IEEE 2012 GLOBECOM
GLOBAL COMMUNICATIONS CONFERENCE, EXHIBITION & INDUSTRY FORUM

Issue 4
Thursday, 6 December 2012

Conference Addresses

IEEE GLOBECOM 2012 began Wednesday's comprehensive agenda with the eagerly awaited keynote addresses of Vinton Cerf, the current Vice President & Chief Internet Evangelist of Google and celebrated computer scientist most commonly known as one of "fathers of the Internet" as well as Krish Prabhu, President of AT&T Labs and Chief Technology Officer.

After brief comments by IEEE GLOBECOM 2012 Executive Chair Pierre Perra, who thanked all the conference's patrons and exhibitors for this year's sponsorship and support, Dr. Cerf shared his views on "Internet Challenges 2012-2020" and the policy and engineering problems underlying the next stage of growth. This began by noting the difference between engineering and policy, the latter of which is "driven by culture and attitude, not computation."

According to Dr. Cerf, the Internet has witnessed an "astonishing evolution" that now includes 2.4 billion users with penetration rates reaching nearly 80 percent in North America, 63 percent in Europe and approximately 30 percent in Asia. However, despite the latest and next wave of smartphone and mobile home innovations, he spoke at length about the numerous challenges currently confronting the Internet. According to Dr. Cerf, the success of global business will be reflected in the understanding of the Internet's diverse nature. This includes adapting IPv6 protocols since the last of the IPv4 addresses were essentially depleted in 2011 as well as more vigorously adopting domain name security extensions designed to ensure the integrity of users.

Stating that the Internet is "one of the most effective tools of the democratization of speech," Dr. Cerf also expressed his concerns about the governments that are currently "threatened by individuals speaking their mind" and their attempts "to control what people say and do" on the Internet. As examples, he referenced the recent "Arab Spring" uprising and the current attempt of some of the officials representing 193 countries at the Dubia Internet conference hosted by the International Telecommunications Union to adopt regulations that will complicate commerce and further cyber-crackdowns.

After this, he spoke of the need for anonymity and ability to speak freely, but also expressed his concerns about the legality of digital signatures and the enforcement of online contracts used in the legal frameworks of e-commerce transactions. In addition, Dr. Cerf expressed the need for "everyone one of us to become access control managers" and for providers to more actively develop system processes that minimize the infections, viruses, worms and Trojan horses can have on our lives and the machines we use daily. Another mandate cited in his speech was the implementation of protocols connecting the various sets of clouds, which as of today "do not know about any other cloud but their own."

In conclusion, Dr. Cerf also talked about the creation of interplanetary networks that will one day link earth to other planets, while admitting that the Internet will "one day be replaced by something else." Although, he "doesn't know what that something will be." As for now, he emphasized that "things that work persist" and that Internet interoperability is essential for providing the ability to talk to billion of users and machines worldwide without interference.

Immediately following this presentation, Krish Prabhu, President of AT&T Labs and the company's Chief Technology Officer spoke about "Connected Life: the Future of Global Communications' and AT&T's vision of the future, which even he admitted "could be wrong despite the logic behind it."

Driven by the evolution of smartphone technologies and its ongoing ability to overcome the challenges of distance, reach and touch, AT&T alone now services 100+ million users in 25 countries, producing $127 billion in revenues. As for the company's vision, it includes connecting life no matter where you are to provide a personalized experience at all times. At the heart of this mission is the access to "ubiquitous broadband connectivity that integrates seamlessly into people's lives and delivers services over an open network no matter who runs it."

In the future, added Prabhu, these capabilities will be enhanced by advances in multiple frontiers. First, "4G LTE will be everywhere by the decade" offering even better latency, higher data speeds and performance. Second, the deploy of small cell clusters will help overcome the challenges associated with the phenomenal wireless data traffic growth that grew by 20,000 percent over the past five years. Next, industry and users alike will benefit from the pervasiveness of cloud services that will ensure customization, offer end-to-end solutions exploiting multiple components and accessibility through handsets "regardless of where you are."

And finally, Prabhu spoke at-length about the services of the future driven by the next class of emerging, empowered devices. This includes sensor networks operated through voice recognition capabilities that will personalize the delivery of content on the go or in the home through open, secure and safe cloud Internet services.

TABLE OF CONTENTS

PROGRAM SPOTLIGHT

Keynote Session
Thursday, 6 December 2012 • 08:00 – 09:30
Center Ballroom/North Ballroom A/B

Stephen B. Alexander

Senior Vice President, Products and Technology & CTO, Ciena

The Performance-on-Demand Application Ecosystem: The Next Phase of Telecom Infrastructure

Today's communication networks are in an intense state of evolution.The explosion of high-bandwidth consumer and enterprise applications as well as the exponential increase in the number of network-connected devices has created a critical need to tightly and seamlessly align computing, storage and connectivity resources together in a single, converged platform that can deliver any service or application when the user demands it.

As a result, the network is taking a more specific and valuable role than ever before. While computing and storage resources have been virtualized for some time, the ability to connect these functions with high-performance network resources that also are virtualized represents the next phase of network modernization: the performance-on-demand application ecosystem.

To achieve this, operators must think of their network not so much in terms of the traditional geographic domains, but as more flexible functional domains that connect users to content and also connect content to content, regardless of location or device. Creating that network means building a programmable infrastructure – with exponential scale at lower cost and software-enabled orchestration and automation – on which applications run to deliver a high-value and superior user experience. This transforms the economics and utility of networks – turning it into highly strategic business asset for value creation, not just a necessary evil in the race to deliver the next "killer app."

This talk will discuss this next phase of network modernization, ways that operators can achieve it and what it means to the future of telecommunications.

Hossein Eslambolchi

Chairman & CEO
2020 Venture Partners

The Power of Technology to Transform the Future

In this talk, Dr. Eslambolchi will present how the power of technology will transform the future. He will describe what he views on the top technology trends over the next several years. He will discuss how these technologies will transform the world and our daily lives. He will present the evolution and future direction of Information Communication integral to every aspect of society: Digital Government, Commerce, Health Care, Education, Security, Military, and Arts & Culture. Dr. Eslambolchi will explore the resulting challenges and research opportunities for academia, industry, business, and society.

Dr. Eslambolchi will also cut through digital clutter and describe what he views as the top technology trends over the next two decade. He will also show how these major trends are already impacting the communications industry in 21st century.

PROGRAM SPOTLIGHT

IF24: Executive Forum: IT Transformation: Clouds, Security, Mobility and Computing

Thursday, 6 December 2012 • 10:00 – 12:00
South Ballroom A

Sponsored by Making Light Work Lighter General Photonics

IEEE CLOUD COMPUTING

According to the latest market research report, over one third of the world's 7 billion population are now active Internet users. New business drivers such as cloud computing, virtualization, security, mobility and business social networking are driving the increased consumerization of IT services in much faster and cheaper manner. Clouds, big data and analytics, workforce mobility are among the key priorities for many large corporates. In this panel, the executives will explore the opportunities and challenges of IT and network transformation and provide insight on how communication will change the way we work, live, play and learn.

Invited Guest Speakers:

Flavio Bonomi, VP & Head, Advanced Architecture and Research Organization, Cisco Systems, USA (Keynote)
The Internet of Everything and the Historic Convergence of Networking, Computing, and Data Management

Steven Gray, CTO, CSR, USA
Location-as-a-Service (LaaS): a Platform for Location Services

Kaushik Arunagiri, Managing Partner, Communications Media & Entertainment, EMC Global Services at EMC, USA
Transform IT, Transform Business

Steve Alexander, SVP & CTO, Ciena, USA
Data Center without Walls

Mahbubul Alam, Head of IoT/M2M, ITS and Auto Mfg, Connected Industries Group, Cisco, USA
Forces Changing our World!


Flavio Bonomi


Steven D. Gray


Kaushik Arunagiri


Stephen B. Alexander


Mahbubul Alam

Only 5 Slots Available at the Lightning Talks Session
Sign up by Noon Today!

The lightning talk session will be held today from 16:00 – 18:00 in North Ballroom A. It is concluding the Industry Forums portion of the conference. ***Lightning talks are short five minute talks on technical topics.*** Any conference related subject can be presented (thoughts triggered by a presentation, a nifty algorithm trick, a thesis project, open source software project, company product, etc.). As of Wednesday noon we have 15 short talks signed up for out of 20 slots. Five slots are still available. If you would like to present a talk sign-up before noon today. Everyone is encouraged to come and listen to this fun, dynamic and interesting session.

Rules:

1. ***Sign up at conference registration before Thursday at noon.***
Speaking slots assigned in order of sign up.
2. Speakers must be present at start of session or slot is forfeited to the next speaker signed up.
3. Each speaker is permitted five minutes to speak
a. Use from zero to three slides.
b. Please no animation on the slides.
c. Use of URLs within the presentations is encouraged.
4. The five minute time limit on talks will be strictly enforced.
Speakers should be prepared to present a concise talk.
5. Email slides to ***reder@ieee.org*** following the session, if you desire them to be posted on the conference web site.

IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS
INDUSTRY FORUM & EXHIBITION

IEEE ICC 2013

Bridging the Broadband Divide
9-13 June 2013 • Budapest, Hungary

IEEE

IEEE COMMUNICATIONS SOCIETY

IEEE ICC 2013 covers the entire range of communications technologies, offering in-depth information on the latest developments in voice, data, image and multimedia.

Technical Symposia
focus on technological trends in recent communication research and development from academia to the industrial laboratories throughout the world.

Tutorials and Workshops
address emerging technical and business issues in communications technologies.

Industry Forums
feature high-level executives addressing challenges, opportunities and the future of the Industry.

Exhibition
showcases the latest technologies, applications and services.

IEEE ICC 2013 PATRONS

T · · Mobile

ERICSSON

Learn more about IEEE ICC 2013 at www.ieee-icc.org.

EVENTS OF THE DAY

08:00 – 09:30
KEYNOTE SESSION
Stephen B. Alexander, Ciena
Hossein Eslambolchi, 2020 Venture Partners
Center Ballroom/North Ballroom A/B

09:30 – 10:00
COFFEE BREAK / South Exhibit Hall
Prize Drawing (must be present to win)

10:00 – 12:00
INDUSTRY FORUMS
IF24: Executive Forum: IT Transformation: Clouds, Security, Mobility and Computing / South Ballroom A
IF25: Cable Industry Access Technology / North Ballroom A

TECHNICAL SESSIONS
AHSN07: Network Connectivity & Relay Selection / North Exhibit Hall A
AHSN16: MAC & Channel Modeling / North Exhibit Hall B
CISS07: Wireless Network Security I / North Exhibit Hall C
CogRN07: Multiuser Access / North Exhibit Hall D
CQ07: Resource Allocation / Magic Kingdom Ballroom 4
CSSM07: Multimedia Application & Service / North Exhibit Hall F
CT06: Cooperative Communications I / North Exhibit Hall G
NGNI07: Traffic Engineering & Congestion Control / North Exhibit Hall H
ONS04: Grooming, RWA, Dimensioning & Survivability / North Exhibit Hall I
SAC-ASN3: Energy Efficiency in Access Networks / Castle A
SAC-GNCS7: Green Cognitive Radio & Energy Harvesting / Castle B
SAC-SSC02: Satellite & Space Communications / Monorail A
SPC09: Wireless Systems & Networks / Monorail B
WC22: LTE / Monorail C
WC23: Wireless Security / Magic Kingdom Ballroom 1
WC24: OFDM II / Magic Kingdom Ballroom 2
WN13: Cognitive Networking / Castle C
WN14: Network Coding / North Exhibit Hall J

12:00 – 13:30
LUNCH / On Your Own

13:30 – 15:30
INDUSTRY FORUMS
IF26: Industry Perspective on Cloud Computing / South Ballroom A
IF27: Technology to Exploit Big Data for Disaster Management / North Exhibit Hall E
I**F28: Optical Wireless Access** / North Ballroom A

13:30 – 15:30 (continued)
TECHNICAL SESSIONS
AHSN08: Power Control & Neighbor Discovery / North Exhibit Hall A
AHSN17: Security & Monitoring / North Exhibit Hall B
CISS08: Wireless Network Security II / North Exhibit Hall C
CogRN08: Spectrum Access / North Exhibit Hall D
CogRN11: Waveform Design & Modulation / Magic Kingdom Ballroom 1
CQ08: Network Design & Management / North Exhibit Hall G
CT07: Cooperative Communications II / North Exhibit Hall F
NGNI08: Topics in Next Generation Networking & Internet I / North Exhibit Hall H
SAC-GNCS8 / Castle A
SAC-PL 2: Power Line Communications & SmartGrid II / North Exhibit Hall J
SPC10: Beamforming / Castle C
WC25: Spectrum Sensing & Cognitive Networks / Monorail A
WC26: Optimization in Wireless Systems / Monorail B
WC27: SC-FDMA & SC-FDE / Monorail C
WN15: Wireless Relay Networks / North Exhibit Hall I
WN16: Routing & Multicasting / Castle B

15:30 – 16:00
COFFEE BREAK / South Exhibit Hall
Prize Drawing (must be present to win)

16:00 – 18:00
INDUSTRY FORUMS
IF29: Cloud Computing Industry Perspectives / South Ballroom A
IF23: Lightning Talks / North Ballroom A

TECHNICAL SESSIONS
AHSN09: Topology Control & Repair / North Exhibit Hall A
AHSN18: Management & Throughput Modeling / North Exhibit Hall B
CISS09: Wireless Network Security III / North Exhibit Hall C
CogRN10: Spectrum Sensing II / North Exhibit Hall D
CQ09: Robustness in Communication Networks / Monorail B
CT09: MIMO System Design & Analysis / North Exhibit Hall F
NGNI09: Routing / North Exhibit Hall G
NGNI10: Topics in Next Generation Networking & Internet II / North Exhibit Hall H
SAC-GNCS9: Green Communications under Quality of Service Constraints / North Exhibit Hall J
SPC11: Signal Processing for Communications III / Castle B
WC28: Wireless Channels & Propagation / Castle C
WC29: Interference Management II / Monorail A
WC30: Random Access Networks / Castle A
WN17: Network Performance Optimization / North Exhibit Hall I

PROGRAM UPDATES

The following are updates to the program guide found in your badge holder. These updates appear in the online final program.

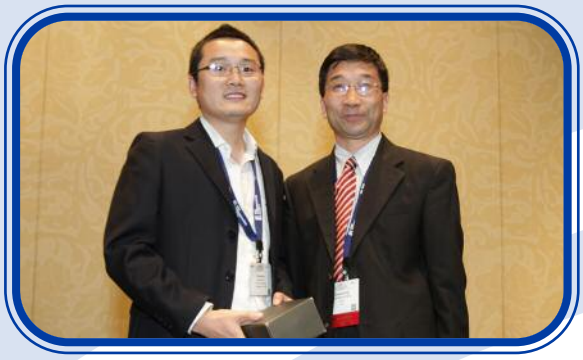
Thursday, 6 December 2012
IF25: Cable Industry Access Technology
from 16:00 – 18:00 will now be held in North Ballroom A.
IF28: Optical Wireless Access
from 13:30 – 15:30 will now be held in North Ballroom A.
IF23: Lightning Talks
from 16:00 – 18:00 will now be held in North Ballroom A.

