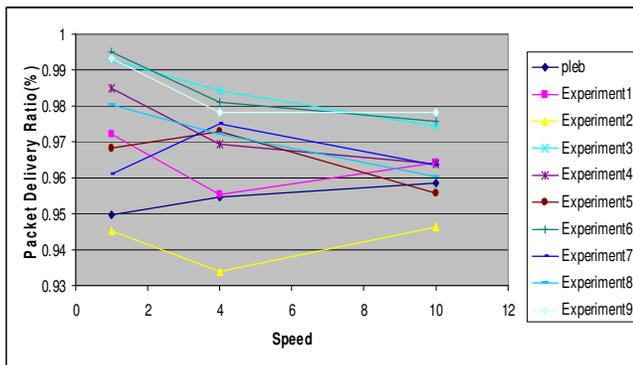


Figure 6. Impact of varying exponential increase factor and resetting mechanism of linear on PDF.

Results show that experiment 3 has a better performance in terms of throughput and packet delivery ratio as compared to PLEB and the delay is approximately the same.

Based on the results of the above figures, in this study our suitable thresholds are  $K= 1.5, T= 5, U= 1.5, V= 5, Y= 2$  as they offer better performance as compared to other possible values, based on these values  $W$  equal 69.75.

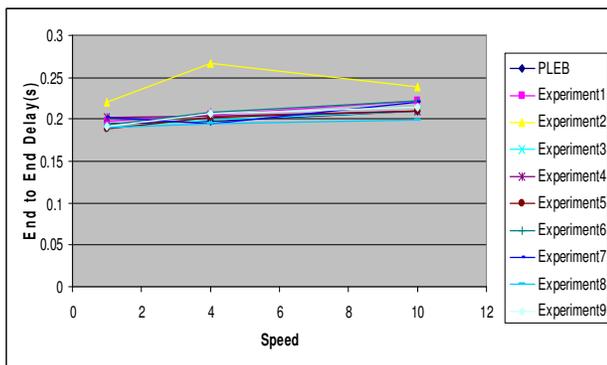


Figure 7. Impact of varying exponential increase factor and resetting mechanism of linear on delay.

#### IV. CONCLUSION

This research presented an experimental study of using different values in a backoff algorithm with different criteria (e.g., number of increments) to select suitable thresholds to be used in new algorithms and compared them with PLEB. The results show that when thresholds are  $K= 1.5, T= 5, U= 1.5, V= 5, Y= 2$  this will offer better performance as compared to other possible values in terms of throughput, packet delivery ratio and delay which is approximately the same compared to PLEB.

#### REFERENCES

- [1] Nieminen, K., "Introduction to Ad Hoc Networking", Networking Laboratory Helsinki University of Technology, pp. 1-10, 2003.
- [2] Chlamtac, I., Conti, M. and Liu, J., "Mobile ad hoc networking: imperatives and challenges", Science Direct, pp 13-64, 2003.
- [3] Wu, H. and Pan, Y., "Medium Access Control in Wireless Networks (Wireless Networks and Mobile Computing)", Nova Science Publishers, 2008.
- [4] Deng, J. and Hass, Z.J., "On Optimizing the Backoff Interval for Random Access Schemes" IEEE Transactions on Communications, volume 51, issue 12, pp. 2081- 2090, 2003.
- [5] Krishna, C., Chakrabarti, S. and Datta, D., "Impact of Contention Resolution Algorithms on the Performance of IEEE 802.11 DCF-based MAC Protocol in a Mobile Ad hoc Network", pp. 1-4, 2005.
- [6] Krishna, C., Chakrabarti, S., Datta, D., "A modified backoff algorithm for IEEE 802.11 DCF-Based MAC protocol in a mobile ad hoc network", IEEE, pp. 664- 667, 2004.
- [7] Bharghavan, V., Demers, A., Shenker, S. and Zhang, L., "MACAW: A media access protocol for wireless LAN's", Proceedings of the ACM SIGCOMM, London, pp. 212-225, 1994.
- [8] Manaseer, S. and Masadeh, M., "Pessimistic Backoff for Mobile Ad hoc Networks ", IEEE, pp 1-10, 2008.
- [9] Manaseer, S., Ould-Khaoua, M. and Mackenzie, L., "Fibonacci Backoff Algorithm for Mobile Ad Hoc Networks ", University of Glasgow, pp. 1-6, 2006.
- [10] Manaseer, S., Ould-Khaoua, M. and Mackenzie, L., "On a Modified Backoff Algorithm for MAC Protocol in MANETs", University of Glasgow, pp. 1-13, 2007.
- [11] Deng, J., Varshney, P., and Haas, Z., "A new backoff algorithm for the IEEE 802.11 distributed coordination function" , In Communication Networks and Distributed Systems Modeling and Simulation (CNDS '04), pp. 1-6, 2004.
- [12] Song, N.O., Kwak, B.J., Song, J., and Miller, L., "Enhancement of IEEE 802.11 distributed coordination function with exponential increase exponential decrease backoff algorithm", in Proc. VTC 2003-Spring, vol. 4, Jeju, Korea, pp. 2775-2778, 2003.
- [13] Choi, N., Seok, Y., Choi, Y., Kim, S. and Jung, H., "P-DCF: Enhanced Backoff Scheme for the IEEE 802.11 DCF", pp. 1-4, 2005.
- [14] Zeng, X., Bagrodia, R. and Gerla, M., "Glomosim: A library for parallel simulation of large-scale wireless networks". In Proceedings of the 12th Workshop on Parallel and Distributed Simulation, 26.29, pp. 154-161.
- [15] GloMoSim: Global Mobile Information Systems Simulation Library. <http://pcl.cs.ucla.edu/projects/gloimosim/>.

# An Interference Avoiding Wireless Network Architecture for Coexistence of CDMA 2000 1x EVDO and LTE Systems

Xinsheng Zhao

National Mobile Communications Research Laboratory,  
Southeast University  
Nanjing, Jiangsu, 210096, P.R. China  
xszhao@seu.edu.cn

Jia You, Tao Jin

National Mobile Communications Research Laboratory,  
Southeast University  
Nanjing, Jiangsu, 210096, P.R. China  
{jiayou, jintao}@seu.edu.cn

**Abstract**—Interference between two RAT (Radio Access Technology) wireless networks will lead to the system capacity deduction. In this paper, system capacity loss was analyzed by using snapshot method and the simulation results show that system capacity loss is closely related to the relative offset and antenna direction between two RAT wireless networks. Based on the interference analysis above, a novel wireless architecture was proposed by rotating antenna direction and adjusting offset between two RAT wireless networks. Simulation results verified that it can effectively avoid system interference.

**Keywords** – Interference; LTE, CDMA2000 1x EVDO; Wireless Networks

## I. INTRODUCTION

Long Term Evolution (LTE) is designed to reduce latency, improve capacity and coverage, and enhance Quality of user Experience (QoE) [1]. However, there will be the scenario of coexistence of LTE network and other RAT (Radio Access Technology) wireless networks due to the huge cost of network building and scarce frequency resource [2]. Previous works have focused on analysis of system interference in two wireless networks coexistence, the problem of wireless network layout influencing system performance still remains open [3-7]. So, it is essential to analyze interference between two wireless networks and explore wireless network architecture to reduce interference.

There are two kinds of interference analysis methods, analytical method and simulation method. Analytical method requires an accurate mathematical model abstracted from system level behavior of wireless network. Interference could be analyzed by mathematical derivation. The simulation method is a snapshot method developed from Monte Carlo method [9]. By establishing the system simulation platform, randomizing user’s parameters such as position, speed, wireless channel state and running program periodically, statistics could be obtained in a cyclic snapshot way for interference analysis.

As it is difficult to get an accurate mathematical model for CDMA 2000 1xEVDO and LTE coexistence scenario, snapshot simulation method could be considered as a proper way for interference analysis. One simulation procedure including several snapshots of LTE system is described as in Figure 1.

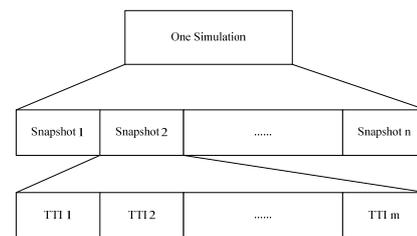


Figure 1. Basic process of snapshot method.

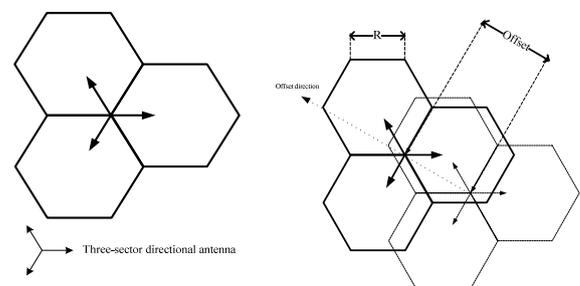
One simulation process is composed of n independent snapshots and each snapshot includes m continues TTIs (Transmission Time Interval). Simulation results of each snapshot are stored and the system performance can be obtained by averaging the stored results.

The rest of the paper is organized as follows. In Section II, interferences in two wireless networks coexistence scenario are briefly described. In Section III, simulation platform is introduced and simulation results were shown. A new wireless network architecture is proposed and verified in Section IV. Section V concludes the work.

## II. INTERFERENCE IN COEXISTENCE OF WIRELESS NETWORKS

This section describes interference types in coexistence of wireless networks of CDMA 2000 1x EVDO and LTE systems.

### A. Coexistence of wireless networks



(a) Single wireless network (b) Two RAT wireless networks  
Figure 2. Wireless network architecture.

Figure 2(a) shows a one cell model for the wireless network architecture of three sector antenna. A cell is composed of three sectors represented as regular hexagon.

Figure 2(b) shows the coexistence of wireless networks architecture. As it can be seen, the LTE network is depicted in dashed line and CDMA 2000 1x EVDO is in solid line.

The network architecture of two systems is the same such as cell radius, antenna direction. But, there is a position offset between LTE and EVDO wireless network. The offset direction is depicted by a dotted line. The maximum of offset is  $\sqrt{3} \cdot R$ , where  $R$  is cell radius as shown in Figure 2(b). After normalization of offset, offset's range is from 0 to 1, where 1 presents the maximum offset  $\sqrt{3} \cdot R$  and 0 indicates no offset between EVDO and LTE network architecture.

### B. Interference types

Interference includes inter-cell interference and adjacent channel interference. Inter-cell interference is caused by inter-cell users which share the same frequency with victim user. Adjacent channel interference is mainly composed of adjacent channel leakage and adjacent channel selectivity [10]. Scenario of CDMA2000 1x EVDO AT interfering LTE eNB can be described as Figure 3 and interference to downlink is the same.

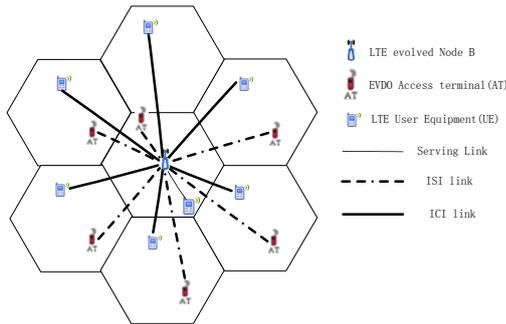


Figure 3. Scenario of EVDO AT interfering LTE UE.

LTE eNB in the center presents the serving station which has established uplink wireless connection to serving UE by dotted line. The serving UE is transmitting uplink data to this serving eNB which is interfered by inter-cell interference and inter-system interference. In addition, black solid line denotes interference from inter cell LTE User equipments (UE) and dashed line represents adjacent channel interference from EVDO system. The SINR (signal interference noise ration) of signal received by serving eNB can be calculated as following.

$$SINR = \frac{P_s}{P_{ISI} + P_{ICI} + N_0} \quad (1)$$

$P_s$ ,  $P_{ICI}$ ,  $P_{ISI}$  present receiving signal power, inter-cell interference power and inter-system signal power respectively.  $N_0$  is noise which is a constant once the working band wide is determined. The calculation methods

for  $P_s$ ,  $P_{ICI}$ ,  $P_{ISI}$  had been described in 3GPP specification [10].

The performance of wireless networks coexistence can be evaluated by system capacity loss which is defined as following.

$$capacity\_loss = \left(1 - \frac{C_{coexistence}}{C_{single}}\right) \cdot 100\% \quad (2)$$

where,  $C_{coexistence}$  denotes sector throughput of LTE (or EVDO) system under coexistence scenario and  $C_{single}$  denotes LTE (or EVDO) sector throughput without inter system interference [10].

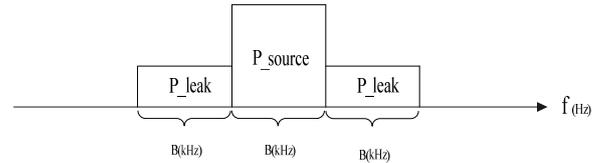


Figure 4. Definition of ACIR.

In order to measure system capacity loss in different inter-system interference level, Adjacent Channel Interference Ratio (ACIR, in dB) is introduced as an independent variable to adjust the adjacent channel interference rate and change interference intensity [10]. In Figure 4,  $P_{source}$  (in dBm) denotes the total power of aggressive system in  $B$  (KHz) bandwidth,  $P_{leak}$  (in dBm) means the leakage power on adjacent  $B$ (KHz) bandwidth channel. Then the ACIR is defined as follow:

$$ACIR = 10 \cdot \log\left(\frac{P_{source}}{P_{leak}}\right) \quad (3)$$

Once ACIR is fixed,  $P_{leak}$  could be calculated as

$$P_{leak} = 10^{-\frac{ACIR}{10}} \cdot P_{source} \quad (4)$$

## III. INTERFERENCE ANALYSIS BY SNAPSHOT

### A. Simulation Platform

The simulation platform is established according to 3GPP and 3GPP2 specifications [10-12]. When calculating inter-system interference, the simulation time period of EVDO sub-frame length (20/3 ms) has to be aligned to LTE system TTI (1ms). Figure 5 represents the process of time alignment. After time alignment, EVDO transmitting power during 1 ms remains unchanged.

Figure 6 shows the flowchart of snapshot process including initial layout of coexistence networks and procedure of each snapshot. A TTI process is composed of modeling the channel fading of LTE and EVDO system, scheduling (or rate controlling), calculating mean SINR of receiving signal considering interference signal power from the other system, and storing the system throughput of current TTI through link level interface.