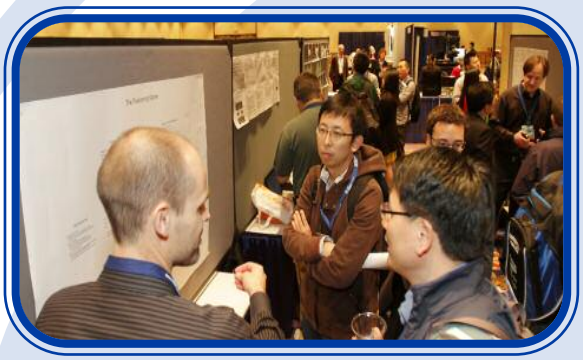
Friday, 7 December 2012
T9: Opportunistic Communication
from 09:00 – 12:00 will now be held in Magic Kingdom Ballroom 1/4.
T12: Cooperative Spectrum Sensing
from 14:00 – 17:00 will now be held in Magic Kingdom Ballroom 1/4.

EXHIBIT HALL

Final Two Prize Drawings
Prize drawings will be held today during the morning and afternoon coffee breaks in the Exhibit Hall.
One prize per attendee. Must be present to win.

From the Exhibit Floor…





CALL FOR PAPERS AND PROPOSALS

The 2013 IEEE GLOBECOM will be held in Atlanta, Georgia in December 2013. This flagship conference of the IEEE Communication Society will feature a comprehensive technical program including numerous symposia, tutorials, workshops and an industrial forum program featuring prominent keynote speakers, technology and industry forums and vendor exhibits. We invite you to submit original technical papers for presentation at the conference, and for publication in the proceedings and in IEEE Xplore®. Proposals for tutorials, workshops and industry forums are also invited. Full details of the submission procedures are available at http://www.ieee-globecom.org/2013.

TECHNICAL SYMPOSIA

— IEEE GLOBECOM 2013 will feature the following technical symposia. —

Selected Areas in Communications Symposium

Data Storage Track
Edward Au, Huawei, China

e-Health Track
Kaoru Sezaki, University of Tokyo, Japan

Internet of Things Track
Latif Ladid, University of Luxembourg

Game Theory for Communications Track
Jianwei Huang, Chinese University of Hong Kong

Power-Line Communications Track
Haniph Latchman, University of Florida, USA

Satellite and Space Communications Track
Igor Bisio, University of Genoa, Italy

Access Networks and Systems Track
Filippo Cugini, CNIT, Italy

Green Communication Systems and Networks
Stefano Bregni, Politecnico di Milano, Italy

Social Networks Track
Kwang-Cheng Chen, National Taiwan University, Taiwan

Optical Networks and Systems Symposium

S. J. Ben Yoo, University of California Davis, USA
Eiji Oki, University of Electro-Communications, Japan

Ad Hoc and Sensor Networking Symposium

Mohamed Younis, University of Maryland Baltimore County, USA
Yu Cheng, Illinois Institute of Technology, USA
Falko Dressler, University of Innsbruck, Austria
Xu Li, INRIA, France

Cognitive Radio and Networks Symposium

Andrea Giorgetti, University of Bologna, Italy
Dusit (Tao) Niyato, Nanyang Technological University, Singapore
Richard Yu, Carleton University, Canada

Communication and Information Systems Security Symposium

Yi Qian, University of Nebraska Lincoln, USA
Yunghsiang Sam Han, National Taiwan University of Science and Technology, Taiwan
Loukas Lazos, University of Arizona, USA

Communications QoS, Reliability and Modeling Symposium

Xavier Masip-Bruin, Technical University of Catalonia, Spain
Stefano Giordano, University of Pisa, Italy
Hideaki Yoshino, NTT, Japan

Communication Software, Services, and Multimedia Application Symposium

Vincent Wong, University of British Columbia, Canada
Liang Zhou, Nanjing University of Posts & Telecommunications, China

Communication Theory Symposium

Nallanthan Arumugam, King's College London, UK
Sang Wu Kim, Iowa State University, USA
Lian Zhao, Ryerson University, Canada

Next-Generation Networking Symposium

Mooi Choo Chuah, Lehigh University, USA
Ying-Dar Lin, National Chiao Tung University, Taiwan
Qian Zhang, Hong Kong University of Science & Technology, China

Signal Processing for Communications Symposium

Huaiyu Dai, North Carolina State University, USA
Octavia Dobre, Memorial University, Canada
Tao Jiang, Huazhong University of Science & Technology, China

Wireless Communications Symposium

David Love, Purdue University, USA
Hai Lin, Osaka Prefecture University, Japan
Claude Oestges, University Catholique de Louvain, Belgium
Liuqing Yang, Colorado State University, USA

Wireless Networking Symposium

Shiwen Mao, Auburn University, USA
Lin Cai, University of Victoria, Canada
Nei Kato, Tohoku University, Japan

INDUSTRIAL FORUM AND EXHIBITION PROGRAM

Submissions are sought for Industry Forums on the latest technical and business issues in communications and networking topics. Submit your proposals to the IF&E Chair Heath Thompson (Heath.Thompson@landisgyr.com).

TUTORIALS

Proposals are invited for half- or full-day tutorials on communications & networking topics. Proposals will be submitted in PDF via EDAS. For questions please contact the Tutorial Co-Chairs: Robert Schober (rschober@ece.ubc.ca) or Matteo Cesana (cesana@elet.polimi.it).

WORKSHOPS

Proposals are invited for half- or full-day workshops on the latest technical and business issues in communications and networking. Please direct questions to the Workshops Co-Chairs: Kwang-Cheng Chen (chenkc@cc.ee.ntu.edu.tw) or Matthieu Bloch (matthieu.bloch@ece.gatech.edu).

ORGANIZING COMMITTEE

General Chair
Branko Bjelajac

Executive Chair
Greg Durgin

Executive Vice Chair
Ruben Salazar

IF&E Chair
Heath Thompson

GITC Advisor
Gerhard P. Fettweis

GIMS Advisor
Richard Miller

Technical Program Chair
John Barry

Awards Chair
Gordon Stüber

Technical Program Vice-Chair
Tommaso Melodia

Technical Program Vice-Chair
Matthew Valenti

Tutorials Co-Chairs
Robert Schober
Matteo Cesana

Workshops Co-Chairs
Kwang-Cheng Chen
Matthieu Bloch

Publications Co-Chairs
Xiaoli Ma
Henk Wymeersch

Student Travel Grant Chair
Yimin Zheng

Panels Chair
Geoffrey Li

Industry Content Chair
Anuj Batra
Bill Lichtensteiger

ComSoc Project Manager
June Leach-Barnaby

To be published in the IEEE GLOBECOM 2013 Conference Proceedings and IEEE Xplore®, an author of an accepted paper is required to register for the conference at the full or limited (member or non-member) rate and the paper must be presented by an author of that paper at the conference unless the TPC Chair grants permission for a substitute presenter. Non-refundable registration fees must be paid prior to uploading the final IEEE formatted, publication-ready version of the paper. For authors with multiple accepted papers, one full or limited registration is valid for up to 3 papers. Accepted and presented papers will be published in the IEEE GLOBECOM 2013 Conference Proceedings and IEEE Xplore®.

IMPORTANT DATES

Workshop Proposal	Industry Forum Proposal	Tutorial Proposal	Paper Submission	Paper Acceptance	Camera-Ready Paper
15 December 2012	15 March 2013	15 March 2013	15 March 2013	1 July 2013	1 August 2013

"Dialog with Industry Leaders" Provided Lively Give-and-take with Audience Members

Tuesday's IEEE GLOBECOM 2012 activities ended with the highly-anticipated "Dialog with Industry Leaders" featuring a lively question-and-answer exchange between leading communications executives and conference attendees.

After a brief introduction by Industry Forum & Exhibition Chair Narisa Chu, the session's co-moderators Hamid Ahmadi of Samsung Information Technology America and Glenn Wellbrook of Verizon provided each of the evening's panelists with the opportunity to discuss their backgrounds and highlight the ways the communications industry is rapidly advancing with new applications and services. Joe Berthold of Ciena Corporation responded by citing the network's evolution into a content-centric service that will eventually lead to the ubiquitous embrace of a new array of cloud computing processes, while Mahbubul Alam of Cisco spoke of the "Internet of Everything" and his company's ongoing mission to build services around connected devices that range from people to data to process and then to cloud services.

When asked about the current state of the Internet and the upper bounds of its limitations, Alam also addressed the shift to a new architectural paradigm that will expand limitations by bringing the cloud closer to the edge and end devices. However, Wellbrook offered a different view that includes featuring self regulation as a method for decreasing the flow of unwanted information over the Internet as Dean Sirovica of Huawei Technologies seconded the notion by citing the need for a one cent fee on all emails.

After this discussion, the panelists were then asked about speeding product introduction cycles. Wellbrook responded by speaking about the need to drive standards that support universal hardware platforms that can easily be updated with new software, while Sirovica stated that "the best way to increase cycle times is through competition."

The question then arose about job opportunities in the communications industry. After noting that "we're all at the mercy of economic cycles that the whole world goes through," Sirovica told the Ph.D student that there will always be a market for talented, well-educated professionals with Alam adding the caveat that he "will always need to keep current throughout life" and Berthold stating the need to "stay flexible, stay curious and work hard."

During the following series of questions highlighting the next wave of architectural changes, the panelists spoke at-length about LTE and its widespread ability to connect the world over the coming decade. This includes providing different ways to service more customers at lower costs with less maintenance and even helping to provide improved mobility and data rates to rural communities that do not have access to fiber and coaxial connections.

Finally, in the session's closing moments the panel responded to the problem of spam again by reinforcing the need to filter unwanted information through economic solutions that will "make the market take care of the problem itself."

YESTERDAY'S NEWS

Wednesday Morning Executive Forum Offered New Paradigms for Network Architectures

On Wednesday following the morning's keynote sessions, Joe Berthold, Network Architecture of CIENA Corporation began the Executive Forum on Data Infrastructure and Services with his presentation on "Content-centric Network Architecture." While bandwidth keeps growing and operators continue their efforts to connect people, places and machines through the use of tens of thousands of servers, they are also increasingly facing the dilemma of "traffics that are growing faster than revenues."

During his presentation, Berthold identified several fundamental approaches for reducing network costs and the unit cost of equipment. This included the "Building of Networks for the Future" that operate under "design, deploy and react" paradigms, minimize complex locations and simplify forwarding and system physical convergence to lower infrastructure costs. Other measures cited include the optimization of networks across layers and the virtualization of network functions running differing treatments.

Following these comments, Christoph Glingener of Adva highlighted the intricacies of Software Defined Optical Networks SD(O)N and its ability to enable new network paradigms and the simple management of receivers, applications and services with open, standardized and flexible programmability. Afterwards, Wupen Yuen of NeoPhotonics continued the topic discussion with his presentation on "Software-Defined Optical Networks Enabled by Photonic Integrated Circuits" and their ability to lower both costs and power consumption.

As the panel's fourth speaker, Hamid Ahmadi of Samsung provided the session's keynote on "Network Trends in Mobile Computing." In his introductions, Ahmadi noted that "infrastructure and traffic were not only the issues" confronting the industry, other solutions are currently confronting the needs of "the changing mobile experience" driven by ever-evolving applications especially in the M2M and Internet of Things.

As a result, he said that "the exponential growth of data usage ranging from tablets to smart tvs can no longer be satisfied by today's wireless infrastructures. In addition, LTE can't do it alone and you can't just keep building new base stations." To achieve "wireless connectivity everywhere" and "ubiquitous always access" that supports 100-fold traffic growth, we must adopt next generation cellular ecosystems utilizing all-IP flat networks and seamless indoor/outdoor operations.

IEEE GLOBECOM 2012 BEST PAPERS

On the following pages, the 4 of 15 best papers featured are from Symposia on Wireless Communications and Wireless Networking.

A Proportional Fair Radio Resource Allocation for Heterogeneous Cellular Networks with Relays

Qian (Clara) Li, Rose Qingyang Hu
Department of Electrical
and Computer Engineering
Utah State University, Logan, UT, USA
Emails: liqian@ieee.org, rosehu@ieee.org

Yi Qian
Department of Computer
and Electronics Engineering
University of Nebraska-Lincoln, NE, USA
Email: yqian@ieee.org

Geng Wu
Wireless Technology Division
Intel Corporation, OR, USA
Email: geng.wu@intel.com

***Abstract*—As a key technology in 4G-LTE, heterogeneous networks can effectively extend the coverage and capacity of wireless networks by deploying multiple micro-nodes on top of the conventional macro base stations (BS). The deployed micro-nodes differ in transmission power and processing capabilities, leading to new challenges in interference management, mobile association, and radio resource management (RRM). In this paper, we consider RRM for heterogeneous networks with relays (RN) where the RNs have full RRM capabilities and can be viewed as micro BSs. A radio resource allocation framework is proposed with the objective to ensure proportional fairness among the UEs. An asymptotically optimal solution is derived by applying the gradient-based scheduling scheme and the Karush-Kuhn-Tucker (KKT) conditions for optimality. To implement RRM in networks with RNs, the resource consumption in the backhaul links, which depends on the demand of the UEs associated with the RN, should be counted at both the BS and the RN. The derived resource allocation scheme gives insight on the optimal radio resource allocation for heterogeneous networks with RNs.**

I. Introduction

With the rapid development of wireless communication devices and applications, the demand on communication network capacity is ever increasing. To meet these demands, in 4G LTE-advanced (LTE-A) networks, heterogeneous networks have been proposed to extend network capacity and coverage [1]. Unlike the conventional homogeneous cellular networks where BSs with similar communication capabilities are deployed, in a heterogeneous network, macro and micro BSs with various transmission powers, coverage and processing capacities coexist. The deployment of micro BSs introduces spatial diversity and cell splittings, which could effectively improve network spectrum efficiency and coverage. Coming along with its benefits, this new form of network architecture introduces new challenges in interference management, mobile association and resource allocation. In this paper, we investigate radio resource allocation for downlink communications in heterogeneous cellular networks with high transmit power BS and low transmit power relays (RN).

Radio resource management (RRM) is a key design issue to ensure efficient use of wireless spectrum resources. For homogeneous networks, this issue has been well investigated with a variety of radio resource allocation schemes [2]-[5]. To ensure fair resource allocation, proportional fair algorithm was proposed [6] and been implemented in Qualcomm's HDR system [7]. For heterogeneous networks, however, due to the difference in network coverage, network node transmit power, and total available radio resources, the scheduling algorithms developed in homogeneous network cannot be directly applied. Moreover, with in-band RNs, the resource required on the wireless backhaul transmission between the BS and the RN should also be counted, which further complicates the the resource allocation problem. Despite the intensive attentions on heterogeneous networks, only a small number of papers have investigated the issue of RRM for such networks [10]. Related works have been done on RRM of relay-aided cellular networks, such as [11]-[13]. However, most of these previous works assume RNs with simple forwarding functions but no ability in buffering and scheduling. By these limitations in RN function, once receiving from the BS, the RNs should immediately forward the received signal to the corresponding UEs in the next subframe. As a result, the data rate of the two-hop link is limited by the compositing link with a lower transmission capability, leading to an inefficient use of the radio resources. To make a more efficient usage of the radio resources, RNs with full RRM functionalities has been adopted in LTE-A standards, where the RNs essentially behave as mini BSs.

In this paper, we propose a radio resource allocation framework with proportional fair considerations for OFDMA-based heterogeneous cellular networks with RNs of full RRM capability. A low-complexity and asymptotically optimal solution is derived using a gradient-based scheduling scheme and the Karush-Kuhn-Tucker (KKT) conditions for optimality. The derived resource allocation solution provides a guideline on resource allocation for heterogeneous networks with RNs. It also helps to gain a better understanding on the impact of backhaul resource consumption on network performance.

II. Heterogeneous Network Model

We consider downlink communication of a cellular network with one BS and multiple RNs deployed in each cell. An example of the considered network is shown in Fig. 1. Each cell is divided into M_c sectors with N_r RNs uniformly deployed in each sector. The total number of sectors in the network is N_c. UEs are uniformly distributed in the network

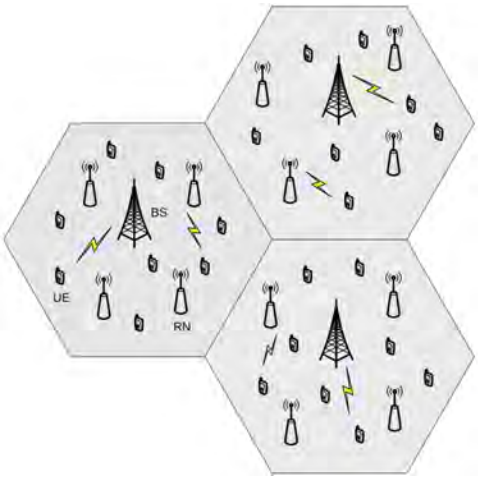


Fig. 1. Network model of a heterogeneous network with relays

with an average of N_u UEs in each sector. Communication between the BS and the UEs can be achieved via direct transmission from the BS or via the forwarding of the RNs. The deployment of RNs introduces smaller cells on top of the conventional cellular system, which effectively expands the cellular network coverage and capacity.

OFDM is used as the physical layer transmission scheme. The total frequency band is equally divided into F resource blocks (RB), with each UE being assigned with an integer number of RBs. The fading channels between the network nodes and the UEs are frequency-selective across different RBs while the channels are frequency flat within the same RB. We denote the frequency-domain channel gain in the fth RB and the tth time instance between the ith BS and the kth UE as $h^f_{k,0,i}(t)$, and between the jth RN in the ith cell and the kth UE as $h^f_{k,j,i}(t)$. The channel gain counts both long-term path loss and shadowing and short-term fading due to multipath and mobility. The received signal-to-interference-noise-ratio (SINR) of the UEs at time tth can be evaluated as follows.

$$\mathrm{SINR}^f_{k,0,i}(t) = \frac{P_b|h^f_{k,0,i}(t)|^2}{\sum_{i'\neq i}|h^f_{k,0,i'}(t)|^2P_b + \sum_{i=1}^{N_c}\sum_{j=1}^{N_r}|h^f_{k,j,i}(t)|^2P_r + N_0}, \quad (1)$$

and

$$\mathrm{SINR}^f_{k,j,i}(t) = \frac{P_r|h^f_{k,j,i}(t)|^2}{\sum_{i=1}^{N_c}|h^f_{k,0,i}(t)|^2P_b + \sum_{i=1}^{N_c}\sum_{j'\neq j}|h^f_{k,j',i}(t)|^2P_r + N_0}, \quad (2)$$

where P_b and P_r are the transmit powers of the BS and the RN, respectively, and N_0 is the variance of the additive white Gaussian noise.

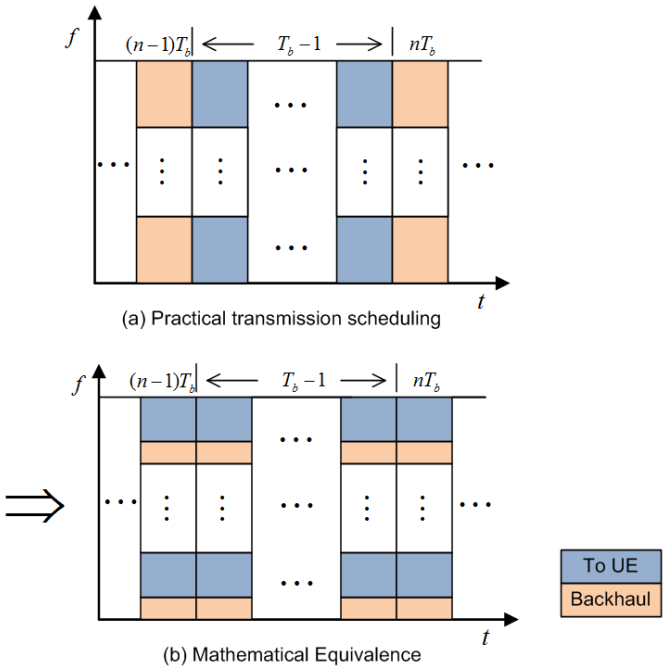


Fig. 2. Illustration of the time-duplex transmission scheduling for the direct/access links and the backhaul links

Based on the SINR values, at the tth subframe, the instantaneous data rate in terms of bit/s/Hz for the kth UE received from the jth RN in the ith cell at the fth RB can be calculated using Shannon formula as

$$R^f_{k,j,i}(t) = \log\left(1 + \mathrm{SINR}^f_{k,j,i}(t)\right). \quad (3)$$

For in-band RNs, the backhaul communication link between the BS and the RN shares the same radio resource as the direct link from the BS to the UE and the access link from the RN to the UE. In practice, communication in the backhaul link and in the direct and access links can be scheduled in a time-duplex fashion as demonstrated in Fig. 2(a), where a subframe is the scheduling interval considered in this paper. Communication in the backhaul link takes place every T_b subframes of direct and access links communication. During the backhaul transmission slots, RNs receive from their donor BSs. The RNs store the received information in its buffer and forward to the corresponding UEs in the appropriate time/frequency resources. In this paper, we assume high-capacity and constant-quality backhaul links. This assumption can be justified by the fact that RNs are usually equipped with multiple antennas and placed in locations with the low shadowing. Denoting $R^f_{0,j,i}$ as the data rate of the backhaul link between the ith BS and the jth RN in the fth RB and tth subframe, the frequency resource required at the backhaul link to support the kth UE with each $R^f_{k,j,i}$ bit/s data transmission in the access link can be calculated as

$$a^f_{k,j,i}(t) = \frac{R^f_{k,j,i}(t)}{R^f_{0,j,i}}. \quad (4)$$

For high-capacity backhauls, we have $a^f_{k,j,i}(t) < 1$.

III. Problem Formulation

To optimize network long-term spectrum efficiency and ensure fairness among the UEs, we should 1) properly associate each UE with a BS or RN, and 2) properly allocate the frequency resource to the UEs at each scheduling interval. Usually mobile association can be done on a longer time scale based on the large scale pass loss and shadowing while resource allocation should be done at each scheduling cycle to address more dynamic conditions such as traffic burstiness and fast fading channel.

In the conventional homogeneous networks, best-power based mobile association scheme is often applied [8], where the kth UE is associated with the best node $(i^*, j^*)_k$.

$$(i^*, j^*)_k = \arg \max_{i \in \{1,\cdots,N_c\}, j \in \{0,1,\cdots,N_r\}} (P_{k,j,i} |\bar{h}_{k,j,i}|^2), \quad (5)$$

where $P_{k,j,i}$ is the corresponding node transmission power and $\bar{h}_{k,j,i}$ is the path loss coefficient of the channel between the kth UE and the jth RN in the ith sector. For slow varying network environment, $\bar{h}_{k,j,i}$ can be regarded as constant over the whole frequency band and in a long time period. In the heterogeneous networks, due to the transmit power disparity between the BS and the RN, most of the UEs will be associated with the BSs if the best-power based association scheme is used. The RN utilization will be low and the advantage of using RN in improving the spectrum efficiency and coverage of the network could not be fully exploited. To balance the traffic load between the BSs and the RNs, we apply range-expansion based association scheme, which uses a bias to compensate the power difference between BSs and RNs [9], so that more UEs can be associated with RNs. In this range-expansion based mobile association, the kth UE is associated with the best node $(i^*, j^*)_k$

$$(i^*, j^*)_k = \arg \max_{i \in \{1,\cdots,N_c\}, j \in \{0,1,\cdots,N_r\}} (|\bar{h}_{k,j,i}|^2 / \delta_{i,j}), \quad (6)$$

where $\delta_{i,0} = 1$ and $1 < \delta_{i,j} < (P_b/P_r)$, for $j > 0$. We use a decision variable $x_{k,0,i}$ to indicate the association status between the kth UE and the BS in ith sector. Specifically,

$$x_{k,0,i} = \begin{cases} 1 & \text{if } k\text{th UE is associated with } i\text{th BS} \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

The decision variable $x_{k,j,i}$ is similarly defined. We assume that each UE can only associate with at most one BS or RN. The $x_{k,j,i}$ values satisfy

$$\sum_{i=1}^{N_c} \sum_{j=0}^{N_r} x_{k,j,i} \leq 1, \text{ for } k = 1, \cdots, N_u. \quad (8)$$

We term the UEs associated with the BSs and RNs as B-UEs and R-UEs, respectively, and denote $\mathcal{K}_{0,i}$ as the set of B-UEs associated with the ith BS and $\mathcal{K}_{j,i}$ as the set of R-UEs associated with the jth RN in the ith sector.

Based on a given mobile association scheme, radio resource scheduling assigns the UEs with proper radio resources at each subframe. For systems with RNs that support in-band backhauls, radio resource management needs to take into account resources consumed in the backhaul link as well as the direct and access link. To facilitate mathematical analysis and to gain more insights on the impact of backhaul resource consumption on the system RRM, instead of counting the backhaul resource consumption in a subframe basis as shown in Fig. 2(a), we virtually distribute the backhaul resource consumption into each RB as shown in Fig. 2(b). In order to formulate the mathematical problem, each of BS's RB can be considered to be shared by downlink transmission towards one of its associated B-UEs in the direct link and backhaul transmission towards its associated RNs. The portion of the radio resource spent in the backhaul link is determined by the amount of information transmitted from the RNs to their associated R-UEs in the counterpart RB. Similarly, each of RN's RB will be shared by downlink transmission towards one of its associated R-UEs and reception from its attached BS in the backhaul link. The portion of the RB required by the backhaul link at the RN is determined by the amount of information transmitted to the served R-UE in the RB. We denote $n^f_{k,0,i}(t)$ as the portion of the fth RB assigned to the kth UE at the ith BS in the tth subframe, $n^f_{k,j,i}(t)$ as the portion of the fth RB assigned to the kth UE at the jth RN in the ith sector and tth subframe, and $n^{b,f}_{k,j,i}(t)$ as the portion of the fth RB assigned to the backhaul link between the jth RN and the ith BS in support of the kth UE. $n^{b,f}_{k,j,i}(t)$ can be calculated as

$$n^{b,f}_{k,j,i}(t) = a^f_{k,j,i}(t) n^f_{k,j,i}(t). \quad (9)$$

To achieve a good balance between spectrum efficiency and fairness, we use proportional fairness as the performance metric. The scheduling problem at each subframe t, can be formulated as follows.

$$[\mathbf{P}_1] \qquad \max U(\mathbf{R}(t)) = \sum_{k=1}^{N_u} \log(R_k(t)) \quad (10)$$

subject to

$$\sum_{k=1}^{N_u} x_{k,0,i} n^f_{k,0,i}(t) + \sum_{j=1}^{N_r} \sum_{k=1}^{N_u} x_{k,j,i} n^{b,f}_{k,j,i}(t) \leq 1, \quad \text{for } i = 1, \cdots, N_c, f = 1, \cdots, F \quad (11)$$

$$\sum_{k=1}^{N_u} x_{k,j,i} n^f_{k,j,i}(t) + \sum_{k=1}^{N_u} x_{k,j,i} n^{b,f}_{k,j,i}(t) \leq 1, \quad \text{for } i = 1, \cdots, N_c, \ j = 1, \cdots, N_r, f = 1, \cdots, F. \quad (12)$$

$$n^f_{k,j,i}(t) \geq 0, \ \forall i, j, k, f \quad (13)$$

where

$$R_k(t) = \frac{1}{T_c} \sum_{\tau = t - T_c + 1}^{t} S_k(\tau), \quad (14)$$

where T_c is the size of the time window for averaging, and

$$S_k(\tau) = \sum_{f=1}^{F} \sum_{i=1}^{N_c} \sum_{j=1}^{N_r} x_{k,j,i} R^f_{k,j,i}(\tau) n^f_{k,j,i}(\tau). \quad (15)$$

In $\mathbf{P}_1$, the objective is to find less importantly the optimal $n^f_{k,j,i}$ values but more importantly the scheduling outcome, i.e., the UE index for each RB, to maximize the time averaged log-scaled system throughput. Constraint (11) is the resource constraint for each RB at the BS. The first term in (11) computes the portion of RB f used by the direct link and the second term in (11) computes the portion of RB f used by the backhaul link. Similarly, constraint (12) gives the resource constraint for each RB at the RN.

As a multicarrier proportional fair scheduling problem, the computational complexity in finding the optimal solution of $\mathbf{P}_1$ is prohibitively high [14]. To fit for practical implementation, we apply the gradient-based scheduling algorithm as proposed in [15], [16]. It was proven in [16] that the gradient-based scheduling algorithm asymptotically converges to the optimal solution. In the next section, based on the gradient-based scheduling algorithm, we show how to optimally allocate resources in such a heterogeneous networks with in-band RNs.