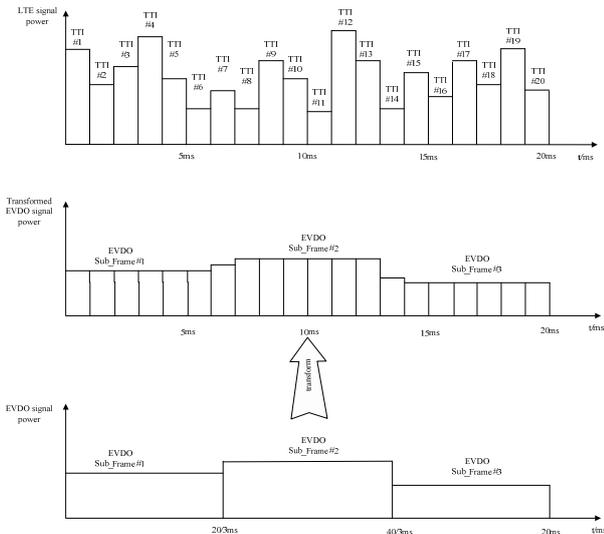


Figure 5. The process of time alignment.

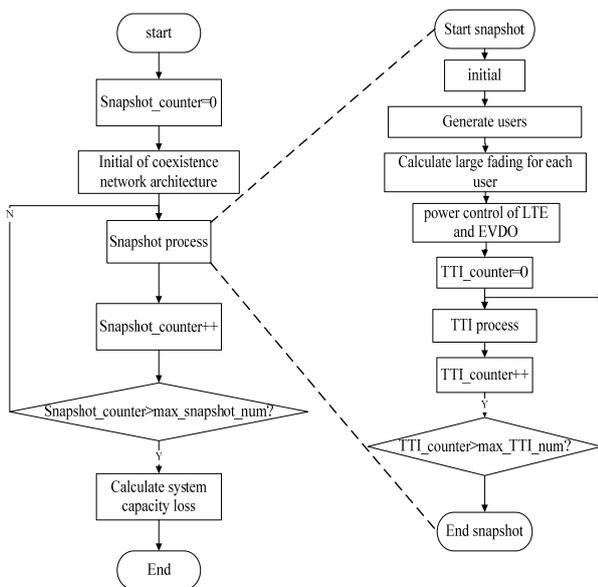


Figure 6. Flowchart of simulation platform.

The system parameters of CDMA2000 1xEVDO and LTE systems in simulation platform are listed in table I and table II according to the related specifications [10-12].

Table I. Parameters of LTE

Parameter	Downlink	Uplink
User Number per sector	16	16
Bandwidth	10MHz	10MHz
Carrier Frequency	2.005GHz	2.005GHz
Power control	N/A	Open Loop power control
Scheduling algorithm	Proportional fair	Proportional fair

Table II. Parameters of CDMA2000 1xEVDO

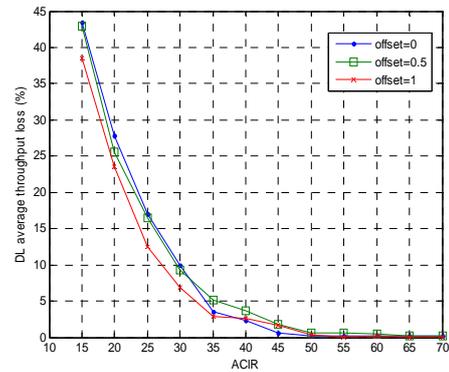
Parameter	Downlink	Uplink
User Number per sector	16	16
Bandwidth	1.25MHz	1.25MHz
Carrier Frequency	1.999375GHz	1.999375GHz
Power control	N/A	Close loop power control
Scheduling algorithm	Proportional fair	N/A
Rate control	N/A	Token bucket

B. Simulation Results

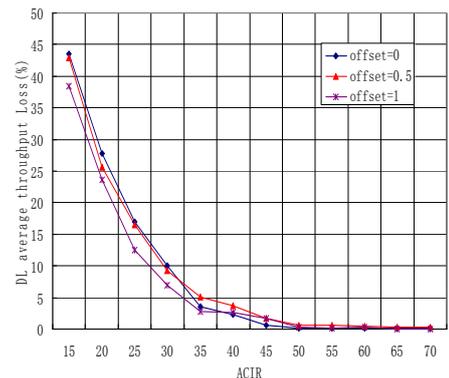
Downlink sector capacity loss and uplink sector capacity loss are taken as the system performance under different inter-interference scenarios.

(i) Downlink

EVDO downlink sector capacity loss interfered by LTE and LTE downlink sector capacity loss interfered by EVDO is illustrated in Figure 7(a) and Figure 7 (b) respectively.



(a) EVDO downlink sector capacity loss



(b) LTE downlink sector capacity loss  
Figure 7. Downlink sector capacity loss.

From each figure, we can find a common phenomenon that downlink capacity loss while offset=0 is more serious than the other two offset situations.

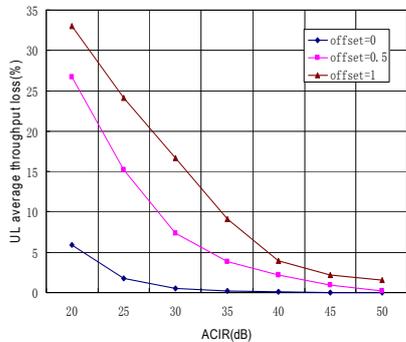
Comparing two figures above, EVDO is much easier to be interfered than LTE. As mentioned on parameter tables

in Section II, LTE is working on a wider band than EVDO and both systems are working on the adjacent channels. The LTE leakage power to the adjacent channel is so wider in frequency domain that it covers EVDO working band totally. Quite different, the power leakage of EVDO in frequency domain could only cover a small part of LTE working band. As a result, EVDO downlink is much more vulnerable than LTE when interfering with each other.

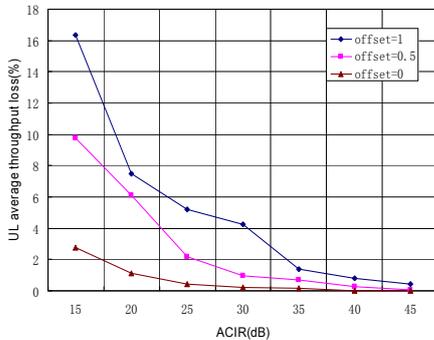
In order to limit capacity loss which should be less than 5% for example, the ACIR of LTE system should be larger than 32.5dB, 31dB, 31.5dB when offset is 0, 0.5, 1 respectively. For EVDO system, the ACIR should be larger than 34dB, 35dB, 32.5dB when offset is 0, 0.5, 1 respectively.

(ii) Uplink

EVDO uplink sector capacity loss interfered by LTE is shown in Figure 8(a) and LTE uplink sector capacity loss interfered by EVDO is illustrated in Figure 8(b).



(a) EVDO uplink sector capacity loss



(b) LTE uplink sector capacity loss  
Figure 8. Uplink sector capacity loss.

The result shows that uplink capacity loss is the most serious when offset=1 among three offsets. After comparing Figure 8(a) with Figure 8(b), we can find that EVDO uplink is easy to be interfered than LTE. The reason is similar to downlink case.

In order to confine capacity loss within 5% for example, the ACIR of LTE should be larger than 15dB, 22dB, 25dB when offset is 0, 0.5, 1 respectively. For EVDO system, the ACIR should be larger than 21dB, 33dB, 39dB when offset is 0, 0.5, 1 respectively.

IV. A NEW WIRELESS NETWORK ARCHITECTURE

In order to reduce the system interference as described above, a new wireless network architecture is proposed, which is different from that presented in 3GPP specification [10]. This wireless architecture is shown in Figure 9.

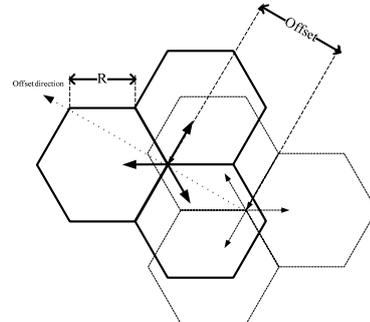
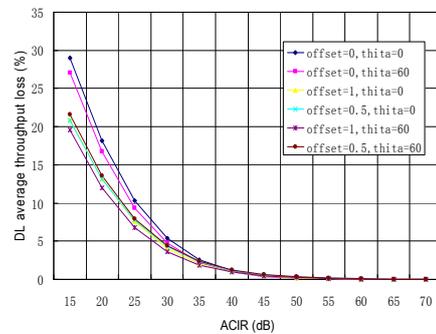


Figure 9. Definition of offset and antenna rotation.

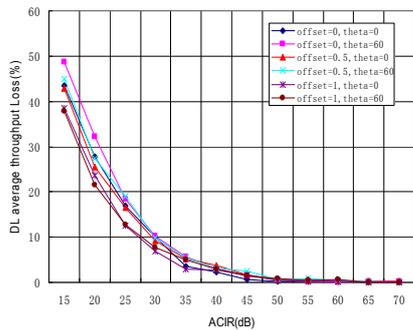
In Figure 9, solid line denotes LTE cell and dashed line means EVDO cell. Comparing to traditional network architecture in Section 2, the antenna of LTE rotates  $\theta = 60$  degrees and other parameters remain unchanged. The offset remains the same definition as defined in Section 2. In order to compare with system performance in traditional architecture, simulation results of capacity loss are shown in the same figure.  $\theta = 0$  represents traditional network architecture, while  $\theta = 60$  represents new coexistence architecture. The result is shown as follows.

A. Downlink

In the new architecture scenario, downlink capacity loss of LTE and EVDO are shown in Figure 10(a) and Figure 10(b). The results show that system capacity loss under new coexistence architecture is reduced comparing to the system performance under traditional architecture when the offset is fixed. To confine capacity loss within 5% under the new architecture, ACIR of LTE should be larger than 30dB, 29dB and 27.5dB when offset is 0, 0.5 and 1 respectively. On the other hand, ACIR of EVDO should meet the minimum requirements of 36.75dB, 35.5dB and 35dB when offset is 0, 0.5 and 1 respectively.



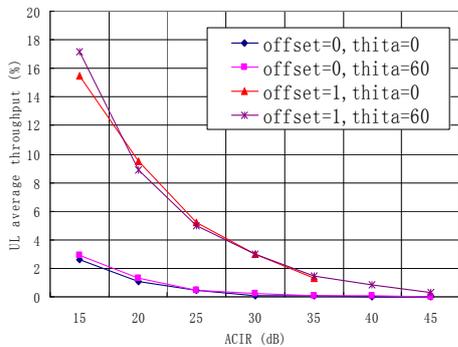
(a) LTE downlink capacity loss



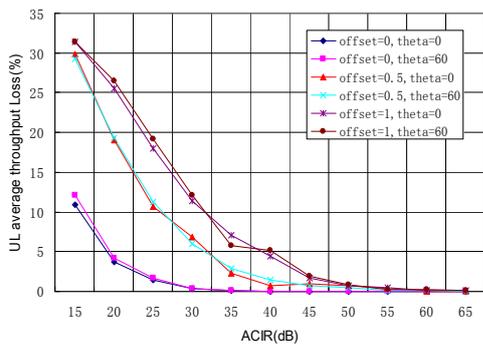
(b) EVDO downlink capacity loss  
Figure 10. Downlink sector capacity loss.

**B. Uplink**

Under the new architecture, uplink capacity loss of LTE and EVDO is shown in Figure 11(a) and Figure 11 (b) respectively.



(a) LTE uplink capacity loss



(b) EVDO uplink capacity loss  
Figure 11. Uplink sector capacity loss.

To confine capacity loss within 5% under the new architecture, ACIR of LTE should be larger than 19.5dB, 31.5dB and 40dB when offset is 0, 0.5 and 1 respectively. On the other hand, ACIR of EVDO should meet the minimum requirements of 15dB, 24.5dB when offset is 0 and 1 respectively. The result shows that under the new

architecture, uplink capacity loss is reduced comparing to traditional wireless network architecture.

**V. CONCLUSION**

In the paper, system performance of LTE and EVDO under different wireless architecture was analyzed by snapshot simulation method. The simulation results show that with increasing offset between two RAT wireless networks, downlink capacity loss is decreased while uplink capacity loss is increased. After rotating antenna direction, interference is relieved in downlink and uplink. Furthermore, comparing to rotating antenna direction of each wireless networks node, adjusting offset between two RAT wireless networks is more effective to reduce inter system interference.

**ACKNOWLEDGMENT**

This work was supported by National Science and Technology Major Project (No.2009ZX03002-013-02) and the National Natural Science Foundation of China (NSFC) No.61071111.

**REFERENCES**

- [1] 3GPP TR25.813, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Radio interface protocol aspects", www.3gpp.org, Oct. 2006
- [2] X. Chen, X.Jin, P.Moorut, R.Love, S. Yakun, W. Xiao, and et al., "Coexistence Analysis Involving 3GPP Long Term Evolution", Vehicular Technology Conference, Fall. 2007 IEEE 66th pp. 225-229
- [3] R.Zheng, X. Zhang, X. Li, Q. Pan, Y.Fang, and D.Yang, "Performance Evaluation on the Coexistence Scenario of Two 3GPP LTE Systems", Vehicular Technology Conference Fall,2009 pp. 1-6
- [4] M. Ng, S. Lin, J. Li, and S. Tatesh, "Coexistence Studies for 3GPP LTE with other mobile systems", IEEE Communications Magazine, Volume: 47, Issue: 4 2009, pp. 60-65
- [5] P. Chang, Y.Chang, J. Zhou, and D.Yang, "Interference Analysis of TD-LTE and TD-SCDMA System Coexistence", Networks Security Communications and Trusted Computing, Second International Conference on 24-25 April 2010 pp. 77-80
- [6] J.Wang, D.Yang, and R. Zheng, "Interference Analysis and Coexistence Studies between E-UTRA and UTRA Systems", Vehicular Technology Conference 2010 IEEE 71st, pp. 1-6
- [7] X. Huang, Q.Cui and X. Tao, "Coexistence studies involving LTE-advanced", Broadband Network and Multimedia Technology, 3rd IEEE International Conference on 26-28 Oct. 2010 pp. 273-277
- [8] Y.R. Zheng and C. Xiao, "Improved Models for the Generation of Multiple Uncorrelated Rayleigh Fading Waveforms", Communications Letters, 2002.6, pp. 256-258
- [9] "Monte Carlo methods", John Michael Hammersley, David Christopher Handscomb, 1964
- [10] 3GPP TR 36.942 (V8.3.0). "Evolved Universal Terrestrial Radio Access (E-UTRA) Radio Frequency (RF) system scenarios", www.3gpp.org, Oct.,2010
- [11] 3GPP2 C.S0010-D V1.0 "Recommended Minimum Performance Standards for cdma2000 Spread Spectrum Base Stations", www.3gpp2.org, Sept. 2010
- [12] 3GPP2 C.S0011-D V1.0 "Recommended Minimum Performance Standards for cdma2000 Spread Spectrum Mobile Stations", www.3gpp2.org, Mar., 2011

## Challenge Token-based Authentication – CTA

Ghassan Kbar

Associate Research Professor

Riyadh Techno Valley, King Saud University

Riyadh, Saudi Arabia

[gkbar@ksu.edu.sa](mailto:gkbar@ksu.edu.sa), [gahkbar@yahoo.com](mailto:gahkbar@yahoo.com)

**Abstract-** A new technique for highly securing the wired and wireless local area network using a Challenge Token-based Authentication as a second authentication factor is presented. This technique is based on two authentication factors, which is in addition to the first factor "user name and password", it also uses the client soft token that will be stored in a mobile phone or USB. The soft token will be obtained during registration and will never be transmitted during the authentication process. This token will be used by a mobile Client Program to generate a secure Authentication Server (AS) public key in order to respond to the AS's challenge. This new authentication mechanism addresses the vulnerabilities existed in the existing weak authentication method that is based only on first authentication factor. It would also solve the Denial of Service attack existed in the second authentication factor techniques because a secure server public key is used instead of well known server public key such as the one used in Extensible Authentication Protocol, and Wireless Application Protocol. In addition it reduces the complexity and associated cost existed in the mobile phone authentication technique since there is no need to send sms messages to authenticate the clients. In addition, the session key will be exchanged using the derived secure AS public key that is correlated to the soft token. This makes the security parameters known only to authentication server, and valid supplicants. Attackers would be unlikely able to know the token and other security keys since the token is only exchanged during registration through a trusted party. Moreover, the use of 2 authentication factors would make the security system stronger and more relevant to sensitive applications particularly for banks.

**Keywords-** wireless; authentication; security.

### I. INTRODUCTION

Strong authentication is extremely needed for e-transaction applications to secure the credential information including the one used in credit card. User name and password are widely used as the main authentication mechanism in computer security systems including the e-transaction applications. This mechanism has various drawbacks including the poor selection of password, the vulnerability to capture and crack password especially in wireless Local Area Network (LAN) and Wireless LAN (WLAN). Hence, the use of one factor authentication that is known as "something you know", and is mainly used in today system security, would not be sufficient for important secure applications.

The issue of authentication becomes more vulnerable in WLANs based on 802.11 standards if no authentication or weak encryption key is used. Although attempts are continuously being made to address the security issues in

the later versions of 802.11 [1], security of WLAN remains challenging. Despite the enhancements provided by Wired Equivalent Privacy (WEP) for WLANs, the demands for a further secured environment still a high priority issues in wireless network. By using the Extensible Authentication Protocol (EAP) for authentication, the access point responds by enabling a port for passing only the EAP packets from the client to an authentication server located on the wired side of the access point. Wi-Fi Protected Access (WPA) [2] also authenticates securely the wireless users to the network. WPA is a subset of the abilities of 802.11i, including better encryption with Temporal Key Integrity Protocol (TKIP), easier setup using a pre-shared key, and the ability to use RADIUS-based 802.1X authentication of users [3]. Wireless security will continue to be a concern for the foreseeable future. Although a single overall solution has yet to be perfected, the best protection is always prevention [4]. Comparison between the different EAP products has been presented in [5]. The issues of flexibility and high speed authentication technique, has been addressed using token-based fast Authentication method for Wireless network [6].

To strengthen the authentication system and improve the system security further, at least two modes of authentication should be used [7]. Hardware token is used in some applications to increase the security during authentication. However, the hardware token has its own limitations such as inconsistent availability, and loose of token sometimes. The primary mode of authentication would rely on user name and password, while the second authentication can be used for emergency in case the primary one is unavailable to users. There are many methods used for emergency authentications [8], but it will still be considered as one way authentication since it would only be used in case of forgetting the user name and password. Passwords and emergency type of life questions are often considered as "something you know," while hardware tokens are considered as "something you have." Another category of authenticator is "something you are," such as the biometric authentication. Some research try to introduce a forth-factor authentication "Somebody you know" [8], but still be considered as the first method by allowing other people to verify the users in case of emergency. A software token can be used as a second method of authentication and would be considered as "something you have". This combination of authentications is easy to implement and can achieve the purpose of having a strong authentication mechanisms.

Unlike the combination of the first method with hardware token which has its own limitation, or the combination with biometric authentication which is difficult to use in e-transaction applications, the combination with Soft token is preferable for ease of implementation as described in section IV. However, the transmission of the soft token has to be secure. Using tokens involves several steps including registration of users, token production and transmission, user and token authentication, and token revocation [10]. Some bank applications use a second authentication factor by sending an SMS message with code to registered mobiles, where the user has to reenter this code in order to continue the authentication process [11, 12]. This mechanism is costly, relies on technology such as GSM which might not be available all the time, and suffers from man-in-the-middle MiM attacks. Encrypting the SMS message would solve the MiM attacks as presented in [13]. Since many people carry a mobile phone at all times, an alternative is to install all the software tokens on the mobile phone [14]. This proposed system involves using a mobile phone as a software token for One Time Password (OTP) generation. However, if the mobile is out of sync with the server, it has to send a SMS message requesting the server to generate OTP and send it back to the mobile. This system might not be reliable and would be costly if it was out of sync where at least 2 SMS will be send to obtain the OTP.

In this paper a new proposed method for transmitting the soft token is used securely, where the token is only transmitted during registration through a trusted Registry Authority (RA) or trusted Authenticators in WLAN, and will be stored in a mobile phone or USB along with a secure key algorithm. The supplicant will *respond to a challenge request from the AS and forward it to the mobile or USB, where it will use the stored soft token to derive a secure server public key to communicate with the AS in order to securely exchange the session key* as will be described in the next section. The use of soft token in this paper is not to generate OTP or to send SMS message with code, but in fact to generate a secure AS public key that is correlated to the temporary key sent by the server, the stored token, and the secure key algorithm which all known only to valid supplicants and server. This will prevent the Denial of Service (DoS) attacks existed in other techniques, where attackers cannot use the well known server public key during authentication and are unable to generate the secure server public key which require a valid token and secure key Algorithm.

## II. CHALLENGE TOKEN BASED AUTHENTICATION (CTA) CONCEPT

In order to have a strong authentication mechanism when users want to access sensitive information from an Application Server (APS), 2 authentication factors should be used. Using user name and password (something you know) as the first authentication factor, in addition to the challenge/response token based authentication (something you have) will strengthen the authentication mechanism. Where the soft token will be issued by the AS to supplicants during the registration, and later can be used

by the supplicant to generate a secure AS public key (SsP) by applying a certain algorithm known to AS and to valid supplicants only. Only supplicants that can respond to the AS's challenge, by generating the SsP, would be able to communicate to AS in order to obtain the session key (SK) which will be used to communicate with the application server APS. In order to implement the concept of the challenge token based authentication (CTA), there is a need for a third party authority such as the Registry Authority (RA) server which is responsible for issuing a Temporary Registry Token (RT) for the valid supplicants. These supplicants will then use the RT to register with the AS and obtain a Permanent Token (PT) and a Secure Key Algorithm (SKA). The RT and SKA will be used by the supplicant to generate an SsP in order to communicate to the AS and to respond to the challenge request.

Figure 1 describes the different parameters that will be created and used by each party involved in the authentication process. At the user/client terminal, there are different authentication parameters used for registration/authentication, where the client terminal passes the username, password, client name and the client registry token (RT) to the RA, and to the Authentication Server (AS). The RA will then respond by passing a Permanent Token (PT) and a Secure Key Algorithm (SKA), which will be stored at the client side for future authentication. In addition, there are different encryption parameters used during registration, authentication, and transfer of messages. The public/private keys of client, registry authority and server are used to encrypt messages during the registration phase. While the secure public/private keys that are generated by the client terminal and server, are used to encrypt messages during the authentication phase. A symmetric session key is used for encrypting the messages sent between the client terminals and application servers.

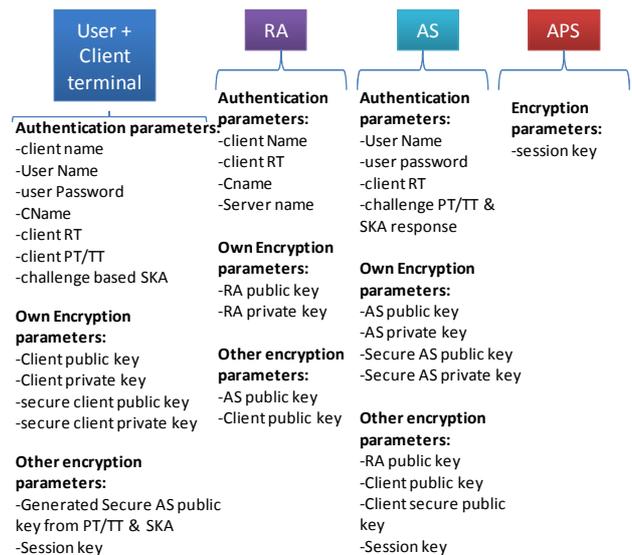


Figure 1. The different encryption parameters used by all parties during authentication

Figure 2 describes the steps required for registration with RA and AS, as well as the steps that are used during

authentication with the AS and during transmission of messages to the APS. The following steps describe the different phases of registration, authentication and access of applications:

- **AS Server to Register Authority (RA) Registration:**
  - (step 1,2): AS uses VPN to register to RA using RA's public/private key.
- **Client to Register Authority RA authentication:**
  - (step 3) Client registers with RA by passing user credential (user name UN, and client public key cP) encrypted by RA's public key.
  - (step 4) RA passes the RT to supplicant which will be used by supplicant to continue registration with the AS.
  - (step 5) Supplicant passes Challenge Name CName to be stored in RA for future retrieval of registry token.
  - (step 6) Supplicant requests the public key of AS by passing UN and AS flag encrypted by RA's public key to register with AS for future communication.
  - (step 7) RA passes the AS's public key to supplicant encrypted by client public key cP.
- **Supplicant to AS Registration:**
  - (step 8) Supplicant registers to AS by passing the RT and its public key cP encrypted by AS's public key sP.
  - (step 9) AS request client registry validation to RA using VPN.
  - (step 10) RA validate supplicant by checking its RT and reply to AS
  - (step 11) Following the supplicant validation, AS will send the PT and the SKA to the supplicant.
  - (step 12) The supplicant will forward the PT and SKA to mobile using Bluetooth technology or to USB which will then be stored in order to be used later to generate a secure server public key (SsP). The SsP will be used by supplicant to respond to challenge received from AS in order to complete the authentication between the supplicant and the AS.
  - (step 13) Supplicant will pass the user name UN and password PW encrypted by SsP to AS in order to be stored at the AS for future accessing the application servers. The AS will store UN and PW for further authentication process.
  - (step 14) AS sends registration complete the supplicant
- **Client Authentication to AS:**
  - (step 15) supplicant request access to the application server APS through AS, by sending its UN and PW encrypted by server public key sP.
  - (step 16) AS sends a challenge authentication request to the supplicant by passing a temporary server public key TsP encrypted by the supplicant public key cP.

- (step 17) supplicant will decrypt the challenge request, extract the TsP and forward it to the mobile or USB in order to get the secure AS public key SsP.
- (step 18) the mobile or USB, will receive the request from the supplicant and if user accepts it by pressing OK, its client token program will generate the SsP from the stored PT and received TsP by using the stored SKA algorithm and pass it to supplicant.
- (step 19) supplicant will generate a secure client public key ScP, and send the ScP encrypted by SsP to the AS.
- (step 20) AS decrypts the challenged message to obtain the ScP, and use it to encrypt and send a session key SK to the supplicant.
- (step 21) the supplicant will send the SK to the application server APS.
- **Exchange of messages between supplicant and APS:**
  - (step 22) the supplicant sends a message m1 to the APS encrypted by the SK.
  - (step 23) the APS will decrypt the message m1 using the SK and send message m2 encrypted by SK to the supplicant. The supplicant will be able to decrypt the message m2 using the same session key SK.

#### Advantages of the CTA mechanism:

- 2 authentication mechanisms are used; user name, password used as first authentication factor and client token for server challenge as second authentication factor.
- Token is never transferred during authentication, but is only transferred during registration to the RA and the AS, or it can be obtained through hardware USB, or it can be retrieved from the RA by passing challenge name CName.
- Used a derived secure AS public key SsP that is based on a stored token and correlation SKA algorithm to respond to AS challenge. This derived SsP can only be derived by the supplicant which it has the valid token and therefore hackers would not be able to communicate with the AS since they cannot generate the SsP.
- Exchange the session key SK using SsP is only known to valid supplicants. This will prevent attackers from using the normal server public key sP to launch a DOS attack at the AS.
- Compared to other second authentication factor used in other methods such as the one using hard or soft token for generating One Time Password or sending SMS message, the complexity in CTA is less, the cost is cheaper, and moreover the DOS attacks are eliminated since a secure server public key is used which is only known to valid supplicants.