IV. An Asymptotically Optimal Radio Resource Scheduling Scheme

Using the gradient-based scheduling framework, the system parameters are chosen to maximize the drift of the objective function at each subframe, given as

$$\begin{aligned} & U(\mathbf{R}(t+1)) - U(\mathbf{R}(t)) \\ &= \sum_{k=1}^{N_u} \Bigg(\log\Big(R_k(t) + \epsilon\big(S_k(t+1) - S_k(t-T_c+1)\big)\Big) \\ &\quad - \log\big(R_k(t)\big)\Bigg) \\ &= \sum_{k=1}^{N_u} \frac{1}{R_k(t)} S_k(t+1)\epsilon - \sum_{k=1}^{N_u} \frac{1}{R_k(t)} S_k(t-T_c+1)\epsilon \\ &\quad + O(\epsilon^2), \end{aligned} \tag{16}$$

where $\epsilon = 1/T_c$ and the second equality is obtained using first order Taylor expansion. Since only the first term in (16) depends on future decisions, the gradient-based scheduling problem can be formulated as

$$[\mathbf{P}_2] \max_{n^f_{k,j,i}(t)} \sum_{k=1}^{N_u} \frac{1}{R_k(t-1)} \sum_{f=1}^{F} \sum_{i=1}^{N_c} \sum_{j=1}^{N_r} x_{k,j,i} R^f_{k,j,i}(t) n^f_{k,j,i}(t), \tag{17}$$

subject to the constraints given in (11)-(13). By independence of summation and since the constraints in (11)-(13) are set in a per RB basis, the optimal solution of $\mathbf{P}_2$ can be found by solving the optimal $n^f_{k,j,i}(t)$ values for each RB using the following optimization formulation.

$$[\mathbf{P}_3] \max_{n^f_{k,j,i}(t)} \sum_{k=1}^{N_u} \sum_{i=1}^{N_c} \sum_{j=1}^{N_r} \frac{1}{R_k(t-1)} x_{k,j,i} R^f_{k,j,i}(t) n^f_{k,j,i}(t), \tag{18}$$

subject to (11)-(13). By gradient-based scheduling, multicarrier proportional fair scheduling can be decomposed into multiple single-carrier scheduling problems.

Note that $\mathbf{P}_3$ is a convex optimization problem, the KKT conditions is necessary and sufficient for optimality. The optimal solution can be solved from the KKT conditions given as follows.

$$\sum_{k=1}^{N_u} x_{k,0,i} n^f_{k,0,i}(t) + \sum_{j=1}^{N_r} \sum_{k=1}^{N_u} x_{k,j,i} n^{b,f}_{k,j,i}(t) \le 1, \quad \text{for } i = 1, \cdots, N_c, f = 1, \cdots, F, \tag{19}$$

$$\sum_{k=1}^{N_u} x_{k,j,i} n^f_{k,j,i}(t) + \sum_{k=1}^{N_u} x_{k,j,i} n^{b,f}_{k,j,i}(t) \le 1 \quad \text{for } i = 1, \cdots, N_c, j = 1, \cdots, N_r, f = 1, \cdots, F, \tag{20}$$

$$-n^f_{k,j,i}(t) \le 0, \forall i, j, k \tag{21}$$

$$\lambda^f_i(t) \ge 0, \forall i \tag{22}$$

$$\mu^f_{j,i}(t) \ge 0, \forall j \tag{23}$$

$$\nu^f_{k,j,i}(t) \ge 0, \forall i, j, k \tag{24}$$

$$\lambda^f_i(t)\Big(\sum_{k=1}^{N_u} x_{k,0,i} n^f_{k,0,i}(t) + \sum_{j=1}^{N_r} \sum_{k=1}^{N_u} x_{k,j,i} a^f_{k,j,i}(t) n^f_{k,j,i}(t) - 1\Big) = 0 \tag{25}$$

$$\mu^f_{j,i}(t)\Big(\sum_{k=1}^{N_u} x_{k,j,i} n^f_{k,j,i}(t) + \sum_{k=1}^{N_u} x_{k,j,i} a^f_{k,j,i}(t) n^f_{k,j,i}(t) - 1\Big) = 0 \tag{26}$$

$$\nu^f_{k,j,i}(t) n^f_{k,j,i} = 0, \forall i, j, k \tag{27}$$

$$\nabla_{n^f_{k,0,i}} L(n^f_{k,j,i}(t), \lambda^f_i(t), \mu^f_{j,i}(t), \nu^f_{k,j,i}(t)) = 0 \tag{28}$$

$$\nabla_{n^f_{k,j,i}} L(n^f_{k,j,i}(t), \lambda^f_i(t), \mu^f_{j,i}(t), \nu^f_{k,j,i}(t)) = 0, \tag{29}$$

where $\lambda^f_i(t)$, $\mu^f_{j,i}(t)$, and $\nu^f_{k,j,i}(t)$ are the Lagrangian multipliers. $L(n^f_{k,j,i}(t), \lambda^f_i(t), \mu^f_{j,i}(t), \nu^f_{k,j,i}(t))$, the Lagrangian of $\mathbf{P}_3$, is given as

$$\begin{aligned} & L(n^f_{k,j,i}(t), \lambda^f_i(t), \mu^f_{j,i}(t), \nu^f_{k,j,i}(t)) \\ &= -\sum_{k=1}^{N_u} \sum_{i=1}^{N_c} \sum_{j=0}^{N_r} \frac{1}{R_k(t-1)} x_{k,j,i} R^f_{k,j,i}(t) n^f_{k,j,i}(t) \\ &\quad + \sum_{i=1}^{N_c} \lambda^f_i \Big(\sum_{k=1}^{N_u} x_{k,0,i} n^f_{k,0,i}(t) \\ &\quad + \sum_{j=1}^{N_r} \sum_{k=1}^{N_u} x_{k,j,i} a^f_{k,j,i}(t) n^f_{k,j,i}(t) - 1\Big) \\ &\quad + \sum_{i=1}^{N_c} \sum_{j=1}^{N_r} \mu^f_{j,i} \Big(\sum_{k=1}^{N_u} x_{k,j,i} n^f_{k,j,i}(t) \\ &\quad + \sum_{k=1}^{N_u} x_{k,j,i} a^f_{k,j,i}(t) n^f_{k,j,i}(t) - 1\Big) \end{aligned}$$

$$-\sum_{i=1}^{N_c}\sum_{j=0}^{N_r}\sum_{k=1}^{N_u}\nu_{k,j,i}^{f}n_{k,j,i}^{f}(t). \quad (30)$$

From (28) and (29), we have

$$-\frac{1}{R_k(t-1)}x_{k,0,i}R_{k,0,i}^{f}(t)+\lambda_i^f x_{k,0,i}-\nu_{k,0,i}^{f}(t)=0, \quad (31)$$

and

$$-\frac{1}{R_k(t-1)}x_{k,j,i}R_{k,j,i}^{f}(t)+\lambda_i^f x_{k,j,i}a_{k,j,i}^{f}(t) +\mu_{j,i}^{f}x_{k,j,i}(1+a_{k,j,i}^{f}(t))-\nu_{k,j,i}^{f}(t)=0. \quad (32)$$

From (31), λ_i^f can be found as

$$\lambda_i^f=\frac{1}{R_k(t-1)}x_{k,0,i}R_{k,0,i}^{f}(t)+\nu_{k,0,i}^{f}(t). \quad (33)$$

For all the $k\in\mathcal{K}_i$, the value of $\nu_{k,0,i}^{f}(t)$ should be chosen to ensure the UEs with different $R_{k,0,i}^{f}(t)$ value achieve the same λ_i^f, i.e.,

$$\nu_{k,0,i}^{f}(t)=\lambda_i^f-\frac{x_{k,0,i}R_{k,0,i}^{f}(t)}{R_k(t-1)}. \quad (34)$$

Considering the KKT condition (27), for the UEs with $x_{k,0,i}n_{k,0,i}^{f}>0$, the corresponding $\nu_{k,0,i}^{f}(t)$ value should be chosen as 0. However, if we choose more than one UE with $x_{k,0,i}n_{k,0,i}^{f}>0$, according to (33), a unique λ_i^f value cannot be guaranteed. Therefore, for each RB of the BS at subframe t, at most one UE can be served, whose corresponding $\nu_{k,0,i}^{f}$ value will be set as 0. For all the other UEs, $n_{k,0,i}=0$ and the corresponding $\nu_{k,0,i}^{f}$ is calculated using (34). Since the KKT condition (24) requires $\nu_{k,0,i}^{f}\geq 0$, the serving UE should be chosen to maximize λ_i^f value, i.e.,

$$k_{0,i}^{f*}=\arg\max_{k\in\mathcal{K}_{0,i}}\frac{R_{k,0,i}^{f}(t)}{R_k(t-1)}. \quad (35)$$

The optimal value of λ_i^f is thus given as

$$\lambda_i^{f*}(t)=\max_{k\in\mathcal{K}_{0,i}}\frac{R_{k,0,i}^{f}(t)}{R_k(t-1)}. \quad (36)$$

The optimal $\nu_{k,0,i}^{f}$ values are

$$\nu_{k,0,i}^{f*}(t)=\begin{cases}0 & \text{if } k=k_{0,i}^{f*}\\ -\frac{R_{k,0,i}^{f}(t)}{R_k(t-1)}+\max_k\frac{R_{k,0,i}^{f}(t)}{R_k(t-1)} & \text{if } k\in\mathcal{K}_{0,i}, k\neq k_{0,i}^{f*}.\end{cases} \quad (37)$$

From (32), we have that

$$\mu_{j,i}^{f}(t)=\frac{1}{1+a_{k,j,i}^{f}(t)}\left(\frac{R_{k,j,i}^{f}(t)}{R_k(t-1)}-\lambda_i^f(t)a_{k,j,i}^{f}(t)+\nu_{k,j,i}^{f}(t)\right). \quad (38)$$

Similar to the above analysis, the $\nu_{k,j,i}^{f}(t)$ value should be properly chosen to compensate for the difference caused by the different $R_{k,j,i}^{f}(t)$ values and ensure a same $\mu_{j,i}^{f}(t)$ value over all $k\in\mathcal{K}_{j,i}$. Moreover, according to conditions (27), when $n_{k,j,i}^{f}(t)\neq 0$, we should have $\nu_{k,j,i}^{f}(t)=0$. Therefore, only one of the $\nu_{k,j,i}^{f}(t)$s' can be chosen as zero and the corresponding UE will be assigned with RB f. All the other $\nu_{k,j,i}^{f}(t)$s' will be chosen to ensure a unique $\mu_{j,i}^{f}(t)$ value, and the corresponding $n_{k,j,i}^{f}(t)$s' are set as zero. Since condition (24) require $\nu_{k,j,i}^{f}\geq 0$. The value of $\mu_{j,i}^{f}(t)$ should be chosen as

$$\mu_{j,i}^{f*}(t)=\left(\max_{k\in\mathcal{K}_{j,i}}\frac{1}{1+a_{k,j,i}^{f}(t)}\left(\frac{R_{k,j,i}^{f}(t)}{R_k(t-1)}-\lambda_i^{f*}(t)a_{k,j,i}^{f}(t)\right)\right)^{+}, \quad (39)$$

where $(x)^{+}=\max\{0,x\}$. Denoting

$$\rho_{j,i,A}^{f}(t)=\frac{1}{1+a_{k_{j,i}^{f*},j,i}^{f}(t)}\frac{R_{k_{j,i}^{f*},j,i}^{f}(t)}{R_{k_{j,i}^{f*}}(t-1)}, \quad (40)$$

and

$$\rho_{j,i,B}^{f}(t)=\frac{a_{k_{j,i}^{*},j,i}^{f}(t)}{1+a_{k_{j,i}^{*},j,i}^{f}(t)}\lambda_i^{f*}(t)=\frac{a_{k_{j,i}^{*},j,i}^{f}(t)}{1+a_{k_{j,i}^{*},j,i}^{f}(t)}\frac{R_{k_{0,i}^{f*},0,i}^{f}(t)}{R_{k_{0,i}^{f*}}(t-1)}, \quad (41)$$

where

$$k_{j,i}^{f*}=\arg\max_{k\in\mathcal{K}_{j,i}}\frac{1}{1+a_{k,j,i}^{f}(t)}\left(\frac{R_{k,j,i}^{f}(t)}{R_k(t-1)}-\lambda_i^{f*}(t)a_{k,j,i}^{f}(t)\right), \quad (42)$$

from (39), we know that $\mu_{j,i}^{f*}=0$ when

$$\rho_{j,i,A}^{f}(t)<\rho_{j,i,B}^{f}(t). \quad (43)$$

The case with (43) corresponds to a scenario where the gain in proportional fairness value obtained by supporting the R-UE is not high enough to compensate for the loss in proportional fairness value at the B-UE due to backhaul resource consumption. In this case, we have

$$\nu_{k,j,i}^{f*}(t)=\lambda_i^{f*}(t)a_{k,j,i}^{f}(t)-\frac{R_{k,j,i}^{f}(t)}{R_k(t-1)}>0, \quad (44)$$

and $n_{k,j,i}^{f*}(t)=0$. The jth RN will not serve any of its R-UEs. Otherwise, for the case with $\rho_{j,i,A}^{f}(t)\geq\rho_{j,i,B}^{f}(t)$, the $k_{j,i}^{f*}(t)$th R-UE can be served by the jth RN in the fth RB at the tth subframe. The value of $\nu_{k,j,i}^{f}(t)$ is chosen as

$$\nu_{k,j,i}^{f*}(t)=\begin{cases}0 & \text{if } k=k_{j,i}^{f*}\\ (1+a_{k,j,i}^{f}(t))\mu_{j,i}^{f*}(t)-\frac{R_{k,j,i}^{f}(t)}{R_k(t-1)}+\lambda_i^{f*}(t)a_{k,j,i}^{f}(t) & \text{if } k\in\mathcal{K}_{j,i}\\ & k\neq k_{j,i}^{f*},\end{cases} \quad (45)$$

and $n_{k^{f*},j,i}^{f}(t)$ can be found as

$$n_{k_{j,i}^{f*},j,i}^{f}(t)=\frac{1}{1+a_{k_{j,i}^{f*},j,i}^{f}(t)}. \quad (46)$$

Based on the above optimization results, the value of $n^f_{k^{f*},0,i}(t)$ can be found as

$$n^f_{k^{f*}_{0,i},0,i}(t) = \Big(1 - \sum_{j\in\mathcal{J}_i} a^f_{k^{f*}_{j,i},j,i}(t) n^f_{k^{f*}_{j,i},j,i}(t)\Big)^+, \quad (47)$$

where $\mathcal{J}_i$ is the set of RNs in the ith sector.

So far, we have obtained the optimal solution of the optimization problem $\mathbf{P}_1$, which is formulated based on the virtual scheduling scheme shown in Fig. 2(b). The solution of $\mathbf{P}_1$ can be translated back for the original time-division based scheduling scheme demonstrated in Fig. 2(a). We summarize the optimization solution of $\mathbf{P}_1$ in the following proposition.