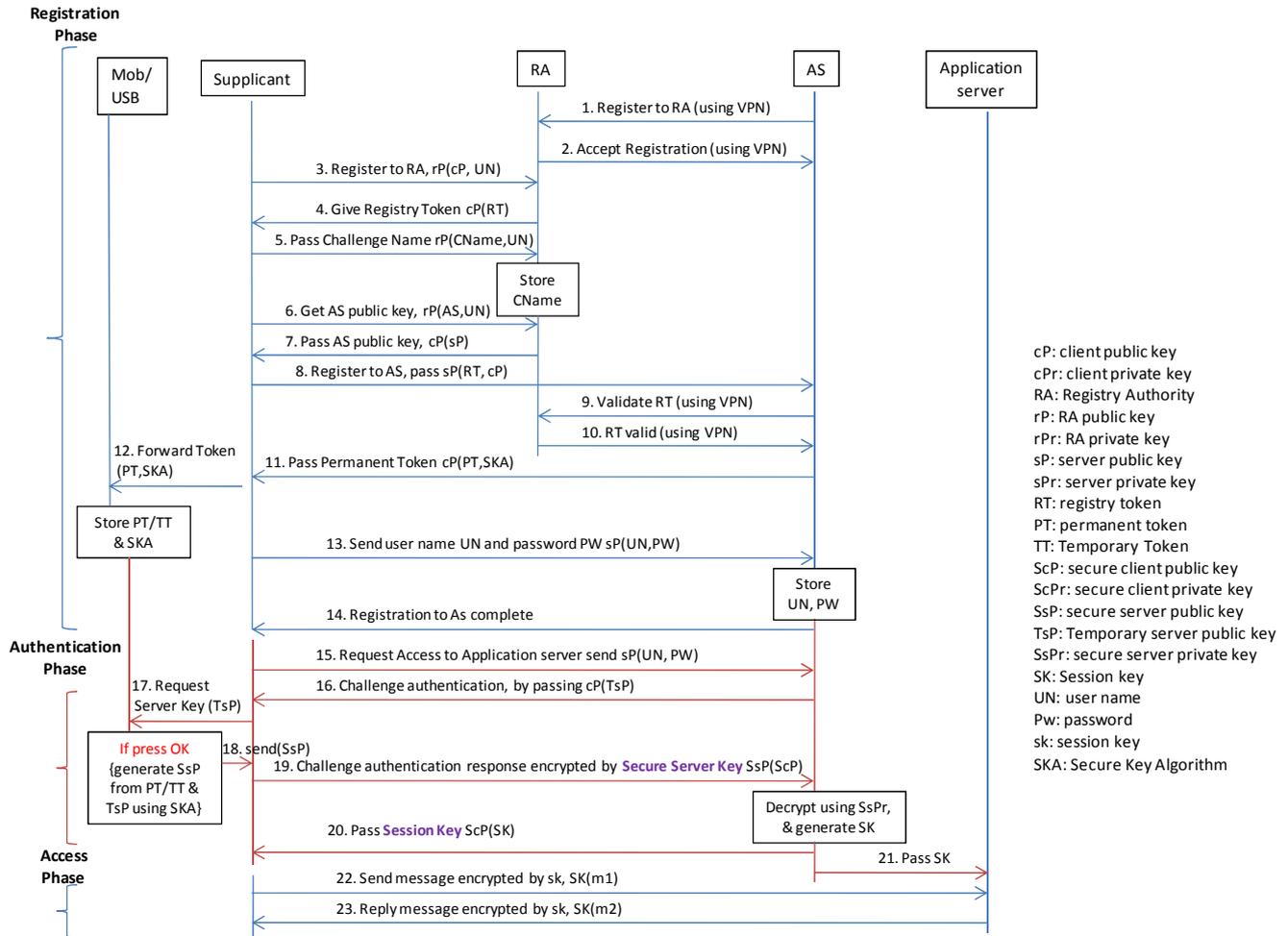


Figure 2. Challenge Token-Based Authentication process at different phases

### III. SECURE KEY ALGORITHM TO LINK PUBLIC KEYS WITH PERMENANT STORED TOKEN

During supplicant's registration with the AS, the supplicant will receive from AS a Permanent token PT as well as a Secure Key Algorithm (SKA) which correlates this token to a secure AS server public key SsP. This will assist the supplicant to generate the SsP later in respond to a challenge request sent by the AS to validate the supplicant.

During authentication, if the registration token is valid, user can submit a user credential which will be saved on the AS, and then will obtain a Permanent Token (PT) from the AS. A generated SsP would be used instead of normal authentication public key that is known to everyone to exchange a session key between the supplicant and the application server.

The following algorithm describes the correlation between the SsP and PT:

$X = PT$  {Permanent Token: stored in client side during registration}

$Y = TsP$  {temporary server Public key: is a random challenge key sent to client to create a Secure server public key SsP}  
 $Z = SsP$  {Secure server Public key: is a random server public key linked to TsP and PT through a KSA algorithm and never be transferred to client}

$$Y = f(X, Z),$$

$$Z = f(X, Y)$$

As example, by using the algorithm:  $X = 10*Y + 20*Z \rightarrow Y = (X - 20*Z)/10, Z = (X - 10*Y)/20$   
 For  $Y = 200, Z = 500 \rightarrow X = 2000 + 10000 = 12000$   
 If AS sends  $Y = 200$  to client, and client knows  $X = 12000$ , then client calculate  $Z = (12000 - 2000)/20 = 10000/20 = 500$

If the AS server is able to decrypt a message that is received from the supplicant using its own private key SsPr, it will extract the user name and password. A Successful

decryption is an indication that client used the same algorithm (by using PT (X) and TsP (Y)) to generate the SsP (Z).

This secure server public key SsP would then be used to send the user credential over a secure session, and once the server verifies the user credential and approve it, it sends a session key to the suppliant and the application server APS to enable them exchanging messages using the same session key SK.

A program that is under development for the supplicant side is intended to generate the SsP in response to the challenge of the server by correlating the received TsP with the stored SKA. The SsP will be used by the supplicant to encrypt the response that will be sent to the server, while the server would be able to decrypt the response and authenticate the supplicant.

Using the secure server public key SsP instead of a well known server public key used in other techniques such as the one used in WAP, would avoid the DoS attack that will try to bring down the AS. The session key will be generated by the AS following a successful authentication of supplicant, and would then be transferred to supplicant using a secure client public key ScP.

#### IV. IMPLEMENTATION ISSUES AND EVALUATION

The two factor authentication would strengthen the security, where a hardware token or software token can be used for this purpose. Hardware Tokens are small devices that are carried by customers, where these tokens usually store cryptographic keys or biometric data. They are used mainly to display a Personal Identification Number (PIN) that changes with time, where customers/users use the PIN displayed on the token in addition to the normal account and password during the authentication. There are several commercial two factor authentication systems using the hardware token [15, 16]. Many banks offered the use of hardware tokens to their customers such as Bank of Queensland, the Commonwealth Bank of Australia and the Bank of Ireland [10]. The hardware token has its own limitation where customers that uses more than one two-factor authentication system requires carrying multiple tokens/cards which are likely to get lost or stolen. In addition, the hardware token solution is costly, where organizations such as banks with a million of customers, have to purchase, handle and maintain a million of tokens, as well as training their customers on how to use the token. Moreover, there is additional cost associated with stolen, lost and broken tokens.

Software tokens on the other hand overcome the limitations associated with the use of hardware tokens, since the device holding the tokens can store multiple tokens which can be used for different authentication applications, and the cost is minimized since the same device can handle multiple tokens. Software tokens are programs that run on computer/device, where they can provide a PIN that change on time such as the One Time Password (OTP). Mobile phone can be used for storing the soft token [14], where it uses a client program to generate OTP locally. However, if the mobile is out of sync with the server, it has to send a SMS

message requesting the server to generate OTP and send it back to the mobile. This system might not be reliable if mobile network does not exist or it is out of coverage, and can be costly if it was out of sync where at least 2 SMS will be send to obtain the OTP.

In the proposed CTA method described in this paper, there is no need to send SMS messages in order to complete the authentication. In CTA, the supplicants respond to a challenge coming from the AS through the generation of a secure server public key (SsP) if the mobile phone can connect to the supplicants using WIFI or Bluetooth as shown in Figure 1, or through USB direct mobile connection in the absence of WIFI or Bluetooth connection. In both cases the stored secure key algorithm and to the stored software token are used to generate the SsP. A program that is under development for the supplicant side is intended to generate the SsP in response to the challenge of the server. This supplicant authentication (SA) program is a web interface, where if users want to login into a secure site using the CTA method, this program detects if WIFI or Bluetooth is enabled at the supplicant device so it passes the username and password entered by the users and waits for a challenge response from the server as shown in Figure 1. Once the challenge response is received from the AS server, it forwards it to the mobile phone, where the Mobile Authentication (MA) program will extract the TsP from the packet sent by the supplicant and generates the SsP using the stored permanent token and SKA, and then send it to the SA program. The SA program uses the SsP to encrypt and send the challenge response to the AS server. However, if the SA program did not detect any WIFI or Bluetooth on its device, users can connect the mobile phone to the device using USB connection and the MA program will interact with the SA program to complete the authentication using the same procedure described above.

#### V. OBTAINING A LOST REGISTRY TOKEN THROUGH REGISTER AUTHORITY

Figure 3 presents the different scenarios for obtaining key from RA for the CTA as follows:

- It can register in a Register Authority RA every time you want to use a new machine so it can store the registry token obtained from RA on the new machine or a mobile phone or USB as shown in Figure 3.a.
- You can store the registry token at mobile phone or secure USB using finger print to avoid losing it and allow the authenticated user to access it from the USB as shown in Figure 3.b.
- Suppliant can complete the registration to AS by obtaining the AS public key through RA as shown in Figure 3.c
- If suppliant uses a public machine or it lost the RT, it can obtain the RT from the RA, and after completing the registration, the RT will be stored in mobile phone of USB, and if it uses a public terminal then it should delete all the sessions. The request can be done by submitting a challenge name CName to RA, as shown in Figure 3.d.

## VI. CONCLUSION

In this paper a new technique "Challenge Token-based Authentication - CTA" has been presented, where, CTA uses 2 authentication factors, a normal user name and password authentication, and a challenge token based authentication. The first authentication used everywhere in current e-transaction applications, which is not secure enough since user name and password can be cracked especially in WLAN. The second authentication method CTA would make the authentication mechanism much stronger by using the soft token stored in a mobile phone or USB to generate a secure server public key SsP following the challenge request/response between supplicants and AS. Since the second authentication mechanism uses a token and a secure key algorithm that will never be transmitted during authentication, the derived SsP would be highly secure. The SsP will be used for exchanging the session key SK which is used for encrypting/decrypting the messages.

The use of SsP instead of known AS server public key will prevent any users from communicating with the AS unless they have a valid token and be able to generate the SsP. Consequently, hackers will not be able to use the cracked user name and password to access the AS, and they cannot launch a Denial of Service (DOS) attack on the AS since the SsP is only known to valid supplicants. The proposed CTA two factors authentication would eliminate the need for sending SMS messages that are used currently in the mobile second authentication techniques and therefore saving cost. Furthermore, the second authentication factor will eliminate hackers who know the User Name and password from accessing the application server since they cannot know the token and they would fail in generating the SsP. Implementing the mobile client program to generate the SsP using the secure key algorithm is in progress.

## REFERENCES

- [1] J. Geier, "802.1X Offers Authentication and Key Management" <http://www.wi-fiplanet.com/tutorials/article.php/1041171/8021X-Offers-Authentication-and-Key-Management.htm>, May 7 2002, accessed 17 April 2011.
- [2] J. Geier, "Wi-Fi Protected Access (WPA)", <http://www.wi-fiplanet.com/tutorials/article.php/2148721>, March 20 2003, accessed 17 April 2011.
- [3] <http://grouper.ieee.org/groups/802/11/>, July 18 2004
- [4] J. Rodrigues, "SECURITY FOR THE WIRELESS NETWORK" WatchGuard Technologies,
- [http://www.docstoc.com/docs/14980698/Wireless-Network-Security Inc, 2006-2007](http://www.docstoc.com/docs/14980698/Wireless-Network-Security-Inc,2006-2007), accessed 17 April 2011.
- [5] IEEE 802.11 "Wireless LAN Security with Microsoft Windows *Microsoft Corporation*" <http://www.microsoft.com/downloads/en/details.aspx?FamilyID=67fdeb48-74ec-4ee8-a650-334bb8ec38a9&displaylang=en> Published: January 2008, accessed 17 April 2011.
- [6] G. Kbar, "Wireless Network Token-Based Fast Authentication", IEEE ICT2010 International Conference on Telecommunication, Doha Qatar 4-7 April 2010, pp. 309-315.
- [7] B. Schneier, "Two-Factor Authentication: Too Little, Too Late," in Inside Risks 178, Communications of the ACM, 48(4), April 2005.
- [8] V. Griffith and M. J. Messin: "Deriving mothers maiden names using public records". In J. Ioannidis, A. D. Keromytis, and M. Yung, editors, Applied Cryptography and Network Security (ACNS), pp. 91–103. Springer-Verlag, 2005. LNCS no. 3531.
- [9] J. Brainard, A. Juels, and R. Rivest, "Fourth-Factor Authentication: Somebody You Know", CCS'06, , October 30–November 3, 2006, Alexandria, Virginia, USA. pp. 168-178, Copyright 2006 ACM1-59593-518-5/06/0010
- [10] D. de Borde, "Two-Factor Authentication," Siemens Enterprise Communications UK- Security Solutions, 2008. Available at [http://www.insight.co.uk/files/whitepapers/Two-factor%20authentication%20\(White%20paper\).pdf](http://www.insight.co.uk/files/whitepapers/Two-factor%20authentication%20(White%20paper).pdf), accessed 17 April 2011.
- [11] A. Kemshall and P. Underwood, "SecurEnvoy -white paper" July 2007, [http://www.securenvoy.com/WhitePapers/white\\_paper\\_two\\_factor\\_authentication.pdf](http://www.securenvoy.com/WhitePapers/white_paper_two_factor_authentication.pdf), accessed 17 April 2011.
- [12] Zhangjiajie, "Mobile Authentication Scheme Using SMS", IITA International Conference on Services Science, Management and Engineering, July 11-12 2009, pp. 161-164, ISBN: 978-0-7695-3729-0
- [13] M. Agoyi and D. Seral, "SMS SECURITY: AN ASYMMETRIC ENCRYPTION APPROACH", Sixth International Conference on Wireless and Mobile Communications ICWMC.2010.87, 20-25 September 2010, Valencia, Spain, pp. 448-452,
- [14] F. Aloul, S. Zahidi, and W. El-Hajj, "Two Factor Authentication Using Mobile Phones", The 7<sup>th</sup> ACS/IEEE International Conference on Computer Systems and Applications (AICCSA-2009), Rabat, Morocco, May 2009, pp. 641-644.
- [15] BestBuy's BesToken, <http://products.esecurityplanet.com/security/identity/BestBuy-Deluxe-BESTOKEN.html>, accessed 22 April 2011.
- [16] RSA's SecurID, <http://www.rsa.com/node.aspx?id=1156>, accessed 22 April 2011.

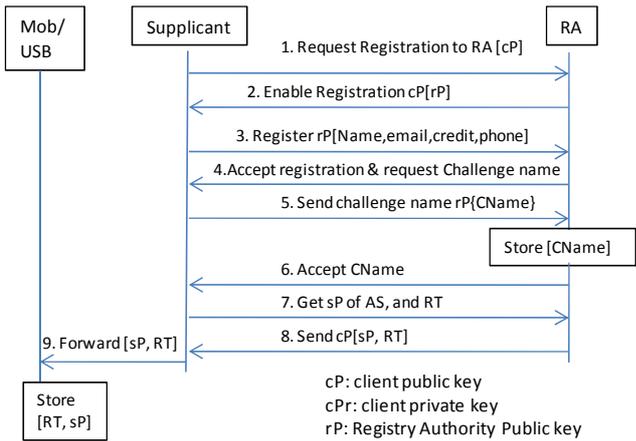


Figure3-a. Obtain Registry Token and AS server public key through Registry Authority through registration



Encrypted [RT, sP]

Figure3-b. Obtain Registry Token and AS server public key through secure USB

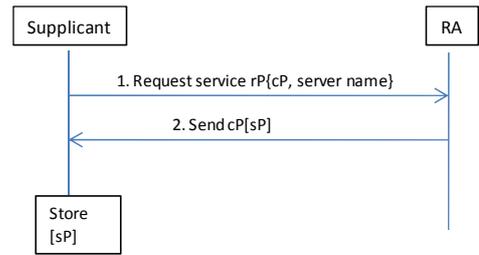


Figure3-c. Obtain AS server public key through Registry Authority after registration

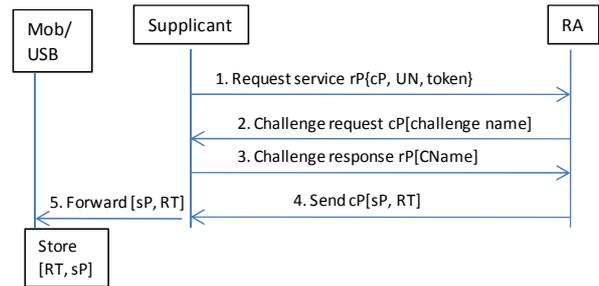


Figure3-d. Obtain lost/un-stored Registry Token and AS server public key through Root Authority after registration

Figure 3. Different scenarios for obtaining the token

## Efficient Privacy Preservation Protocol Using Self-certified Signature For VANETS

Bidi Ying<sup>1,2</sup>, Dimitrios Makrakis<sup>1</sup>, Hussein T. Mouftah<sup>1</sup>

Broadband Wireless & Internetworking Research Laboratory,

School of Information Technology and Engineering, University of Ottawa, Ottawa, Canada<sup>1</sup>

Zhejiang Gongshang University, Hangzhou, China<sup>2</sup>

{byiung, dimitris, mouftah} @site.uottawa.ca

**Abstract**—Privacy and security are two important issues in vehicular networks. Traditional authentication methodologies such as public/private keys & their corresponding certificates are not feasible to protect location privacy and security due to large computing overhead. In this paper, we propose an Efficient Privacy Preservation (EPP) protocol in vehicular ad-hoc networks, which, uses smart card functionality to authenticate users and employs bilinear pairings method to generate the public and private key. The public key is derived from the signature of the user's pseudonym identity and private key signed by a trust authority and roadside units, hence, users can verify signatures without their corresponding certificates. Performance analysis shows the proposed EPP protocol can authenticate vehicular users and data messages with low time complexity and preserve users' privacy.

**Keywords**-Vehicular networks, self-certified, privacy

### I. INTRODUCTION

Vehicular ad-hoc network (VANET) is a promising technology expected to play an important role in road safety, traffic management, and information dissemination to drivers and passengers [1]. A VANET mainly consists of On-Board Units (OBUs) and Roadside Units (RSUs) [2]. OBUs are installed on vehicles while RSUs are deployed to act as base stations providing connectivity to properly equipped vehicles located in their area of coverage.

Authentication is an important feature in a VANET as the source of the information should be verified to ensure the legitimacy of the data communicated [3-4]. Compare to wired network applications, VANET applications typically have more stringent authentication requirements. First of all, authentication should be done in short time in VANETs to ensure enough time for the drivers to take action. For e.g. a message update slower than once every 500 msec is probably too slow. Driver reaction time to stimuli like brake lights can be of the order of 0.7 Sec and higher [5]. Thus if updates come in slower than every 500msec, the driver may realize something is wrong before the safety system. This would make the driver think the safety system is not effective. Second, a certain degree of anonymity is typically required to ensure privacy of drivers, and the authentication model must ensure that this anonymity is maintained. For example, a rouge user could target and succeed to collect messages generated by other vehicular users and obtain sensitive information such as the driver's name, license plate, speed, location of vehicle, route of travel without successful

security and privacy guarantee mechanisms in place. Third, VANETs are highly mobile and the mobility should be considered when designing the authentication protocol for possible network partitioning. For instance, if two cars drive in opposite directions with 90 Km/h each, and if we assume a theoretical wireless transmission range of 300meters, communication is only possible for 12 seconds.

Most of the existing security proposals for secure VANETs are based on the use of an asymmetric algorithm [1, 3-4]. For example, an algorithm using public/private keys and their corresponding anonymous certificates to authenticate messages requires larger storage of a huge number of keys and larger computing overhead. A short-time signature using bilinear pairings was proposed in [6] to use in the electronic cash system, which allows a user to get a signature without giving the signer any information about the actual message or the resulting signature. However, this short-time signature is impractical for vehicular networks due to high mobility. Girault [7] first introduced self-certified public key, where the private key of each user is only known to the user himself, while the corresponding public key is derived from the signature of the user's identity and private key. This self-certified method can implicitly validate the user's public keys, and cannot need extra corresponding certificates. Hence, it can reduce storage space. Shuo [8] proposed a further research about Self-Certified Signature (SCS) where users can choose their private keys and the actual public key consists of the public key of a Trust Authority (TA) and the partial public key chooses by the user, along with the identity of the user explicitly. However, when sending the signed message together with the public key  $Y_{TA}$ , its partial public key  $Y_{ID}$  and its identify  $ID$ , an attacker still can link its  $ID$  to the user by monitoring messages. Thus, it can reveal private information regarding the activities of the user. Besides, real-time data from users is to be accessed directly by an external party (eg. attacker), which will leak sensitive information to the external party.

In order to solve above problems, we introduce an efficient privacy preservation protocol, which is based on bilinear pairings. EPP scheme uses smart card to authenticate users and RSUs before allowing the user to access data; and employs bilinear pairings to generate the partial private key. The actual private key including two partial private keys is used to sign a data message, and the corresponding public key derived from the user's pseudo ID/ partial public key and

the TA's public key is used to verify the signature. This strategy helps to avoid some active attacks such as forgery attack and guess attack, as will be shown in the security analysis section. Compared to the SCS scheme, the time complexity of EPP scheme remains lower, as will be shown in the time complexity analysis section.

The remainder of the paper is organized as follows. Section II presents background. In Section III, the EPP scheme is described in detail. Section IV provides performance analysis. In Section V, the related work is surveyed. Finally, Section VI concludes the paper.

## II. BACKGROUND

### A. Definition of Bilinear Pairings

Bilinear pairing is an important cryptographic primitive and has recently been applied in many positive applications in cryptography [9].

Let  $G_1$  and  $G_2$  be two cyclic groups of the same prime order  $q$ . We write the group laws of  $G_1$  and  $G_2$  additively and multiplicatively, respectively. Let  $P$  be a generator of  $G_1$ , assume that the discrete logarithm problems in  $G_1$  and  $G_2$  are hard. An efficient admissible bilinear map  $e: G_1 \times G_1 \rightarrow G_2$  with the following properties: ①Bilinear: for all  $P, P' \in G_1$  and  $a, b \in Z_q^*$ ,  $e(aP, bP') = e(P, P')^{ab}$ ; ②Non-degenerate: there exist  $P, P' \in G_1$  such that  $e(P, P') \neq 1$ ; ③Computable: there is an efficient algorithm to compute  $e(P, P')$  for any  $P, P' \in G_1$ .

We review related underlying mathematics problems [9] in  $G_1$ , which will server as the basis for our proposed EPP scheme.

① *Bilinear Diffie-Hellman (BDH) parameter generator*: a randomized algorithm  $IG$  is a BDH parameter generator if  $IG$  takes a security parameter  $k > 0$ , runs in time polynomial in  $k$ , and outputs the description of two groups  $G_1$  and  $G_2$  of the same large prime order  $q$  and the description of an admissible pairing  $e: G_1 \times G_1 \rightarrow G_2$

② *Discrete Logarithm (DL) Problem*: Given  $P, P' \in Q_1$ , for unknown  $n \in Z_q^*$ , compute  $P' = nP$ .

③ *Computational Diffie-Hellman (CDH) Problem*: Given  $P, aP, bP \in G_1$ , for unknown  $a, b \in Z_q^*$ , compute  $abP \in G_1$ .

④ *Decisional Diffie-Hellman (DDH) Problem*: Given  $P, aP, bP, cP \in G_1$ , for unknown  $a, b, c \in Z_q^*$ , decide whether  $c = ab \pmod q$ . It is known that DDH problem in  $G_1$  is easy and can be solved in polynomial time according to  $e(P, P)^{ab} = e(P, P)^c$  [9].

⑤ *Gap Diffie-Hellman (GDH) Problem*: If DDH problem in  $G_1$  is easy and CDH problem in  $G_1$  is hard, call  $G_1$  is GDH.

### B. Self-certified Signature by Bilinear Pairings

Shuo [8] proposed the SCS scheme using bilinear pairings, which includes KeyGen, Extract, Sign, and Verify. ① *KeyGen*: It takes a security parameter  $k$  as input and returns system parameters. The system parameters include two cryptographic hash functions  $H$  and  $H_1$ . The TA chooses a master-key  $s$  and computes the corresponding public key  $Y_{TA}$ . Each user chooses partial private key  $x_{ID}$  and computes the corresponding partial public key  $Y_{ID}$ . ② *Extract*: The TA generates the partial private key  $d_{ID}$  by input the system parameters, the master-key  $s$ , the partial public key  $Y_{ID}$  and an arbitrary  $ID \in \{0,1\}^*$ , the infinite set of all binary strings. Then TA sends  $d_{ID}$  securely to the user. The user can get the actual private key  $\langle x_{ID}, d_{ID} \rangle$  and the actual public key  $\langle Y_{TA}, ID, Y_{ID} \rangle$ . ③ *Sign*: An user signs any message  $M$  with its actual private key  $\langle x_{ID}, d_{ID} \rangle$ . ④ *Verify*: Any verifier can validate the signed  $M$  by checking the verification equation with respect to the actual public key  $\langle Y_{TA}, ID, Y_{ID} \rangle$ .

In the SCS method, the actual public key  $\langle Y_{TA}, ID, Y_{ID} \rangle$  is verified implicitly through the subsequent use of the public key  $Y_{TA}$ , hence, two users can directly use the partial public keys and the corresponding partial private keys to work. It can decrease much storage space and reduce authentication message time. However, when sending the signed message together with the public key  $Y_{TA}$ , its partial public key  $Y_{ID}$  and its identify  $ID$ , an attacker still can link its  $ID$  to the user by monitoring messages. Thus, it can reveal private information regarding the activities of the user.

## III. EFFICIENT PRIVACY PRESERVATION PROTOCOL USING SELF-CERTIFIED SIGNATURES

### A. System Formulation

Fig.1 illustrates the network architecture, which consists of three entities: the top Trusted Authority (TA), the RSUs located at the road side, and the OBUs located on the vehicles. The TA is responsible for the registration of RSUs and OBUs and is assumed that is having sufficient computation and storage capabilities. RSUs are assumed to connect with the TA by wire or wireless links. Wireless access between vehicles as well as vehicles and RSUs is conducted through networks complying with the IEEE 802.11p standard [5].

Made assumptions are the following.

- 1) TA is fully trusted by all parties in the system, and is not possible for an adversary to compromise it.
- 2) RSUs are semi-trusted entity. The TA can inspect all RSUs at the high level. Once an RSU is compromised in one time slot, the TA can detect and take action to recover it in the next time slot [11].

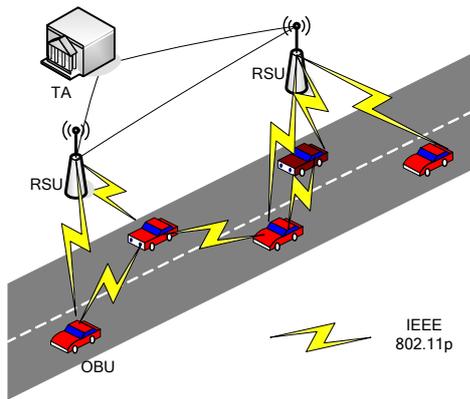


Figure 1. System model

### B. EPP scheme

We assume that RSUs and OBUs are not trust. Hence, before entering into the VANET, RSUs and OBUs should be registered to the TA, and then OBUs obtain their actual public through the TA and RSUs. Therefore, the proposed EPP protocol consists of four parts shown in Fig. 2.

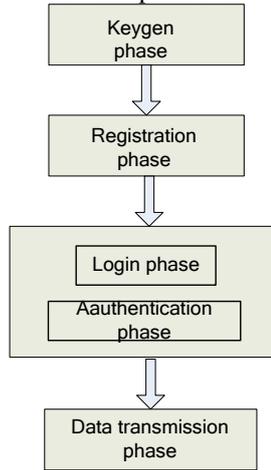


Figure 2. EPP scheme

① Keygen phase, which generates system parameters for the self-certified signatures; ② Registration phase, where RSUs & users register into TA and obtain their smart cards; ③ Login/authentication phase, where a vehicular user should be authenticated before entering into the vehicular network and obtain its actual public key from the TA and RSUs and its actual private key generated by the TA and RSUs; ④ Data transmission phase: where each vehicular user signs any messages by its actual private key, receivers verify the signature by using its actual public key.

#### 1) Keygen phase

Let  $k$  be the security parameter and a randomized algorithm  $IG$  be a BDH parameter generator satisfying the GDH Problem. Let  $G_1$  and  $G_2$  be two cyclic groups of the same prime order  $q$ . We write the group laws of  $G_1$  and  $G_2$  additively and multiplicatively, respectively. Let  $P$  be a generator of  $G_1$ , assume that the discrete logarithm problems

in  $G_1$  and  $G_2$  are hard. An efficient admissible bilinear map is  $e: G_1 \times G_1 \rightarrow G_2$ . The TA first generates the bilinear parameters  $(q, G_1, G_2, e, P)$  by running the randomized algorithm  $IG$ . Then, the TA chooses two cryptographic hash functions:  $H, H_1: \{0,1\}^* \rightarrow Z_q^*$ , and random selects  $s \in Z_q^*$  as its private key, and computes  $Y_{TA} = sP$  as its public key. The system parameters will be published, which include  $\{q, G_1, G_2, e, P, H, H_1, Y_{TA}\}$ .