Proposition 1: For the tth subframe of direct/access link transmission in heterogeneous relay networks, with proportional fairness considerations, it is asymptotically optimal to allocate radio resources using the following strategy. At the ith BS, the fth RB is assigned to the B-UE with index

$$k^{f*}_{0,i} = \arg\max_{k\in\mathcal{K}_{0,i}} \frac{R^f_{k,0,i}(t)}{R_k(t-1)}. \quad (48)$$

At the jth RN in the ith sector,

1) when $\rho^f_{j,i,A}(t) \geq \rho^f_{j,i,B}(t)$, the fth RB is assigned to the R-UE with index

$$k^{f*}_{j,i} = \arg\max_{k\in\mathcal{K}_{j,i}} \frac{1}{1+a^f_{k,j,i}(t)}\left(\frac{R^f_{k,j,i}(t)}{R_k(t-1)} - \lambda^{f*}_i(t) a^f_{k,j,i}(t)\right), \quad (49)$$

where $\lambda^{f*}_i(t)$ is given in (36), and

2) when $\rho^f_{j,i,A}(t) < \rho^f_{j,i,B}(t)$, no R-UEs will be served in the fth RB.

Here, the values of $\rho^f_{j,i,A}(t)$ and $\rho^f_{j,i,B}(t)$ are given in (40) and (41), respectively.

V. Numerical Results

We simulate the performance of the proposed radio resource scheduling scheme in a cellular network with a 19-cell 3-sector three-ring hexagonal cell structure. Four RNs are uniformly deployed in each sector. Simulation setup follows the guidelines for Case 1 described in the 3GPP technical reports [1]. The simulated multipath channel model is chosen to be the extended typical urban (ETU) model. Transmit power of the BS is 46dBm (40W) and transmit power of the RN is 30dBm (1W). The UEs are uniformly distributed in the network. The UEs are traveling at a speed of 3 km/h. The total bandwidth is 10MHz with 180 kHz for each frequency resource block (RB). The whole frequency band is divided into 50 RBs.

In Fig. 3, the proposed resource allocation scheme is simulated under different bias values in mobile association. An average number of 50 UEs are distributed within each sector. The backhaul can support a data rate of 5 b/s/Hz. From the simulation results, it can be seen that the value of the objective function achieves its maximum at $\delta = 8$ dB. Note that $\delta = 0$ dB corresponds to path loss-based association and $\delta = 16$ dB corresponding to best power-based association. This result demonstrates the importance of a proper range-extension based association in improving the spectrum efficiency in heterogeneous networks. Best power-based association under-utilizes the RNs while path loss-based association overly extends the coverage range of the RNs, resulting in either a high burden on the BSs or a high interference to the R-UEs. A proper range extension-based association exploits the resources of the RNs in sharing the network traffic and also watches out the interference level of the R-UEs.

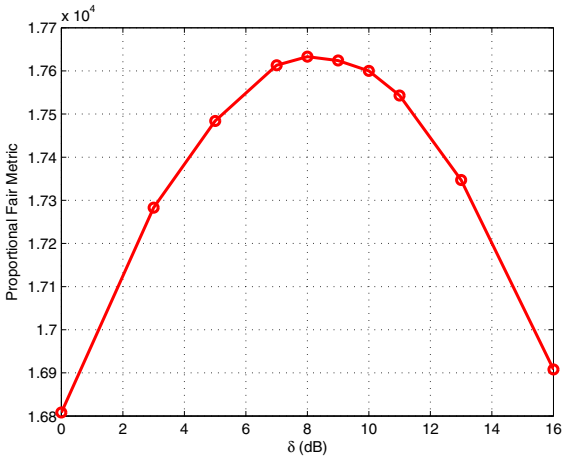


Fig. 3. The value of the objective function with respect to different mobile association biases.

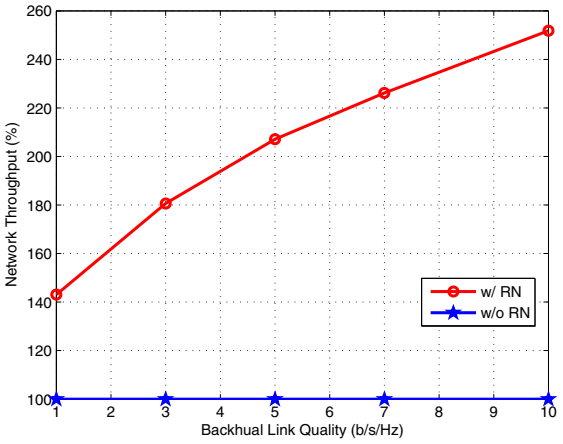


Fig. 4. Network throughput with respect to backhaul link quality

In Fig. 4, we evaluate the network throughput using the proposed resource allocation scheme under different backhaul link qualities. The average number of UEs in each sector is 50 and $\delta = 8$ dB. As a comparison, we also simulate the network throughput for homogeneous networks without RNs. We express the network throughput of the heterogeneous network as the relative percentage of the throughput of the non-relay network. It can be seen that the throughput of the heterogeneous network improves as the backhaul link quality improves. In the case with a 10 b/s/Hz data transmission capability on the backhaul, a 252% capacity gain can be

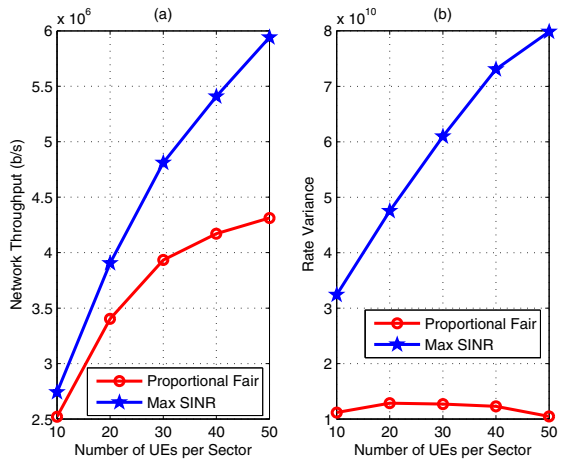


Fig. 5. Comparison in (a) network throughput and (b) rate variance between proportional fair and max SINR-based resource allocation schemes

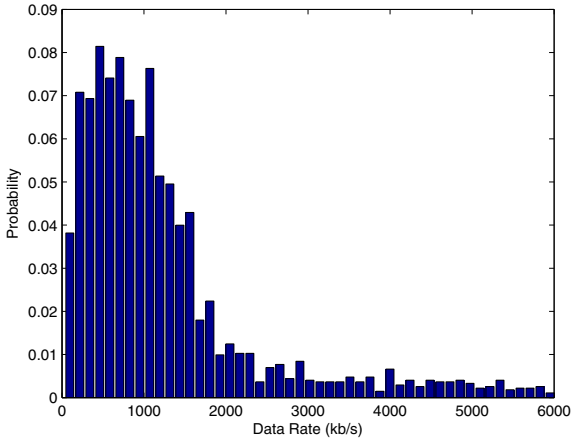


Fig. 6. UE rate distribution

obtained by the network with RNs. In a heterogeneous network with in-band RNs, the better the backhaul link quality, the lower the cost of supporting the RN transmission, and thus the more salient the advantages of using RNs.

In Fig. 5, we compare the proportional-fair based resource allocation scheme with the max-SINR based scheme, where each RB is allocated to the UE with the highest received SINR. The figures are plotted with respect to the average number of UEs in each sector. The mobile association bias value is $\delta = 8$ dB. From Fig. 5(a), we can see that since max SINR-based scheme fully exploits the multiuser diversity, it yields a better network throughput than the proportional fair-based scheme. This throughput gain comes at the cost of fairness among the UE, as can be observed from Fig. 5(b), where the variance of the UE rates are plotted. Max SINR-based scheme prefers the UEs with better channel condition and results in a large derivation among the rate of the UEs. Proportional fair-based resource allocation takes care of the disadvantageous UEs and achieves a good tradeoff between throughput and fairness. The rate distribution of the UEs for the proportional fair-based scheme is shown in Fig. 6. It can be seen that most of the UEs achieve a data rate in the range 100kbps to 1500kbps while a small portion of UEs with good geometrical achieve a better data rate. The rate distribution chart further demonstrates the effectiveness of the proportional-fair based scheme.

VI. Conclusion

In this paper, we investigated the radio resource allocation for OFDMA-based heterogeneous networks with RNs by considering proportional fairness as the performance metric. Based on the gradient-based scheduling framework and KKT conditions for optimality, an asymptotically optimal resource allocation scheme is derived. Our derived resource allocation scheme evaluated the impact of in-band backhaul on radio resource allocation in a RN network. The study provides the guidelines on how to incorporate backhaul radio resource consumption into network radio resource management.

References

[1] 3GPP TR36.814, "Further advancements for E-UTRA physical layer aspects," v9.0.0, Mar. 2010.

[2] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-Part I: Chunk Allocation," *IEEE Trans. Commun.*, vol. 57, pp. 2734-2744, Sept. 2009.

[3] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems-Part II: Joint chunk, power and bit allocation," *IEEE Trans. Commun.*, vol. 60, pp. 499-509, Feb. 2012.

[4] G. Song and Y. G. Li, "Cross-layer optimization for OFDM wireless networks-Part I: Theoretical framework," *IEEE Trans. Wireless Commun.*, vol. 4, pp. 614-624, Mar. 2005.

[5] S. Sadr, A. Anpalagan, and K. Raahemifar, "Radio resource allocation algorithms for the downlink of multiuser OFDM communication systems," *IEEE Commun. Surveys Tutorials*, vol. 11, pp. 92-106, 3rd Quarter 2009.

[6] F. P. Kelly, A. K. Maulloo and D. K. H. Tan, "Rate control in communication networks: shadow prices, proportional fairness and stability," *Journal of Operational Research Society*, vol. 49, pp. 237-252, Mar. 1998.

[7] A. Jalili, R. Padovani, and R. Pankai, "Data throughput of CDMA HDR a high Efficient high data rate personal communication wireless system" in *Proc. IEEE VTC 2000 Spring*, pp. 1854-1858, Tokoy, May 2000.

[8] S. Kyuho, C. Song, and G. Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, pp. 3566-3576, Jul. 2009.

[9] Qualcomm Europe, "Range expansion for efficient support of heterogeneous networks," TSG-RAN WG1 #54bis R1-083813, Sept. 2008.

[10] M. Salem, A. Adinoyi, M. Rahman, H. Yanikomeroglu, D. Falconer, Y. Kim, E. Kim, and Y. Cheong, "An overview of raido resource management in relay-enhanced OFDMA-based networks," *IEEE Commun. surveys and tutorials*, vol. 12, pp. 422-438, Q3 2010.

[11] Z. Tang and G. Wei, "Resource allocation with fairness consideration in OFDMA-based relay networks," in *Proc. IEEE WCNC2009*, pp. 1-5, Budapest, Apr. 2009.

[12] T. C. Ng and W. Yu, "Joint optimization of relay strategies and resource allocations in cooperative cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 25, pp. 328-339, Feb. 2007.

[13] Ö. Oyman, "Opportunistic scheduling and spectrum reuse in relay-based cellular networks," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 1074-1085, Mar. 2010.

[14] H. Kim and Y. Han, "A proportional fair scheduling for multicarrier transmission systems," *IEEE Commun. Letters*, vol. 9, pp. 210-212, Mar. 2005.

[15] R. Agrawal and V. Subramanian, "Optimality of certain channel aware scheduling policies," *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, 2002.

[16] A. L. Stolyar, "On the asymptotic optimality of the gradient scheduling algorithm for multiuser throughput allocation," *Operations Research*, vol. 53, no. 1, pp. 12-25, 2005.

Performance Analysis of Cooperative ADHOC MAC for Vehicular Networks

Sailesh Bharati[†] and Weihua Zhuang[‡]
Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1
E-mail: [†]sbharati@uwaterloo.ca and [‡]wzhuang@bbcr.uwaterloo.ca

***Abstract*—The paradigm of vehicular ad-hoc networks (VANETs) emerges as a promising approach to provide road safety, vehicle traffic management, and infotainment applications. Thus, it is important to develop a VANET medium access control (MAC) protocol that provides an efficient and reliable delivery of packets for diverse applications. Cooperative communication, on the other hand, can enhance the reliability of communication links in VANETs, thus mitigating wireless channel impairments due to a poor channel condition. Recently, a cooperative scheme for MAC in VANETs based on time-division multiple access, referred to as Cooperative ADHOC MAC (CAH-MAC), has been proposed [1]. CAH-MAC is an efficient protocol capable of increasing the network throughput by reducing the wastage of time slots. In CAH-MAC, neighboring nodes cooperate by utilizing the unreserved time slots, for retransmission of a packet which failed to reach its target receiver due to a poor channel condition. In this paper, we study the reliability of CAH-MAC in terms of packet transmission delay (PTD) and packet dropping rate (PDR). Through mathematical analysis and computer simulation, we show that CAH-MAC provides reliable communication by decreasing the PTD and PDR as compared with existing approaches.**

I. Introduction

Increasing road accidents and user demands for a drive-thru Internet connection have led to the evolution of intelligent transportation systems [2] and other applications that improve road safety, increase transportation efficiency, and provide on-board infotainment services. To make these applications possible, vehicles can be equipped with sensors and communication devices to form a communication network called vehicular ad-hoc network (VANET). In a VANET, a smart vehicle uses advance sensors for gathering information and wireless medium for exchanging the information with other vehicles. Such vehicles are equipped with an on-board unit (OBU) and/or one or multiple application units (AUs) [3]. An OBU is a device with a wireless networking interface which enables vehicles to communicate. AUs, on the other hand, are devices which run application(s) and make use of OBUs to exchange information with other vehicles. Vehicles communicate independently either with each other or with stationary wireless stations. These wireless stations are known as road side unit (RSU) and can be any equipment such as traffic lights, roadside monitors, and information traffic gateways which are connected to the Internet. Thus, VANETs will support both vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications.

In addition to various obstacles due to unreliable wireless transmission medium, development and operation of VANETs have unique challenges when compared with other forms of wireless networks. High node mobility, dynamic topology changes with frequent link breakage, and strict delay constraints of high priority safety messages are some common challenges in VANETs. These issues must be considered in developing networking protocols for VANETs. Recently, the IEEE 802.11p [4] has been proposed for medium access control (MAC) in VANETs to address the aforementioned issues. However, in the IEEE 802.11p, even successful broadcast messages are left unacknowledged. Further, with the random channel access, it suffers from unbounded latency and broadcast storm [5], [6]. On the other hand, as high priority safety messages are short range, uncoordinated, and broadcast in nature [7], they have a strict delay requirement and demand a reliable broadcast service. Distributed time division multiple access (TDMA) based MAC protocols, namely the ADHOC MAC [5] and the VeMAC [8], are proposed to facilitate reliable broadcast and point-to-point (P2P) communication in VANETs. However, due to VANET dynamic topology, the TDMA MAC protocols may lead to wastage of time slots. The wastage occurs when there are not enough nodes in a neighborhood to use all the time slots of a frame. In addition, upon a transmission failure, the source node has to wait until the next frame for retransmission even if the channel is idle during unreserved time slots. Hence, both the IEEE 802.11p and the existing TDMA based MAC approaches are not free from packet dropping and throughput reduction due to a poor channel condition. Further, these approaches can be inefficient in utilizing the available radio resources.

Various techniques such as diversity and channel coding are effective to mitigate wireless channel impairments and to improve network throughput. An alternative approach is cooperative communication, which makes use of nearby nodes to improve transmission performance between a pair of source and destination ($s - d$) nodes via diversity gain. Existing works on link layer cooperation focus on cooperation in the IEEE 802.11 based networks [9]–[11] and/or infrastructure based TDMA networks [12]–[14]. Different from the existing works, here we consider distributed TDMA MAC for VANETs referred to as Cooperative ADHOC MAC (CAH-MAC) [1].

This work was supported by a Strategic Project research grant from the Natural Science and Engineering Research Council (NSERC) of Canada.

In CAH-MAC, all operations such as cluster formation, slot allocation, cooperation decision and cooperation itself are performed in a distributed manner. Also, the helper is not fixed and changes with channel condition and network topology. As each node has reserved a time slot to transmit its own packets, we propose cooperation in the unreserved time slots. In this way, relay transmission in cooperation does not stop direct transmission from neighboring nodes, and hence does not increase the waiting time of neighboring nodes to access the channel. In this paper, we study the reliability of CAH-MAC in terms of packet transmission delay (PTD) and packet dropping rate (PDR). The CAH-MAC allows a neighboring node to use an unreserved time slot to retransmit the packet that failed to reach the destination in the same frame. Since the packet is retransmitted earlier in CAH-MAC as compared with ADHOC MAC, packet transmission delay decreases. Consequently, for a given maximum retransmission limit, the packet dropping rate is significantly reduced in CAH-MAC as compared with ADHOC MAC in similar networking and channel conditions.

This paper is organized as follows. Section II describes the system model and assumptions made to evaluate the performance of the CAH-MAC. The CAH-MAC protocol is described in Section III. Section IV presents performance analysis of CAH-MAC, which is verified in Section V with simulations. Finally, Section VI provides a summary of our contributions and identifies some issues for further investigation.

II. System Model

Consider a VANET consisting of N vehicles moving along a multi-lane road. Vehicles are distributed randomly. Let L be the number of lanes, each with width w_l, $l \in \{1, 2, 3, ..., L\}$. All vehicles move with negligible relative movements over an observation period. Hence, they are stationary with respect to each other, maintaining a fixed network topology. All vehicles are identical with respect to their communication capabilities with transmission range r. Therefore, vehicles with Euclidean distance more than r cannot communicate directly with each other. Vehicles within the transmission range of a source node can successfully receive the transmitted packets with probability p, taking account of a possible poor channel condition. The probability p depends on channel characteristics. The smaller the p value, the poorer the channel quality. The parameter p does not account for transmission errors due to the collision when multiple nodes within an interference range transmit simultaneously.

The channel time is partitioned into frames and each frame is further partitioned into time slots. Each time slot is of a constant time interval and each frame consists of a fixed number of time slots, denoted by F. Each vehicle is capable of detecting the start time of a frame and, consequently, the start time of a time slot. Accessing a time slot thus demands precise time synchronization among nodes. When a vehicle is equipped with a Global Positioning System (GPS) receiver, the one-pulse-per-second (1PPS) signal [15] that a GPS receiver gets every second can be use for the synchronization. Nodes support broadcast, multicast, or point-to-point modes of communication. However, to evaluate the performance of CAH-MAC, we consider nodes communicating in a point-to-point mode only. A helper node performs cooperation to retransmit an overheard packet from the source node.

Each vehicle maintains a list of its one-hop and two-hop neighbors. One-hop and two-hop nodes are those which can be reached at maximum one and two hops of transmission respectively from a reference node. Sets of these nodes are called one-hop set (OHS) and two-hop set (THS) respectively. All nodes in the same THS can communicate with each other with maximum two hops. Nodes form clusters of two-hop neighbors. Here a cluster refers to a group of nodes which are at a maximum two-hop transmission distance from each other. There is no cluster head, and a node can be a member of multiple clusters. Formation of a cluster stops simultaneous usage of a time slot by more than one node within the same interference range, thus avoids hidden and/or exposed node problems. Nodes belonging to the same THS contend with each other to reserve a time slot. To contend for a time slot, a node first listens to the channel over the period of F consecutive time slots (not necessarily in the same frame), then attempts to reserves one time slot among the unreserved ones if available. Access collisions occur when multiple nodes within the same interference range attempt to reserve the same time slot. After successfully reserving a time slot, a node transmits a packet in its own time slot in every frame until it encounters a merging collision [5] due to relative mobility. Merging collision occurs when nodes using the same time slot but belonging to different clusters approach each other, resulting in a transmission collision in the corresponding time slot [16]. In [16], it is shown that ADHOC MAC suffers from throughput reduction due to node mobility. To overcome the throughput reduction, VeMAC is proposed in [8]. In VeMAC, time slots are separated into three disjoint groups, dedicated to vehicles moving in opposite directions and to RSUs respectively. Separation of the time slots into three disjoint groups alleviates throughput reduction due to node mobility.

Here, with a focus on cooperation to improve transmission reliability, we consider a network where all nodes are perfectly synchronized and have already reserved their time slots. Hence, access collisions do not occur and cooperation is performed by only those nodes which have their own slots for transmission. Also, as relative mobility among nodes is negligible, merging collisions do not occur; hence a reserved time slot is always dedicated to its owner. All operations such as reserving a time slot, synchronization among nodes, cooperation decision, and cooperative transmission are done in a distributed manner, making it suitable for VANETs.

III. Cooperative ADHOC MAC

In this section, we review the operation of CAH-MAC as proposed in [1], including cooperation decision and helper selection. A node in its own time slot transmits a packet that consists of frame information, cooperation header, packet

header, pay load data, and cyclic redundancy check (CRC). Fig. 1 shows the structure of a packet that a node transmits. The packet header, payload data, and CRC are the same as in ADHOC MAC and VeMAC, whereas frame information is different. In addition, cooperation header is a new field that is introduced specifically for cooperation in CAH-MAC.

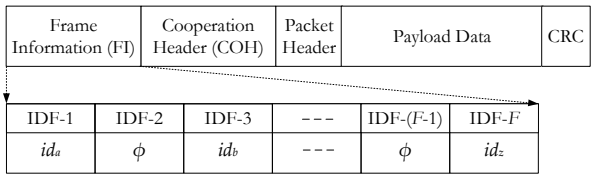


Fig. 1. Structure of a packet and a frame information field in CAH-MAC, where ϕ indicates an empty field.

A. Frame Information (FI)

The FI is a collection of ID fields (IDFs). The number of IDFs in an FI field is equal to F, i.e., the number of time slots per frame. Each IDF is dedicated to the corresponding time slot of a frame. The basic FI field structure is shown in Fig. 1. Destination node D, upon receiving a packet successfully from source node S in the s^{th} time slot, concludes that the s^{th} time slot belongs to S. Node D then puts the ID of node S in the s^{th} IDF of its FI. Hence by successfully receiving FIs from all of its one-hop neighbors, a node maintains a neighbor-table which includes: (i) all of its one-hop neighbors, (ii) all of its two-hop neighbors, and (iii) the owner of each time slot in a frame. If there is no signal in a time slot, then a node considers it as an unreserved time slot. In such a case, corresponding IDFs of unreserved time slots are left empty in an FI field.

A node can identify an unreserved time slot in which it can transmit without causing any collision in its one-hop neighborhood. Note that a node updates its neighbor-table based on any packets received successfully from new neighbors. These packets can be broadcast, unicast, or multicast packets. In addition to the neighborhood discovery and time slot reservation, the FI also helps for transmission acknowledgement. For example, consider that node D does not include the ID of node S in the IDF-S of its FI. Upon receiving FI from D, node S concludes a transmission failure between itself and D in the s^{th} time slot, which is basically a negative acknowledgement (NACK). Similarly, inclusion of the node S ID in the FI of node D serves as acknowledgement of a successful transmission from S to D.

B. Cooperation Among Neighboring Nodes

In the following, we discuss how a node decides and performs a cooperation. Denote S, D and H as the source, destination and helper nodes respectively. Cooperation decision and cooperative relay transmission are perform only if: (i) the direct transmission between source node S and destination node D fails, (ii) the helper node H successfully receives a packet from the source node S for cooperative relay transmission, (iii) the destination node D is reachable from the helper node H, and finally (iv) there is an available unreserved time slot for cooperative relay transmission.

If all the preceding conditions are satisfied, the helper node H offers cooperation to the source and destination, and the cooperative transmission is performed in one of the available unreserved time slots. Let the h^{th} slot of the frame be choosen for cooperation by node H. Fig. 2 shows necessary information exchanges for cooperation in the CAH-MAC. When the destination node D fails to receive a packet from source node S (in Fig. 2(a)), it announces transmission failure through its FI as shown in Fig. 2(b). Upon deciding to cooperate, the helper node H transmits its intention of cooperation using cooperation header (COH) as in Fig. 2(c). In the h^{th} time slot, after receiving a cooperation acknowledgement (C-ACK) from the destination node D, helper node H transmits the packet that node D failed to receive (in Fig. 2(d)).

Once a node decides to cooperate, it transmits its decision via cooperation header (COH) in its packet as shown in Fig. 1. If there are multiple potential helper nodes, the one which first announces to help will relay the packet while all other potential helpers will not proceed with cooperation for the same packet. Hence, helper H is the one which first offers cooperation in the frame and performs a cooperation for the $s-d$ pair. The information included in the cooperation header are (i) its intention to cooperate, (ii) the index of time slot of the source during which transmission failure occurred, and (iii) the selected unreserved time slot in which the packet will be retransmitted from the helper to the destination. The information is embedded in the cooperation header and transmitted in the helper's time slot. Collisions may occurs at the destination node when two or more potential helpers, which are not in each other's OHS, offer cooperation at the same unreserved time slot. In order to avoid such a collision, a cooperation acknowledgement (C-ACK) from the destination node is transmitted during the selected unreserved time slot. In C-ACK, the destination node puts the ID of the node H and acceptance of cooperation. Transmission of a C-ACK from the destination node forces other potential helper nodes to suspend their transmissions, thus avoiding any possible collision.

IV. Performance Analysis

In this section, we develop a mathematical model for the performance analysis of the CAH-MAC protocol. We will study the reliability of CAH-MAC in terms of packet transmission delay (PTD) and packet dropping rate (PDR) as compared with that of ADHOC MAC.

A. Vehicle Distribution

Vehicles are distributed randomly on the road with an exponentially distributed inter-vehicular distance over each lane. Let ρ_l, $l \in \{1, 2, 3, ..., L\}$, be the vehicle density of lane l in terms of the number of vehicles per unit length. Thus the counting of vehicles follows a Poisson process over a given length of road, such that the probability of finding m vehicles

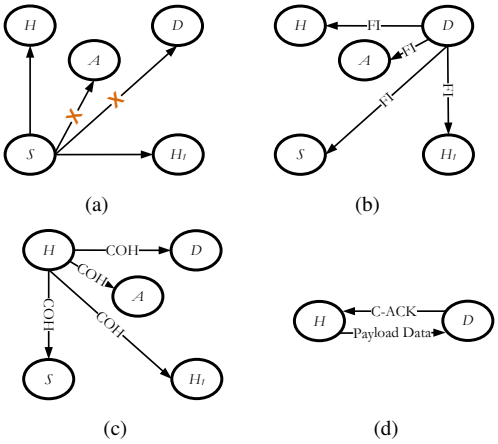


Fig. 2. Information exchanges in the CAH-MAC: (a) Phase 1: Source node transmits a packet to the destination; (b) Phase 2: Neighboring nodes detect transmission failure after examining the FI from the destination; (c) Phase 3: Helper node H, offers cooperation; (d) Phase 4: Helper node H, re-transmits the packet that failed to reach the destination after receiving a cooperation acknowledgement from the destination.

along a given length z of the road segment is given by

$$p(m,z) = \frac{(\rho z)^m e^{-\rho z}}{m!}, \qquad m = 0, 1, 2, ... \tag{1}$$

where $\rho = \sum_{l=1}^{L} \rho_l$.

Note that (1) is an approximation for tractable analytical framework, considering a vehicle as a point in a line representing a roadway. In reality, the inter-vehicular distance follows a shifted negative exponential distribution, such that a minimum safety distance (MSD) is always maintained by two adjacent vehicles in a lane to avoid any vehicle collision between them.

B. Direct Transmission

Let p_s denote the probability of successful transmission during a time slot. As channel condition (characterized by p) and transmission collision are independent of each other, p_s is given by

$$p_s = (1 - p_c)p \tag{2}$$

where p_c is the probability of transmission collisions in a given time slot. Collisions can be merging collisions [5] due to relative mobility between nodes. Since nodes are relatively stationary with respect to each other in the system model under consideration, there are no collisions among packets transmitted by different nodes, i.e, $p_c = 0$ and $p_s = p$.

C. Cooperative Relay Transmission

If a transmission failure occurs, cooperation may be triggered. Based on the criteria explained in Section III-B, upon a transmission failure between an $s - d$ pair, cooperation gets triggered if all of the following events occur:

1) Event E_1: There exits at least one potential helper which can help an $s - d$ pair to relay the packet that failed to reach the destination. Potential helpers are those nodes which are in the same OHS of the source and the destination. In addition, potential helpers must have successfully received the packet that failed to reach the destination. *Event* E_1 occurs if there is at least one potential helper. The probability of *Event* E_1 occurrences can be written as in (3).

2) Event E_2: There exists at least one unreserved time slot in which a potential helper can transmit without causing any collision in its OHS neighborhood. For nodes belonging to the same THS, an unreserved time slot for one node is unreserved for all of them. Hence, a potential helper can help an $s-d$ pair if there exists at least one unreserved time slot in the frame belonging to the corresponding THS. *Event* E_2 occurs if there exists at least one unreserved time slot in the frame, which is being shared by the source, the destination, and the potential helpers. The probability of *Event* E_2 occurrences is given by

$$\Pr\{E_2\} = \sum_{j=1}^{F-1} \frac{(2\rho r)^j e^{-2\rho r}}{j!}. \tag{4}$$

Events E_1 and E_2 are independent of each other. Hence, the probability of cooperation decision for each failed direct transmission, p_{coop}, is given by

$$p_{coop} = \Pr\{E_1\}\Pr\{E_2\}. \tag{5}$$

With the introduction of cooperation, transmission is successful either direct or cooperative relay transmission is successful. Hence, the probability of a successful transmission with cooperation, p_s^{coop}, is given by

$$p_s^{coop} = p_s + p_s(1 - p_s)p_{coop}. \tag{6}$$

D. Packet Transmission Delay

Upon transmission failure, a source attempts retransmission of a packet until it successfully reaches the destination. In this work, the packet transmission delay (PTD) is defined as the number of frames that is required to successfully transmit a packet to the destination. Since in the system model under consideration, the probability of successful transmission during a time slot depends only on channel characteristics, the number of retransmission attempts is independent of collision probability and follows a geometric distribution [17], [18]. Let random variables M and M_{coop} represent PTD of ADHOC MAC and CAH-MAC respectively. Hence, the probability mass function (pmf) of M with parameter p_s, is given by

$$\Pr\{M = i\} = (1 - p_s)^{i-1} p_s, \quad i = 1, 2, 3, ... \tag{7}$$

Similarly, the pmf of M_{coop} with parameter p_s^{coop} is given as

$$\Pr\{M_{coop} = i\} = (1 - p_s^{coop})^{i-1} p_s^{coop}, \quad i = 1, 2, 3, ... \tag{8}$$

Consequently, the expected values of M and M_{coop} are

$$E[M] = \frac{1}{p_s}, \; E[M_{coop}] = \frac{1}{p_s^{coop}} \tag{9}$$

respectively.