#### 2) Registration phase

When a RSU or vehicle submits its identity to the TA for registering itself, the TA will do the following function:

① For a vehicle:  $V_i$  submits its identity ( $ID_{V_i}$ ) and

password ( $PW_i$ ) to the TA. Upon receiving the registration request, TA will compute vehicle  $V_i$ 's pseudonym  $PVID_{V_i} = h(ID_{V_i})$ , pseudo password  $\alpha_i = h(PW_i)$ ,  $\gamma_i = \alpha_i \oplus x_s$ ,  $N_i = h(PVID_{V_i} \| x_s)$ . Note that we use  $\gamma_i$  and  $N_i$  to hide the parameter  $x_s$ . Then it sends a smart card with  $\langle h(\bullet), PVID_{V_i}, \alpha_i, \gamma_i, N_i \rangle$  to  $V_i$ . Here,  $h(\bullet)$  is a Hash function;  $x_s$  is a secret parameter generated securely by the TA;  $\oplus$  is XOR operation.

② For the RSU:  $R_i$  submits its identity ( $ID_{R_i}$ ) and

location ( $L_i$ ) to the TA. After receiving the registration request, the TA computes  $\Phi_i = h(ID_{R_i} \| L_i)$  and  $\Psi_i = h(L_i) \oplus h(ID_{R_i}) \oplus x_b$ , then sends a smart card with  $\langle h(\bullet), h(ID_{R_i}), \Phi_i, h(L_i), \Psi_i \rangle$  to  $R_i$ . Note that we use  $\Psi_i$  to hide the parameter  $x_b$ . Here,  $h(\bullet)$  is a Hash function;  $x_b$  is a secret parameter generated securely by the TA.

#### 3) Login/authentication phase

Login and authentication phase is invoked when  $V_i$  wants to enter into this network. The login and authentication phase is as follows.

##### ① Login phase:

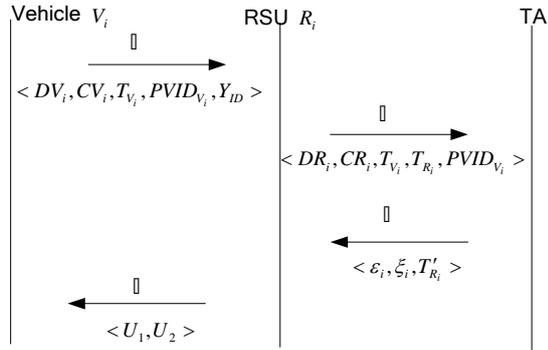
$V_i$  inserts its smart card to a terminal and  $ID_{V_i}$  and  $PW_i$ . The smart card validates  $ID_{V_i}$  and  $PW_i$  with the stored ones in it by the functions by the functions  $h(ID_{V_i})^* = PVID_{V_i}$  and  $h(PW_i)^* = \alpha_i$ , and then computes  $x_s^* = \gamma_i \oplus h(PW_i)^*$  and verifies  $h(h(ID_{V_i})^* \| x_s^*) = N_i$ . If it is not true, the smart card terminates the process and send rejection message to  $V_i$ .

$R_i$  inserts its smart card to a terminal and  $ID_{R_i}$  and  $L_i$ . The smart card validates  $ID_{R_i}$  and  $L_i$  with the stored ones in it by the function  $h(L_i)^* = h(L_i)$  and  $h(ID_{R_i})^* = h(ID_{R_i})$ . If it is not true, the smart card terminates the process and send

rejection message to  $R_i$ . Finally,  $R_i$  can obtain  $x_b^*$  by computing  $x_b^* = \Psi_i \oplus h(L_i) \oplus h(ID_{R_i})$ .

### ② Authentication phase

Authentication phase is invoked when  $V_i$  wants to enter into this network. The authentication phase is shown as Fig.3. The details are as follows.



$$\square : DV_i = h(\alpha_i \parallel PVID_{V_i}) \oplus h(T_{V_i} \parallel x_s^*) \quad CV_i = h(N_i \parallel \gamma_i \parallel T_{V_i})$$

$$\square : DR_i = h(DV_i \parallel CV_i \parallel \Phi_i) \oplus h(x_b^* \parallel T_{R_i} \parallel T_{V_i}) \\ CR_i = h(h(\Psi_i \parallel DV_i \parallel CV_i) \parallel x_b^* \parallel T_{R_i} \parallel T_{V_i})$$

$$\square : \varepsilon_i = s \oplus h(ID_{R_i} \parallel L_i \parallel T'_{R_i} \parallel PVID_{V_i}) \\ \xi_i = h(s \oplus h(ID_{R_i} \parallel L_i))$$

$$\square : U_1 = rP \quad U_2 = d_{ID} + rY_{ID}$$

Figure 3. Authentication phase

**Step 1:** The smart card in  $V_i$  picks a random  $x_{ID} \in Z_q^*$  as its partial private key and sets  $Y_{ID} = x_{ID}P$ , then will perform

· Compute  $V_i$ ' dynamic login identity

$$DV_i = h(\alpha_i \parallel PVID_{V_i}) \oplus h(T_{V_i} \parallel x_s^*)$$

· Compute  $CV_i = h(N_i \parallel \gamma_i \parallel T_{V_i})$

· Send  $\langle DV_i, CV_i, T_{V_i}, PVID_{V_i}, Y_{ID} \rangle$  to the RUS  $R_i$ , where  $T_{V_i}$  is the current timestamp when sending the message.

**Step 2:** After receiving the login message,  $R_i$  will perform as follows:

· Verify  $(T - T_{V_i}) \leq \Delta T$ , if it dose not hold, abort the process. Otherwise, store  $PVID_{V_i}$  and  $Y_{ID}$ , where  $T$  is the current timestamp when receiving the message,  $\Delta T$  denotes the expected time interval for the transmission delay.

· Compute  $R_i$ ' dynamic login identity

$DR_i = h(DV_i \parallel CV_i \parallel \Phi_i) \oplus h(x_b^* \parallel T_{R_i} \parallel T_{V_i})$ , where  $T_{R_i}$  is the current timestamp when sending the message.

· Compute  $CR_i = h(h(\Psi_i \parallel DV_i \parallel CV_i) \parallel x_b^* \parallel T_{R_i} \parallel T_{V_i})$

· Send  $\langle DR_i, CR_i, T_{V_i}, T_{R_i}, PVID_{V_i} \rangle$  to the TA.

**Step 3:** Upon receiving the message, the TA will perform as follows:

· Verify  $(T - T_{R_i}) \leq \Delta T$ , if it dose not hold, abort the process.

· Compute

$$DV_i^* = h(h(PW_i) \parallel PVID_{V_i}) \oplus h(T_{V_i} \parallel x_s)$$

· Compute

$$CV_i^* = h(h(PVID_{V_i} \parallel x_s) \parallel (h(PW_i) \oplus x_s) \parallel T_{V_i})$$

· Compute

$$h(DV_i^* \parallel CV_i^* \parallel h(ID_{R_i} \parallel L_i))^* = DR_i \oplus h(x_b \parallel T_{R_i} \parallel T_{V_i})$$

· Compute

$$CR_i^* = h(h(DV_i^* \parallel CV_i^* \parallel (x_b \oplus h(ID_{R_i}) \oplus h(L_i))) \parallel x_b \parallel T_{R_i} \parallel T_{V_i})$$

· Verify  $CR_i^* \stackrel{?}{=} CR_i$ , if it is not true, reject the message.

· Compute  $\varepsilon_i = s \oplus h(ID_{R_i} \parallel L_i \parallel T'_{R_i} \parallel PVID_{V_i})$

· Compute  $\xi_i = h(s \oplus h(ID_{R_i} \parallel L_i))$

· Send message  $\langle \varepsilon_i, \xi_i, T'_{R_i} \rangle$ , where  $T'_{R_i}$  is the current timestamp when sending the message.

**Step 4:** Upon receiving the message,  $R_i$  will perform as follows:

· Verify  $(T - T'_{R_i}) \leq \Delta T$ , if it dose not hold, abort the process.

· Compute  $s^* = \varepsilon_i \oplus h(ID_{R_i} \parallel L_i \parallel T'_{R_i} \parallel PVID_{V_i})$

· Compute  $\xi_i^* = h(s^* \oplus \Phi_i)$

· Verify  $\xi_i^* \stackrel{?}{=} \xi_i$ , if it dose not hold, abort the process.

· Compute  $H_{ID} = H(Y_{TA}, PVID_{V_i}, Y_{ID}) \in G_1^*$ , and sets the partial private key  $d_{ID} = s^* H_{ID}$ .

· Choose random integer  $r \in Z_q^*$  and compute  $U_1 = rP$ ,

$$U_2 = d_{ID} + rY_{ID}$$

· Send message  $\langle U_1, U_2 \rangle$  to  $V_i$ .

**Step 5:** After receiving the message,  $V_i$  will perform as follows:

· Compute  $d_{ID}^* = U_2 - x_{ID}U_1$

· Verify  $e(d_{ID}^*, P) \stackrel{?}{=} e(H(Y_{TA} \parallel PVID_{V_i} \parallel Y_{ID}), Y_{TA})$ , if it dose not hold, abort the process. Otherwise,  $d_{ID}^*$  is the secret certificate of the TA's public key  $Y_{TA}$ , the partial public key  $Y_{ID}$  and the vehicle  $V_i$ '  $PVID_{V_i}$ . Thus, the  $V_i$  obtains his actual private key  $(x_{ID}, d_{ID}^*)$ . Hence, the actual public key  $(Y_{TA}, PVID_{V_i}, Y_{ID})$  is used as the private key for signing.

4) Data transmission phase

**Step 1:** To sign a message  $M_j$ ,  $V_i$  randomly chooses an integer  $a \in Z_q^*$  and performs

- Compute  $H_{ID} = H(Y_{TA} \parallel PVID_{V_i} \parallel Y_{ID})$
- Compute  $R = aY_{TA}$
- Compute  $f = H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})$
- Compute  $\beta = fad_{ID} + x_{ID}H_{ID}$
- Send the message  $\langle \beta, R, PVID_{V_i}, Y_{ID}, M_j, T_1 \rangle$

**Step 2:** To verify the signature  $(R, \beta)$ , the verifier  $V_{i+1}$  will perform

· Verify  $(T - T_1) \leq \Delta T$ , if it dose not hold, abort the process.

· Compute  $H_{ID} = H(Y_{TA} \parallel PVID_{V_i} \parallel Y_{ID})$

· Compute

$$e(\beta, P) = e(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})R + Y_{TA}, H_{ID}), \text{ if}$$

it dose not hold, abort the process.

Hence, if two vehicles follow this protocol, the verifier will always accept the signature  $(R, \beta)$  and be convinced of the authenticity of the partial public key of  $V_i$ .

#### IV. PERFORMANCE ANALYSES

##### A. Security Analysis

###### 1) Resilience to stolen smart attack

Assume that the smart card of a vehicle is stolen or lost, then the attacker can extract the secret information  $\{h(\bullet), PVID_{V_i}, \alpha_i, \gamma_i, N_i\}$  from it by side channel attacks and invasive attacks [12]. However, even though an adversary taking control of the smart can obtain  $\gamma_i$ , it is practically infeasible for the adversary to know  $x_s$  by inferring  $\gamma_i$  or  $N_i$ , because of the one-way property of  $h(\bullet)$ . Therefore, the attacker cannot generate a valid login message  $DV'_i$  ( $DV'_i = h(\alpha_i \parallel PVID_{V_i}) \oplus h(T_{V_i} \parallel x_s)$ ).

For the RSU  $R_i$ , the attacker can obtain secret information  $\{h(\bullet), h(ID_{R_i}), \Phi_i, h(L_i), \Psi_i\}$  from the smart card, however, he cannot forge a fake message  $DR'_i$  ( $DR'_i = h(DV_i \parallel CV_i \parallel h(ID_{R_i} \parallel L_i)) \oplus h(x_b^* \parallel T_{R_i} \parallel T_{V_i})$ ) without knowing  $x_b$ .

###### 2) Resilience to guessing attack

Guessing attack is a crucial concern to any password-protected system [13]. Our scheme can resist the guessing attack, since the communication units within vehicles do not contain password and IDs. The attacker might try different passwords in its effort to construct  $DV'_i$ , however, the probability of failing is very high, because dose not have knowledge of  $x_s$ .

###### 3) Resilient to replay attack

Assume that the attacker intercepts a valid login message  $\langle DV_i, CV_i, T_{V_i} \rangle$  and tries to login to the RSUs by replaying the same message. The verification of this login message fails because the interval  $(T' - T_{V_i}) > \Delta T$  ( $\Delta T$  denotes the expected time interval for the transmission delay), where  $T'$  is RSU's system time when receiving the replayed message.

We also include a timestamp in each data packet in order to verify during the data transmission phase the message's validity. Thus, the replay attack is also prevented.

###### 4) Validity

In the data transmission phase, the receiver can compute

$$\begin{aligned} e(\beta, P) &= e(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})ad_{ID} + x_{ID}H_{ID}, P) \\ &= e((H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})as + x_{ID})H_{ID}, P) \\ &= e((H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})as + x_{ID})P, H_{ID}) \\ &= e(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})R + Y_{TA}, H_{ID}) \\ &= e(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})R + Y_{TA}, H(Y_{TA} \parallel PVID_{V_i} \parallel Y_{ID})) \end{aligned}$$

Hence, the proposed EPP is validity.

###### 5) Resilient to forgery attack

**Theorem1.** The proposed EPP scheme is unforgeable under the assumption of the DL problem.

**Proof.** If the attacker wants to forge a signature  $(R', \beta')$ , he has to make sure the following verification correct:

$$e(\beta', P) = e(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})R + Y_{TA}, H(Y_{TA} \parallel PVID_{V_i} \parallel Y_{ID}))$$

If the attacker knows  $(H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})R + Y_{TA})$  be the discrete logarithm problem in  $P$ , given  $r \in Z_q^*$ , then assume that  $\beta = rH_{ID}$ , where  $r = (H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})as + x_{ID})$ . The attacker knows  $a$  and computes  $H_1(M_j \parallel R \parallel H_{ID} \parallel PVID_{V_i})$ , however, he dose not know  $s$  and  $x_{ID}$  by computing  $Y_{TA} = sP, Y_{ID} = x_{ID}P$  because of the discrete logarithm problem in  $G_1$  and  $G_2$ .

##### B. Time Complexity Analysis

Assume that  $T_{pmul}$  represents the time for one point multiplication computation in  $G_1$ ;  $T_{padd}$  represents the time for one point addition computation in  $G_1$ ;  $T_{pair}$  denotes the time for one pairing computation, and  $T_{hash}$  denotes that the time for one hash function. Note that the time complexity for other computation operations, such as the multiplication in  $Z_q^*$ , are ignored, since they are much smaller than  $T_{pmul}$ ,  $T_{padd}$ ,  $T_{pair}$  and  $T_{hash}$  [10]. Table I shows the time complexity of the proposed EPP scheme and the SCS scheme [8]. Compared to the SCS scheme in signing a message, the proposed EPP scheme reduces the time of  $(T_{pair} + T_{padd})$ . From the paper [10], we can see that  $T_{pair}$  is much larger than other operations.

TABLE I TIME COMPLEXITY

Scheme	Sign Message	Verify Message
SCS scheme	$T_{pair} + 3T_{pmul}$ $+ 2T_{padd} + 2T_{hash}$	$2T_{pair} + 2T_{pmul}$ $+ T_{padd} + 2T_{hash}$
EPP scheme	$3T_{pmul}$ $+ T_{padd} + 2T_{hash}$	$2T_{pair} + T_{pmul}$ $+ T_{padd} + 2T_{hash}$

## V. RELATED WORK

User authentication is very important feature in a VANET as the source of the information should be verified to ensure the legitimacy of the communicated data [3-5]. To address such issues in VANETs, Raya et al. [1] introduced a security protocol for VANETs by installing a large number of private keys and their corresponding anonymous certificates to each vehicle. Instead of taking any real identity information of the drivers, these anonymous certificates are generated by taking the pseudo IDs of the vehicles. Existing PKI-based security schemes are prohibitively inefficient due to their computational complexity and obviously cannot scale to large vehicle populations.

A possible approach to reduce the overhead of the PKI-based security schemes is to improve the verification efficiency. We can do so by using the short group signatures method [14, 15], since it can quickly verify a large number of signatures simultaneously instead of sequentially by decreasing the number of some principal time-consuming operations. However, these methods assume that all verified signatures are authentic, and therefore, they need to be optimized for realistic applications, where bogus signatures commonly exist.

A short-time signature using bilinear pairings was proposed in [6] to use in the electronic cash system, which allows a user to get a signature without giving the signer any information about the actual message or the resulting signature. Liu et al. [11] proposed an efficient conditional private preservation protocol which authenticates users & RSUs by using Wei or Tate pairings on the elliptic curves and authenticates data messages by the short-time anonymous keys within certificates. However, the computing overhead to authenticate users & RSUs is still much high (the execution time (computing time) of authenticating users & RSUs is about 34.8msec in paper [11]). Liu et al. [16] proposed a secure protocol based on group signature and identity-based signature techniques by bilinear pairings. A signature based on identity is adopted in the RSUs to digitally sign each message by the RSUs, which reduces signature overhead.

## VI. CONCLUSIONS

In this paper, we presented an efficient privacy preservation protocol by using smart card functionality and bilinear pairings method for secure communications in vehicular network. Since users sign messages by their actual

private keys and verify these messages only with their actual public keys no corresponding certificates, the EPP protocol has been identified to be not only capable of providing the conditional privacy preservation that is critically demanded in the VANET applications, but also achieves high efficiency in terms of time complexity.

## ACKNOWLEDGMENT

This work was supported by the Government of Ontario under the ORF-RE WISENSE project (3074600) and the Natural Sciences and Engineering Research Council (NSERC) of Canada under NSERC Grant 193961-2006.

## REFERENCES

- [1] M. Raya and J. P. Hubaux, "Securing vehicular ad hoc networks," *Journal of Computer Security*, vol. 15, no. 1, pp. 39-68, 2007.
- [2] Y. Peng, Z. Abichar, and J. M. Chang, "Roadside-aided routing (RAR) in vehicular networks," in *Proc. IEEE ICC 2006, Istanbul, Turkey*, vol. 8, pp. 3602-3607, June 2006.
- [3] K. Ren, W. Lou, R. H. Deng, and K. Kim, "A novel privacy preserving authentication and access control scheme in pervasive computing environments," *IEEE Transaction on Vehicular Technology*, vol. 55, no. 4, pp.1373-1384, July 2006.
- [4] H. Moustafa, G. Boudron, and Y. Gourhand, "AAA in vehicular communication on highways with ad hoc networking support: a proposed architecture," in *Proc. International Workshop on Vehicular Ad Hoc Networks (VANET)*, Germany, 2005.
- [5] "Dedicated Short Range Communications (DSRC)," [Online]. Available: <http://grouper.ieee.org/groups/scc32/dsrc/index.html>.
- [6] D. Chaum, "Blind signatures for untraceable payments," in *Proc. Advances in Cryptology - Crypto'82*, Santa Barbara, California, USA, pp. 199-203, Aug. 1982.
- [7] M. Girault, "Self-certified public keys," in *Proc. Advances in Cryptology- Eurocrypt'91*, LNCS 1440, Springer-Verlag, Brighton, UK, pp. 491-497, 1991.
- [8] Z.H. Shao, "Self-certified signature scheme from parings," *The Journal of Systems and Software*, 2007, 80 (2) : 388 - 395.
- [9] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the Weil paring," in *Proc. Advances in Cryptology - Asiacrypt'01*, LNCS 2248, Springer-Verlag, Gold Coast, Australia, pp. 514-532, 2001.
- [10] "Crypto++ 5.6.0 Benchmarks," <http://www.cryptopp.com/benchmarks.html>.
- [11] R. X. Liu, et al., "ECP: efficient conditional privacy preservation protocol; for secure vehicular communications," *The 27th Conference on Computer Communications*. IEEE, 2008, pp. 1229-1237.
- [12] K. Markantonakis, et al., "Attacking smart card systems: theory and practice," *Information Security Technical Report*, Vol. 14, Issue 2, May 2009, pp. 46-56.
- [13] M.L. Das, "Two-factor user authentication in wireless sensor networks," *IEEE Trans. Wireless Comm.* 2009, 8, pp. 2450-2459.
- [14] D. Chaum and E. van Heyst, "Group signatures," in *Advances in Cryptology - EUROCRYPT 1991*, LNCS 547, pp. 257-265, Springer-Verlag, 1991.
- [15] D. Boneh, X. Boyen, and H. Shacham, "Short group signatures," in *Advances in Cryptology - CRYPTO 2004*, LNCS 3152, pp. 41-55, Springer-Verlag, 2004.
- [16] X. D.Liu, et al., "GSIS: a secure and privacy-preserving protocol for vehicular communications," *IEEE Transactions on Vehicular Technology*, 56(6), pp. 3442-3456, 2007.

# Security Enhanced Authentication Protocol for UHF Passive RFID System

Suna Choi, Sangyeon Lee, Heyungsub Lee

RFID Basic Technology Research Team

Electronics and Telecommunications Research Institute (ETRI)

Deajeon, Korea

sunachoi@etri.re.kr, lseyoun@etri.re.kr, leehs@etri.re.kr

**Abstract**—The passive RFID system was spotlighted as a future technology for automatic identification, but it has a possibility of eavesdropping and leaking of private information. We propose a security enhanced protocol with mutual authentication and data cryptographic mechanism between secure reader and tag for the UHF passive RFID system. We use the OFB-like mode of AES as an effective encryption method. The proposed security enhanced protocol is designed to satisfy the demands of ISO/IEC WD 29167-6, namely the new international standard for the RFID security services. In addition, we present that the proposed security process conforms to the time limitation of the ISO/IEC 18000-6C.

**Keywords**—RFID security; AES; OFB-like mode; mutual authentication.

## I. INTRODUCTION

The RFID is one of the most important technologies which bring enormous benefits in applications where objects have to be identified automatically. It can be applied various applications including supply chain management, product tracing, building access control, public transportation, airport baggage, express parcel logistics and automatic product checkout, etc [1].

The ISO/IEC 18000-6C is the representative RFID standard for the UHF-band passive RFID system. But it does not provide security mechanism between the tag and reader in wireless environment, so the conventional RFID system has a possibility of eavesdropping and leaking of private information.

Each RFID system consists of a tag which is attached to a product for identification and a reader which can access individual data of tags. An unauthorized RFID reader might access to the tag and steal the private contents, and it could be used to trace the movements of a consumer who has a product with an RFID tag. Furthermore, RFID tags can be forged and abused when it is applied services such as proof of origins. The security problems can threaten the development of the RFID system.

The ISO/IEC 18000-6C standard allows the Kill command to protect privacy [2]. But the tag cannot be used anymore after killed, so the applying services can be restricted. As concerns regarding security and privacy issues are raised highly, ISO/IEC JTC 1 SC31 WG7 has been

organized. It is a working group for preparing international standard of the security services and file management of RFID by the classified frequency bands. The ISO/IEC 29167-1 defines the architecture for RFID security framework and security service and the ISO/IEC 29167-6 defines the secured air interface and file management for 860 – 960 MHz UHF band.

Many researchers in the standard group have discussed a cryptographic security system which provides the untraceability, secure communication, authentication, and compatibility with the ISO/IEC 18000-6C standard as the requirements.

There are cryptographic primitives using hash-based methods, symmetric encryption methods, etc. Hash-based methods are conceptually simple and considered a good choice for RFID [3]~[5]. But the symmetric encryption method like Advanced Encryption Standard (AES) is suggested as a better choice on RFID tags from the implementation point of view [6]~[8]. Furthermore, AES is largely accepted in industry and actually mentioned as a strong candidate of the security method for standardization.

The response times and link frequencies of a tag and a reader are specified in the ISO/IEC 18000-6C standard and the security protocol should obey them. But it is hard for a cryptographic process to satisfy the specified time, especially the time from reader transmission to tag response (T1 time), because of the limited resources of the passive tags. In other words, the allowed power of a passive tag is approximately less than several tens of uW. As the operating frequency is higher, the processing time of a cryptographic process can be decreased, however, the consuming power of a tag increases. So the operating frequency is limited (generally less than several MHz).

In this paper, we propose a security enhanced protocol with mutual authentication and cryptographic process which conforms to the ISO/IEC 18000-6C and meets the demands of the ISO/IEC 29167-6. In addition, we present the proposed cryptographic process satisfy the time limitation.

The paper is organized as follows. The key generation method and encryption/decryption process of the proposed cryptographic method are described in Section II. The structure of memory and proposed security enhanced protocol are given in Section III. Then, the discussion and

the simulation results showing that the proposed security process satisfies the time limitation of ISO/IEC 18000-6C are presented in Section IV. Finally, conclusion and further works are followed in Section V.

II. PROPOSED CRYPTOGRAPHIC METHOD

The AES algorithm was chosen in 2001 as an encryption standard. It provides strong security and is well suited for hardware implementation [8]. It operates on a symmetric data block with variable key and block length. The key and block length can be specified to 128, 192, 256 bits. In this paper, we applied the AES algorithm using fixed 128 bit data block and key length.

In order to use a cipher to protect the confidentiality or integrity of messages, the mode of operation of a block cipher must be specified [9]. We apply the modified Output Feedback (OFB) mode of AES, named OFB-like mode. Similar to the OFB mode, the AES engine generates key streams and the messages are encrypted or decrypted by means of bitwise XOR with the generated key streams. Because of the symmetry of the XOR operation, the encryption and decryption processes are technically similar and the extra decryption engine is not required. So, it is appropriated to implement a lightweight secure RFID system.

In the OFB-like mode, however, all messages transmitted between a reader and a tag are considered as a long message and a new session key is generated using the previous key instead of the output of bitwise XOR. It reduces the processing time and enables to satisfy the time limitation specified in the ISO/IEC 18000-6C standard.

A. Session key generation method

Figure 1 shows the generation method of the session keys in the AES OFB-like mode.

The encryption engine that generates a session key is initiated by the first data and the master key. The first data is randomly generated by the reader and the tag during security

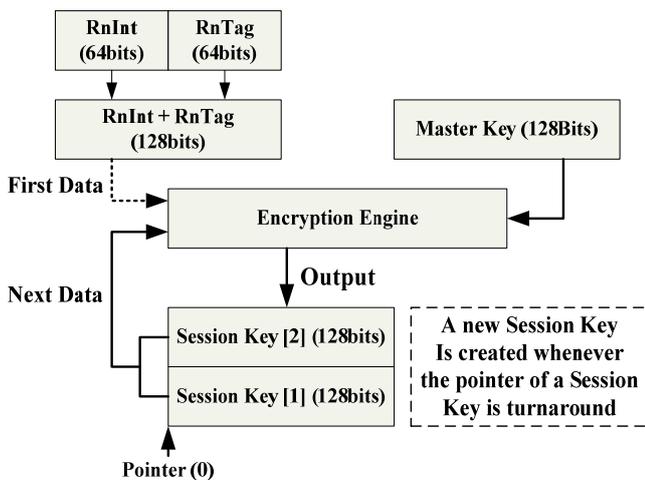


Figure 1. Key generation method in the OFB-like mode

protocols. The RnInt is a random number transmitted from a reader to a tag and the RnTag is a random number transmitted from a tag to a reader. The addition of the RnInt and RnTag becomes the first data. The AES crypto engine generates the first session key and then generates the second session key using the first session key and the master key. The crypto engine takes the previous session key as the next data in every generating routine of a new session key. The crypto engine generates firstly two session keys to prevent the exhaustion of the session key. And then, the crypto engine generates a new session key whenever one session key of the two is exhausted.

B. Encryption and decryption process

Figures 2 and 3 show the proposed encryption and decryption process.

The encryption process is bitwise XOR operations of the plain data and generated session keys. Sequentially, the command and CRC16 is created and added to the encrypted data. The pointer is moved by a bit as the encryption is performed. And the decryption process is similar to the encryption process. The command and CRC16 is checked and removed and only encrypted data takes the XOR operation with the session key.