E. Packet Dropping Rate

In a communication system, a packet is dropped by a source node from its buffer memory, when it fails to deliver the packet to the destination within the predefine time limit. In our system, we consider this time limit in terms of the number

$$\Pr\{E_1\} = \sum_{k=3}^{F} \left(1-(1-p_s)^{k-2}\right) \frac{(1.5\rho r)^k e^{-1.5\rho r}}{k!} + \left(1-(1-p_s)^{F-2}\right)\left(1 - \sum_{k=0}^{F} \frac{(1.5\rho r)^k e^{-1.5\rho r}}{k!}\right) \quad (3)$$

of frames. Let M_{max} denote the maximum number of frames that a source node attempts to transmits a packet, referred as the maximum transmission limit. Hence, for a given M_{max} value, the packet dropping rate (PDR) of ADHOC MAC is given by

$$PDR = 1 - \sum_{i=1}^{M_{max}} (1-p_s)^{i-1} p_s. \quad (10)$$

With the cooperation, PDR as in (10) changes to

$$PDR_{coop} = 1 - \sum_{i=1}^{M_{max}} (1-p_s^{coop})^{i-1} p_s^{coop} \quad (11)$$

where PDR_{coop} is the packet dropping rate of CAH-MAC.

In the next section, we present numerical results to validate the performance analysis of CAH-MAC.

V. Analytical and Simulation Results

Computer simulations are performed in MATLAB. A road segment with two lanes, each of 5 m width, is considered, i.e., $L = 2$ and $w_l = 5$ m. Vehicles density, ρ_l, is kept 0.05 vehicle/m in both lanes, hence $\rho = L\rho_l = 0.1$ vehicle/m. For the fair comparison, the total number of time slots per frame is kept 60, i.e., $F = 60$ time slots/frame. Transmission range, r, is varied to change the number of THS members sharing a frame, and consequently, the number of unreserved time slots in a frame. The value of p is varied to characterize different channel conditions. The PTD and PDR of CAH-MAC are obtained and compared with ADHOC MAC for different scenarios.

Fig. 3 compares the PTD of CAH-MAC with that of ADHOC MAC. Figs. 4-6 compare the PDR of CAH-MAC with that of ADHOC MAC for different M_{max} values. It is observed that, at two extreme channel condition (i.e., $p = 0$ and 1), both protocols perform equally. When $p = 0$, all transmissions fail due to channel errors; thus there are no potential helpers and cooperation will not be trigged (i.e., $p_{coop} = 0$), resulting in PTD to be infinite and PDR to be 1 for both protocols. On the other hand, at $p = 1$, all packets reach to the destination directly from the source. Thus cooperation is not needed, resulting in PTD to be 1 and PDR to be 0 for both protocols. The advantage of cooperation starts as p increases from zero, such that a source node can get potential helpers upon a transmission failure.

Fig. 3 shows that PTD of CAH-MAC is almost reduced by 40% at a poor channel condition (for $p \leq 0.25$) as compared with that of ADHOC MAC when $r = 200$. However, for the case $r = 300$, reduction in PTD is only about 20%. This is due to the fact that, the advantage of cooperation can be achieved when there are a moderate number of THS members as compared with F. In such a case, there are potential helpers and sufficient unreserved time slots to perform cooperative relay transmission. When $r = 300$, the average number of THS

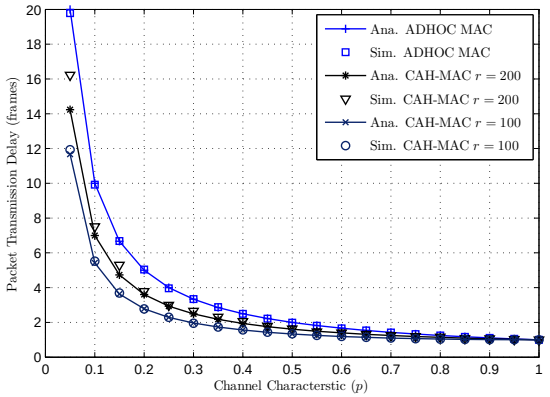


Fig. 3. Packet transmission delay of ADHOC MAC and CAH-MAC.

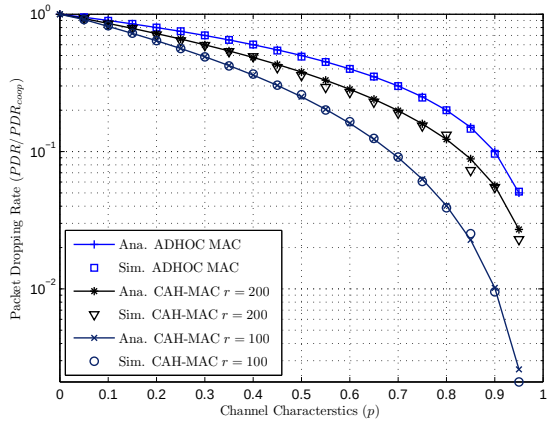


Fig. 4. Packet Dropping Rate of ADHOC MAC and CAH-MAC with $M_{max} = 1$ frame.

nodes sharing a frame is almost equal to the total available time slots. i.e., ($2\rho r = 60$), hence few unreserved time slots. The smaller the number of unreserved time slots or the larger the number of THS members, the smaller the p_{coop} value, which decreases the cooperation gain. Hence, a less number of unreserved time slots for the cooperative relay transmission results in a higher PTD for $r = 200$. As p increases, the delay improvement starts to decrease. The probability of a successful direct transmission increases with the improvement in the channel condition (i.e., $p \geq 0.85$), the gap between PTD values for both protocols decreases as the channel condition deteriorates.

Figs. 4-6 show PDR for both protocols at various M_{max} values. It is observed that the PDR of CAH-MAC is always less than that of ADHOC MAC for a given channel condition. However, the gap between the PDR values for CAH-MAC and ADHOC MAC increases as the channel gets better in a similar networking condition. Also, the gap increases with the probability of cooperative relay transmission p_{coop}. Hence, the gap between two protocols when $r = 200$ is higher than that

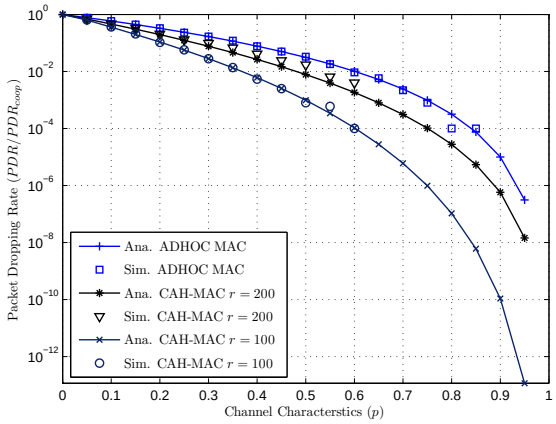


Fig. 5. Packet Dropping Rate of ADHOC MAC and CAH-MAC with $M_{max} = 5$ frames.

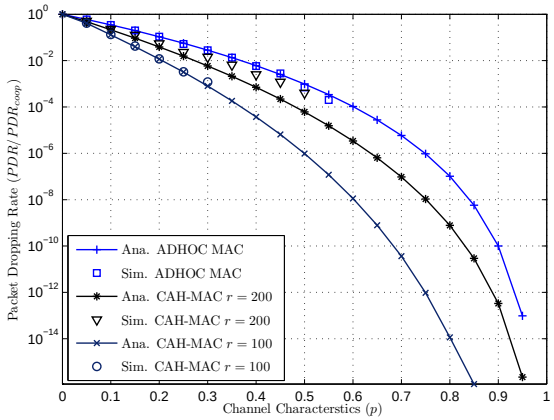


Fig. 6. Packet Dropping Rate of ADHOC MAC and CAH-MAC with $M_{max} = 10$ frames.

when $r = 300$. For the same channel condition, the larger the M_{max} value, the larger the gap between the PDR values of two protocols. Upon a transmission failure, in CAH-MAC, the helper node retransmits a packet in an unreserved time slot. Hence, with a larger M_{max}, a node in CAH-MAC gets more retransmission attempts than that of ADHOC MAC. This increases the probability of successful packet delivery to the destination within M_{max} frames, preventing it from being dropped from the buffer memory.

VI. Conclusion

In this paper, we study the reliability of cooperative ADHOC MAC protocol (CAH-MAC) for VANETs based on ADHOC MAC. In CAH-MAC, upon detecting a transmission failure between an $s - d$ pair, a neighboring node offers cooperation to relay the packet to the destination during an unreserved time slot. As a packet is retransmitted by a helper node, delay improvement is achieved and chances of packet being dropped decrease. We derive a close-form expression for the packet transmission delay and packet dropping rate of the CAH-MAC protocol, which are verified using simulations. Our analysis shows that the CAH-MAC protocol achieves lower packet delay and packet dropping rate than that of the ADHOC MAC under a similar networking condition. Numerical results demonstrate that CAH-MAC performs better in the presence of a moderate number of nodes in a two-hop neighborhood as compared with the total number of time slots available in a frame. Also, the packet dropping rate is smaller for a larger maximum retransmission limit value.

In this work, we have not considered relative mobility among nodes. Effects of dynamic network topology changes due to the relative mobility and a more realistic link model (other than the unit disk model) on the performance of CAH-MAC need further investigation.

References

[1] S. Bharati and W. Zhuang, "CAH-MAC: Cooperative ADHOC MAC for Vehicular Networks," *Submitted to IEEE J. Select. Areas Commun.*, Jan. 2012. [Online]. Available: https://engine.lib.uwaterloo.ca/ojs-2.2/index.php/pptvt/article/view/647

[2] United States Department of Transportation, "Intelligent Transportation Systems." [Online]. Available: http://www.its.dot.gov/index.htm

[3] H. Moustafa and Y. Zhang, *Vehicular Networks: Techniques, Standards, and Applications*. Boston, MA, USA: Auerbach Publications, 2009.

[4] IEEE Standard for Information Technology, "Telecommunications and information exchange between systems–Local and metropolitan area networks–Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Wireless Access in Vehicular Environments," 2010.

[5] F. Borgonovo, A. Capone, M. Cesana, and L. Fratta, "ADHOC MAC: New MAC Architecture for Ad Hoc Networks Providing Efficient and Reliable Point-to-Point and Broadcast Services," *Wireless Networks*, vol. 10, pp. 359–366, 2004.

[6] M. Hassan, H. Vu, and T. Sakurai, "Performance Analysis of the IEEE 802.11 MAC Protocol for DSRC Safety Applications," *IEEE Trans. Veh. Technol.*, vol. 60, no. 8, pp. 3882–3896, Oct. 2011.

[7] D. Jiang, V. Taliwal, A. Meier, W. Holfelder, and R. Herrtwich, "Design of 5.9 Ghz DSRC-Based Vehicular Safety Communication," *IEEE Wireless Communications*, vol. 13, no. 5, pp. 36–43, Oct. 2006.

[8] H. Omar, W. Zhuang, and L. Li, "VeMAC: A TDMA-based MAC Protocol for Reliable Broadcast in VANETs," *IEEE Trans. Mobile Comput.*, to appear.

[9] J. Zhang, Q. Zhang, and W. Jia, "VC-MAC: A Cooperative MAC Protocol in Vehicular Networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 3, pp. 1561–1571, March 2009.

[10] T. Zhou, H. Sharif, M. Hempel, P. Mahasukhon, W. Wang, and T. Ma, "A Novel Adaptive Distributed Cooperative Relaying MAC Protocol for Vehicular Networks," *IEEE J. Select. Areas Commun.*, vol. 29, no. 1, pp. 72–82, Jan. 2011.

[11] S. Moh and C. Yu, "A Cooperative Diversity-Based Robust MAC Protocol in Wireless Ad Hoc Networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 3, pp. 353–363, March 2011.

[12] A. Sadek, K. Liu, and A. Ephremides, "Collaborative Multiple-Access Protocols for Wireless Networks," in *Proc. IEEE ICC*, vol. 10, June 2006.

[13] Z. Yang, Y.-D. Yao, X. Li, and D. Zheng, "A TDMA-Based MAC Protocol with Cooperative Diversity," *IEEE Commun. Lett.*, vol. 14, no. 6, pp. 542–544, June 2010.

[14] G. Yuan, M. Peng, and W. Wang, "Opportunistic User Cooperative Relaying in TDMA-Based Wireless Networks," *Wireless Communication and Mobile Computing*, vol. 10, no. 7, pp. 972–985, July 2010.

[15] W. Ding, J. Wang, Y. Li, P. Mumford, and C. Rizos, "Time Synchronization Error and Calibration in Integrated GPS/INS Systems," *ETRI Journal*, vol. 30, no. 1, pp. 59–67, 2008.

[16] F. Borgonovo, L. Campelli, M. Cesana, and L. Fratta, "Impact of User Mobility on the Broadcast Service Efficiency of the ADHOC MAC Protocol," in *Proc. IEEE VTC Spring*, vol. 4, June 2005.

[17] R. Fantacci and S. Nannicini, "Multiple Access Protocol for Integration of Variable Bit Rate Multimedia Traffic in UMTS/IMT-2000 Based on Wideband CDMA," *IEEE J. Select. Areas Commun.*, vol. 18, no. 8, pp. 1441–1454, Aug. 2000.

[18] D. Vassis and G. Kormentzas, "Delay Performance Analysis and Evaluation of IEEE 802.11e EDCA in Finite Load Conditions," *Wirel. Pers. Commun.*, vol. 34, no. 1-2, pp. 29–43, July 2005.