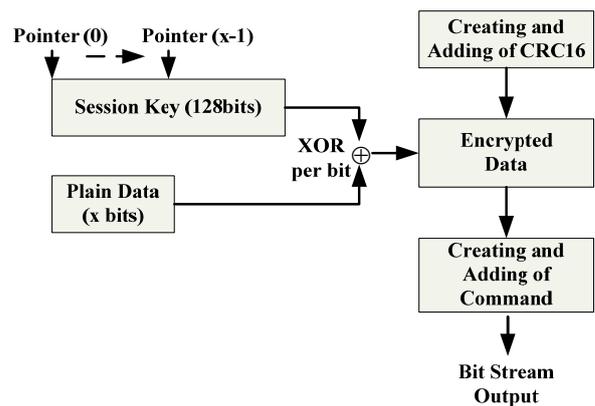


Figure 2. Encryption process

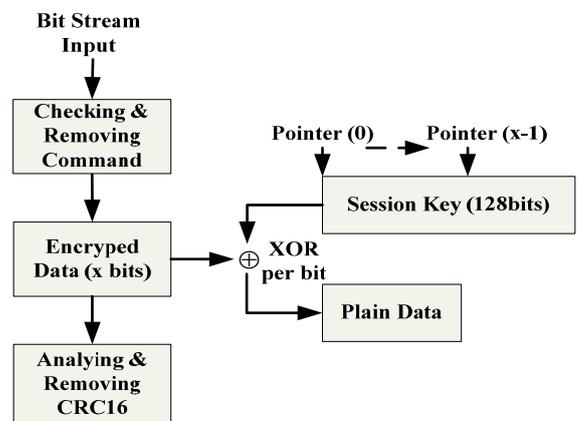


Figure 3. Decryption process

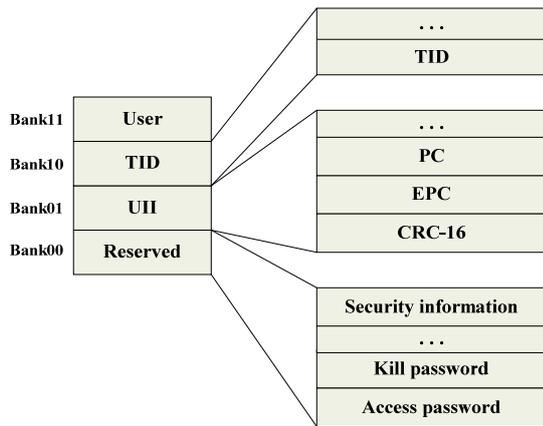


Figure 4. Structure of the RFID tag memory

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
XPC_W1	XEB		MIIM	SA	SS	FS	BM	TC	U	S	RFU	REC[2:0]				
XPC_W2	FMFF				SFF				CSI [3:0] (AES OFB-like)							

Figure 5. XPC in the memory of the secure tag

III. PROPOSED SECURITY ENHANCED PROTOCOL

A. Memory Map of the Secure Tag

As shown in Figure 4, a RFID tag has the memory which is logically separated into four distinct banks. They are Reserved bank, Unique Item Identifier (UII) bank, Tag identification (TID) bank, and User bank.

The External Protocol Control (XPC) of the ISO/IEC 18000-6C contains XPC-W1 and XPC-W2 as presented at Figure 5.

If the Extension bit (XEB) is '0', it means that the Tag does not implement an XPC\_W2. The XEB value of the proposed secure tag is always '1'.

XPC-W1 of the proposed secure tag has U (Untraced) and S (Secure) flags additionally. The U bit indicates whether the corresponding tag supports the untraced function. When the U bit value is '0', it means that the UII which is transmitted to the reader in the first inventory process is fake. When the U bit value is '1', it means that the UII is true. The S bit indicates whether the corresponding tag supports the security function. When the S bit value is '0', it indicates that the corresponding tag operates as a conventional passive tag. When the S bit value is '1', it indicates that the tag operates as a secure tag supporting the security functions.

XPC-W2 contains 4-bits Crypto-graphic suite identifier (CSI). The CSI indicates the kind of cryptographic algorithm used in the secure tag. The default value of CSI is '1', and it means the proposed AES OFB-like mode is used in the RFID system.

The additional security information stored in the tag memory is depicted at Figure 6. SecParam and key index (KI) and master key are contained in this area.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
SecParam	SM	KS	Num of KI [2:0]			RFU										
KI (16*N bits)	Key Index [(16*N-1):(16*N-16)]															
	...															
AES Key (128 bits)	Key Index [15:0]															
	AES Key [127:112]															
...																
AES Key [15:0]																

Figure 6. Security information in the memory of the secure tag

SecParam is composed of SM, KS, Num of KI and RFU. The functions are presented as the followings:

- SM (Security Mode): 1bit SM represents whether the corresponding tag supports the security functions. When the SM bit value is '0', it indicates that the tag operates as a passive RFID tag according to the ISO/IEC 18000-6C standard and when the SM bit value is '1', it indicates that the tag operates as a secure tag supporting the security functions.
- KS (Key Setting): 1bit KS shows whether the Master key is set in the tag. When the KS bit value is '0', it indicates that the Master key is not in the tag and when the SM bit value is '1', it indicates that the Master key is set in the tag.
- Num of KI: It means the number of Key index. Default value is '1', and it means that 1 word (16bit) of key index is assigned and the size of key pool is 2<sup>16</sup>. When the value is '0', it means 8 words is assigned. In this case, the size of key pool is 2<sup>128</sup> as the maximum.
- RFU (Reserved for Future Use): RFU is reserved bits for future use.

For the strong security level, the security RFID reader has a number of AES keys in the key pool. The KI indicates where the AES key is stored in the key pool.

Additionally, 128 bit AES key is stored in the tag memory. The AES key is a private key for generating an output key used for data encryption.

B. Security enhanced protocol

We propose the security enhanced protocol with mutual authentication and data cryptographic process. We assume that the secure reader maintains the database of master keys and key index, and the secure tag and the secure reader have the identical master key.

Figure 7 shows the proposed secure protocol.

First, the secure reader sends Select, Query, or Query Rep commands and then the secure tag transmits a RN16 when the slot counter of the tag is '0'. This selection procedure follows the ISO/IEC 18000-6C standard (step 1~4). When receiving an ACK with respect to the RN16 from the secure reader, a secure tag replies PC, XPC and untraced UII.

Because UII has the information of product where the tag is attached, the proposed secure protocol protects the information of UII from the access of illegal readers using the untraced UII, which is a fake UII composed of random

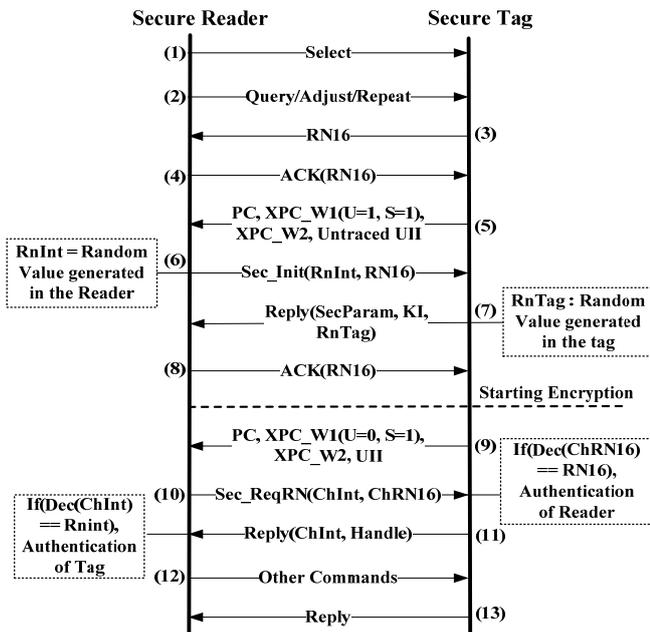


Figure 7. Proposed security protocol

values of the same length with the real UII. The real UII is encrypted and transmitted to the reader in step 9. An attacker can neither get the real UII nor trace the tag (step 5). Subsequently, the secure reader transmits Sec\_Init command. It is defined to initialize the encryption engine by sending the RnInt and require the security information of tag. The secure tag creates the RnTag and starts to generate session key using the RnInt and RnTag as the first data (step6). Thereafter, tag replies SecParam, KI, and RnTag (step7). The secure reader generates the session key using the RnInt and RnTag and sends ACK with RN16 (step8).

Afterwards, the commands and replies between secure reader and tag are transmitted after encrypting with the session key. The tag sends encrypted Protocol Control (PC), XPC, UII as a reply and only secure reader which has the Master Key can decrypt the real UII (step 9). When the received PC and XPC are identified as the appropriate value, the secure reader transmits Sec\_ReqRN command containing ChInt (encrypted data of RnInt) and ChRN16 (encrypted data of RN16) to the secure tag (step10).

In the proposed protocol, the identical PC, XPC, and RN16 are transmitted as both plain data (step 6, 8) and encrypted data (step 9, 10). Even though an attacker eavesdrops on the messages, only some bits of session key are exposed by XOR operation of plane text and encrypted data. The session key is changed continuously and other data are encrypted using the other bits of session key, so no more data is exposed.

The secure tag decrypts the received ChRN16 and checks whether the value is matched. If it is successful, the authentication of the reader is completed. When the reader is considered as an authorized secure reader, the secure tag

replies re-encrypted ChInt and a new 16 bit random number (Handle) (step11). As the session key value is changed continuously, the re-encrypted ChInt has the different values from the prior ChInt in step10. The secure reader decrypts the reply and checks whether the received ChInt is matched to the RnInt. When they are identical, the secure reader determines that the secure tag is authenticated. Through the process, the secure tag and the secure reader can authenticate each other. When the authentication process is failed, the RFID tag goes to the initial state. When the authentication is completed, encrypted access commands including read/write commands and replies can be transmitted between the secure reader and the secure tag (step12~13).

#### IV. SIMULATION AND DISCUSSION

Even if an attacker eavesdrops the whole messages between a reader and a tag, the information of the tag can't be recognized. And movements of a consumer who have the tag are untraced. The proposed secure protocol protects the UII from the access of illegal readers by using the untraced UII and encryption of real UII.

In addition, if an illegal reader attempts to fake as legal reader, it can't perform the mutual authentication process because there is no Master Key. Only authorised reader and tag can execute all access commands including read/write commands. When the authentication process is failed, the RFID tag goes to the initial state.

On the other hand, we compare the operating time of crypto engine with allowed time of the security protocol both in the tag and the reader. The results present the proposed security process satisfies the time limitation of the ISO/IEC 18000-6C standard.

As described in the Section II-A, the crypto engine generates firstly two session keys using the RnInt and RnTag, and then generates a new session key whenever one session key of two session keys is exhausted. As shown in Figure 7, the crypto engine of the tag is initiated and starts to generate first session keys when receiving the Sec\_Init (step6), and should be completed before sending the encrypted PC, XPC, UII (step9). The same procedure is started when receiving the reply and should be completed before sending Sec\_ReqRN in the reader.

The proposed security process is designed following the ISO/IEC 18000-6C standard. The time from reader transmission to tag response is defined as T1 and the time from tag transmission to reader response is defined as T2 in the ISO/IEC 18000-6C standard, and the transmission time of messages is determined by the link frequency.

The simulation condition for analysing the operating time is as follows:

- Link frequency from Reader to tag: 160khz
- Link frequency from tag to reader: 640khz
- Operating frequency of crypto engine: 1.25MHz.

As seen below, the operating time of crypto engine for initialization and generation of first two session keys is less

than the allowed time of the security protocol both in the tag and the reader. The allowed time in the tag and the reader includes the T1 and T2 specified in the ISO/IEC 18000-6C standard.

- Allowed Time in the Tag  
=  $T_1 + T(\text{Reply}) + T_2 + T(\text{Ack}) = 462.875\mu\text{s}$
- Allowed Time in the Reader  
=  $T_2 + T(\text{Ack}) + T_1 + T(\text{PC, XPC, UII})$   
=  $461.3125\mu\text{s}$
- Simulated Time for Initialization and Generation of first two session keys (256bit)  
=  $458\mu\text{s}$

In the proposed security process, a spare session key is always kept in the tag and reader to prevent from exhausting. In case when the message to encrypt is longer than 128bit, a new session key should be generated during the transmission of the message. As seen below, the time for generating a session key is less than the transmission time in the ISO/IEC 18000-6C standard.

- Transmission Time for 128 bit data  
=  $200\mu\text{s}$
- Simulated Time for generating a session key(128bit)  
=  $185\mu\text{s}$

#### V. CONCLUSION AND FURTHER WORKS

The proposed secure protocol conceals the UII using untraced UII to protect the information of product where the tag is attached. Also, it provides the mutual authentication process using the encrypted random values generated in the secure reader and tag and the cryptographic process using the AES crypto engine. Only secure reader and tag which share the same master key can authorize each other and decrypt the encrypted messages. In addition, the proposed secure process improves the operating speed using the AES OFB-like mode. Moreover, the proposed secure protocol and encryption process satisfy the requirements of the ISO/IEC 29167-6 and the compatibility with the ISO/IEC 18000-6C standard.

In the future, we plan to implement the RFID security system applying the proposed security protocol. And more

detailed analysis will be studied against the specific security risks such as relay attack.

#### ACKNOWLEDGMENT

This work was supported by the IT and R&D program of MKE/KEIT [10035239, Development of ultralight low-power RFID secure platform].

#### REFERENCES

- [1] K. Finkenzeller, "RFID Handbook : Fundamentals and Application in Contactless Smart Cards and Identification", 2nd ED, New York :Wiley, 2003
- [2] ISO/IEC 18000-6C standard, "Information Technology-Radio frequency identification for item management – Part 6 : Parameters for air interface communications at 860MHz to 960MHz", July 2007
- [3] H. M. Sun and W. C. Ting, "A Gen2-based RFID authentication protocol for security and privacy", *IEEE transaction on Mpbile Computing*, vol8, no8, pp 1052-1062, 2009
- [4] G. Tsudik, "YA-TRAP: Yet Another Trivial RFID Authentication protocol", *IEEE Annual Conference on Pervasive Computing and Communications*, pp. 640-643, March 2006
- [5] S. A. Weis, S. E. Sarma, R. L. Rivest and D. W. Engels, "Security and privacy aspects of low-cost radio frequency identification systems", *International Conference on Security in Pervasive Computer science*, pp. 201-212, 2004
- [6] M. Feldhofer, C. Rechberger, "A Case against Currently used Hash Functions in RFID Protocols", *OTM 2006 Workshops. LNCS*, vol. 4277, pp. 372-381. Springer, Heidelberg (2006)
- [7] M. Kim, J. Ryou, Y. Choi and S. Jun, "Low-cost Cryptographic Circuits for Authentication in Radio Frequency Identification Systems", *IEEE 10<sup>th</sup> international symposium on Consumer Electronics*, pp. 1-5, 2006
- [8] M. Feldhofer, S. Dominikus, and J. Wolkerstorfer, "Strong authentication for RFID systems using the AES algorithm", in *Proceeding Workshop on Cryptographic Hardware and Embedded System*, vol. 3156, pp. 357-370, 2004
- [9] J. Daemanm, V. Rijmen, "The design of Rijndael: AES- the Advanced Encryption Standard", Springer, 2002
- [10] A. Juels, "RFID security and privacy : a research survey" *IEEE Journal on Selected Areas in Communications*, pp. 381-394, Feb, 2006