A Raytracing Model for Wireless Propagation in Tunnels with Varying Cross Section

Camillo Gentile, Fabien Valoit, and Nader Moayeri

Advanced Network Technologies Division
National Institute of Standards and Technology
Gaithersburg, MD, USA

{}

***Abstract*— Mandated by the 2006 United States Miner Act, reliable two-way communications in mines has drawn the interest of network engineers in recent years. Critical to the design of these systems is an accurate channel propagation model. Given the elementary geometry seen in most tunnels, models that approximate them as a rectangular waveguide have been developed. These models are extremely accurate in vehicular tunnels because – since the tunnel is typically cast from concrete – the cross section is uniform throughout and the surface roughness is negligible. Mines, however, do not conform to these two conditions. In this paper, we extend the waveguide model to tunnels with varying cross section and measurable surface roughness. The effectiveness of the proposed model is validated through in-house field measurements collected in a vehicular tunnel and in a coal mine. We show that while the original model performs well in the former, it falters in the latter. The extended model, however, predicts reliably in the mine as well.**

***Index Terms*— Mine; surface roughness; convex; concave**

I. Introduction

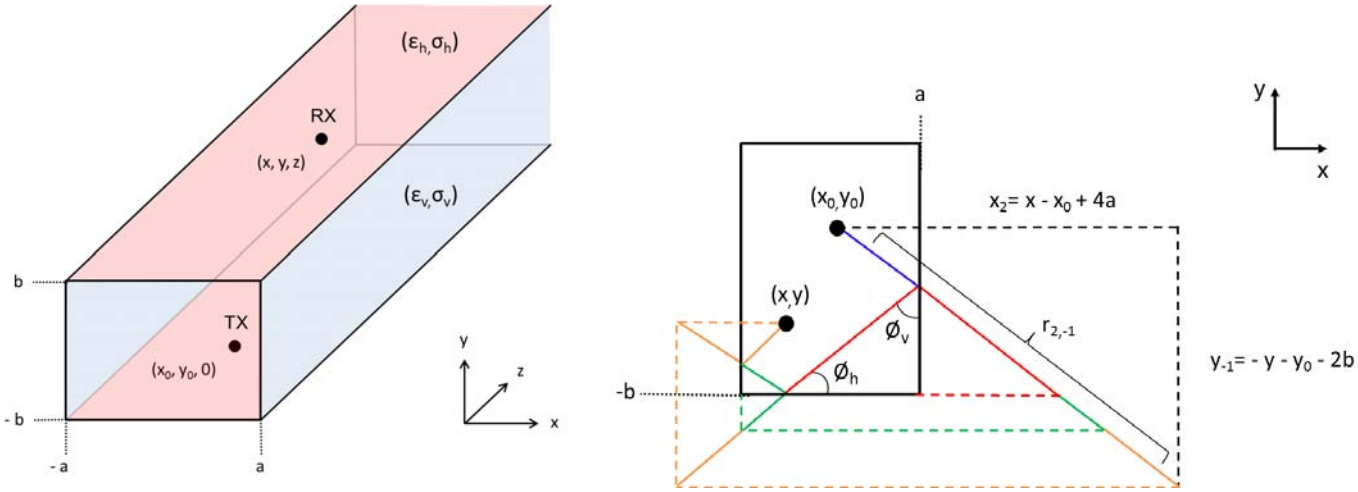
Prompted by two mining accidents which caused the deaths of 14 miners in West Virginia in January 2006, the United States Congress passed the most sweeping legislation in close to 30 years. Amongst other provisions, the 2006 Miner Act calls for two-way communications between underground miners and rescuers, as well as location tracking. Critical to the design of these systems is an accurate channel propagation model. In this context, recent years have seen the development of analytical models to predict the radio propagation of electromagnetic waves in tunnels [1], [2], [3]. The papers leverage the elementary geometry seen in most tunnels so as to approximate them as a rectangular waveguide, for which well-defined theory has been established. The theory permits closed-form expressions for the electromagnetic field by way of raytracing the reflected paths.

Associated with these analytical models are two main underlying assumptions. The first is that the radio waves behave like light, hence that the effects of diffraction can be ignored. This assumption is valid when the dimensions of the tunnel are much larger than the wavelength of the signal. This translates to a center frequency of a few hundred Megahertz for tunnels with cross-sectional dimensions on the order of several meters. The second assumption, of course, is that the tunnel is well approximated by a rectangular cuboid. For example, the results in [2], [3] are presented for a vehicular tunnel in France at center frequencies of 450 MHz and 915 MHz. Since the cross section of the tunnel is close to rectangular – and moreover uniform throughout the shaft – the predictions are extremely accurate. However, when the cross section of the tunnel varies throughout the shaft – a prime example is in mines – the predictions break down. The breakdown does not arise simply from the variation in the dimensions, but also because they are typically much smaller than the dimensions of vehicular tunnels. As the dimensions decrease, the model is more sensitive to such variations as they represent a larger percentage of the nominal cross section.

To our knowledge, [3] is the most recently published work in this specific research area. Their model, intended for uniform cross sections, is described in Section II. In this paper, we propose two major improvements to it. The first is a raytracing model for tunnels with varying cross sections throughout the shaft. It is presented in Section III. The second improvement, in Section IV, is an extension of the model to include surface roughness. Again, as opposed to vehicular tunnels, this factor is much more significant in coal (iron, precious metals, etc.) mines due to the natural roughness of the rock. In order to substantiate the proposed model, we compare predicted results to in-house field measurements in two tunnels: a train tunnel and a coal mine. In Section V we show that in the mine – with varying cross section and surface roughness – our proposed model predicts reliably while the uniform model falters. The results are summarized in the Conclusions section.

II. The uniform cross section model

In this section, we present the raytracing model described in [3]. Figure 1(a) illustrates the model parameters. The cross-sectional profiles of the tunnel are given as (a, b) in reference to the coordinate system with origin at the center of the tunnel, $(x = 0, y = 0, z = 0)$, The cross section is said to be uniform because the profiles do not vary with z. The vertical (v) and horizontal (h) walls have electromagnetic properties (ϵ_v, σ_v) and (ϵ_h, σ_h) respectively, where ϵ denotes the permittivity and σ the conductivity. The air in the tunnel has the permittivity of free space, ϵ_0, and zero conductivity. The permeability of the air and the walls are all assumed to be equal to that of free space. The transmitter is positioned at $(x_0, y_0, 0)$ and the receiver at (x, y, z), where z is the range between the two.



(a) The tunnel parameters (b) The image reflection model

Fig. 1. The uniform cross section model. The vertical and horizontal walls have different electromagnetic properties and their profiles are symmetric about the respective axes. The cross section is uniform throughout the shaft of the tunnel.

The model traces all the rays (or paths) emitted from the TX which are received by the RX. Each path is indexed as (p, q) according to the total number of $|p|$ reflections off the vertical walls and $|q|$ reflections off the horizontal walls. The reflections alternate between the positive and negative walls along the shaft. A positive value of p indicates that the first reflection hits the positive vertical wall while a negative value indicates that the negative wall is hit first. The sign convention is the same for the horizontal walls.

The field attenuation associated with the collection of all propagation paths can be written as

$$\alpha(x,y,z) = \sum_{p=-\infty}^{p=\infty} \sum_{q=-\infty}^{q=\infty} \left[\frac{\exp(-jkr_{p,q})}{r_{p,q}}\right] \cdot R_v^{|p|} \cdot R_h^{|q|}, \quad (1)$$

where $r_{p,q}$ denotes the propagation distance of the indexed path. The bracketed term represents the complex attenuation due purely to distance traveled. As shown, it is inversely proportional to $r_{p,q}$. The numerator accounts for the phase shift, where $k = \frac{2\pi f_0}{c}$ is the wavenumber, f_0 is the center frequency of the signal, and c is the speed of light. The model assumes that reflection is the sole specular effect present in the tunnel. As expressed in (1), with each reflection off the walls the signal is attenuated by the corresponding reflection coefficient. The vertical and horizontal walls have reflection coefficients [4]:

$$R_v = \frac{\cos\phi_h - \sqrt{\bar{k}_h - \sin^2\phi_h}}{\cos\phi_h + \sqrt{\bar{k}_h - \sin^2\phi_h}} \quad (2)$$

and

$$R_h = \frac{\bar{k}_v\cos\phi_v - \sqrt{\bar{k}_v - \sin^2\phi_v}}{\bar{k}_v\cos\phi_v + \sqrt{\bar{k}_v - \sin^2\phi_v}}, \quad (3)$$

where $\bar{k}_v = \left(\epsilon_v + \frac{\sigma_v}{j2\pi f_0}\right)/\epsilon_0$ and $\bar{k}_h = \left(\epsilon_h + \frac{\sigma_h}{j2\pi f_0}\right)/\epsilon_0$. The incident angles off the vertical and horizontal walls, ϕ_v and ϕ_h respectively, are calculated in the next subsection. Implicit to the coefficients above is that the dipole antennas are vertically polarized. Analogous equations for horizontally polarized antennas are provided in [3]. In both cases, only the electromagnetic field in the main polarization direction is considered, assuming that the effects of cross polarization are minimal.

A. The uniform image reflection model

The propagation distance, $r_{p,q}$, is calculated through the image reflection model [5]. Fig. 1(b) shows the images for path $(p = 2, q = -1)$ in the (x, y)-plane. For every vertical reflection, the image is displaced by a value $2a$ in the x-direction; likewise, for every horizontal reflection, the image is displaced by a value $2b$ in the y-direction. Then the total x-displacement for p reflections is

$$x_p = (\text{-}1)^p x - x_0 + p \cdot 2a \quad (4)$$

and the total y-displacement for q reflections is

$$y_q = (\text{-}1)^q y - y_0 + q \cdot 2b. \quad (5)$$

Knowing z as well, the three can be combined as:

$$r_{p,q} = \sqrt{x_p^2 + y_q^2 + z^2}. \quad (6)$$

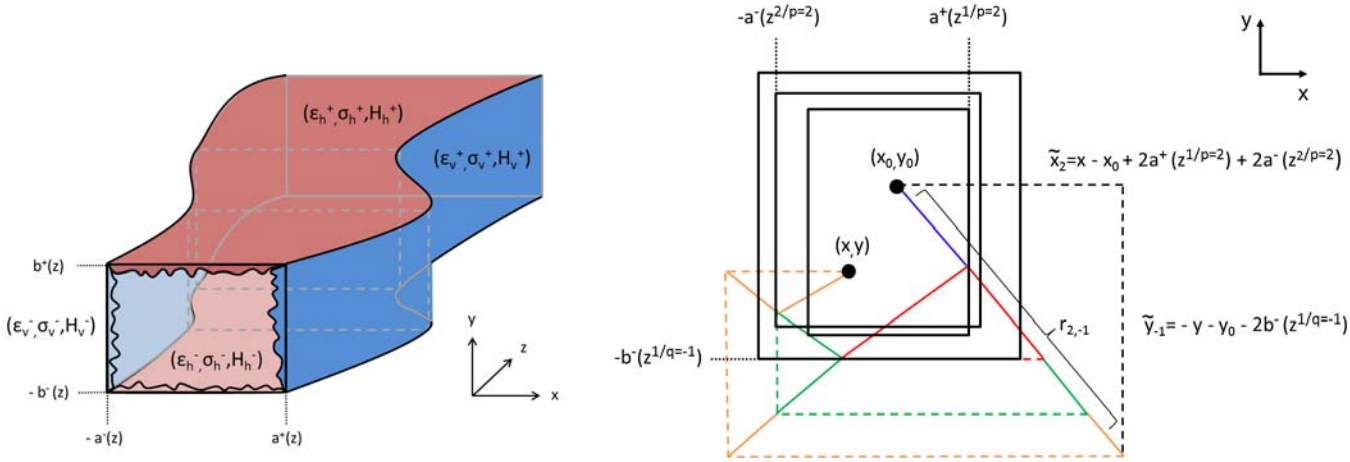
Also, from the (x, y)-displacements, the incident angles are calculated as

$$\phi_v = \arcsin\frac{|x_p|}{r_{p,q}} \quad (7)$$

and

$$\phi_h = \arcsin\frac{|y_p|}{r_{p,q}}. \quad (8)$$

Given the rectangular shape of the tunnel, the tunnel behaves as a rectangular waveguide. In fact it is shown in [3] how the raytracing model can be decomposed into a waveguide model expressed in terms of all the propagation modes.



(a) The tunnel parameters (b) The image reflection model

Fig. 2. The varying cross section model. The positive / negative walls may have different electromagnetic properties and their profiles may be asymmetric about the respective axes. The cross section varies throughout the shaft of the tunnel.

III. The varying cross section model

In this section, we extend the uniform cross section model to account for large-scale variation in the tunnel profiles along the shaft. As a means to describe this variation, the vertical and horizontal profiles, $\big(a^{\pm}(z), b^{\pm}(z)\big)$, are written explicitly in terms of z. Figure 2(a) illustrates the model parameters. Notice that the positive / negative profiles are labeled separately, meaning that they may now be asymmetric, as shown. The attenuation is computed as in (1):

$$\tilde{\alpha}(x,y,z) = \sum_{p=-\infty}^{p=\infty}\sum_{q=-\infty}^{q=\infty}\left[\frac{\exp(-jkr_{p,q})}{r_{p,q}}\right]\cdot \tag{9}$$
$$(R_v^+)^{\lfloor\frac{p+1}{2}\rfloor}(R_v^-)^{\lfloor\frac{p}{2}\rfloor}(R_h^+)^{\lfloor\frac{q+1}{2}\rfloor}(R_h^-)^{\lfloor\frac{q}{2}\rfloor},$$

however the expressions for the displacements x_p and y_p in (4) and (5) must be modified as explained in the sequel. Also notice that positive / negative profiles may have different electromagnetic properties with designated reflection coefficients ($R_v^{\pm}$, $R_h^{\pm}$) given respectively from (2) and (3). Based on the values of (p,q), (9) accounts for the number of times the signal is incident with each of the four walls.

A. The varying image reflection model

To see the effect of varying cross section on the displacements, we reconsider the image reflection model. Recall that for a uniform cross section, the x-displacement is augmented by $2a$ with each reflection off the vertical wall. This translates to the total displacement in (4) for p total reflections. Since the profile is uniform, where the reflections occur has no bearing on x_p. This is not the case when the profiles are varying. As such, we denote the vertical profile at the $\tilde{p}^{th}$ out of p total reflections as $a^{\pm}(z^{\tilde{p}|p})$, where $z^{\tilde{p}|p}$ is the z-coordinate at which the reflection occurs and the sign is given from the values of $\tilde{p}$ and p as $\text{sgn}(p)(\text{-}1)^{\tilde{p}+1}$. Now the x-displacement is augmented instead by $2a^{\pm}(z^{\tilde{p}|p})$ with reflection $\tilde{p}$. Of course, the same relationship applies for the y-displacement with a total of q reflections off the horizontal wall. Then the (x,y)-displacements follow as

$$\tilde{x}_p = (\text{-}1)^p x - x_0 + \text{sgn}(p)\cdot 2\sum_{\tilde{p}=\text{sgn}(p)}^{p} a^{\pm}(z^{\tilde{p}|p}) \tag{10}$$

and

$$\tilde{y}_q = (\text{-}1)^q y - y_0 + \text{sgn}(q)\cdot 2\sum_{\tilde{q}=\text{sgn}(q)}^{q} b^{\pm}(z^{\tilde{q}|q}). \tag{11}$$

An example for $(p=2, q=-1)$ is illustrated in Fig. 2(b). As opposed to the uniform cross section in Fig. 1(b), shown here are the three cross sections corresponding to the three reflection points. Notice that the positive / negative profiles are asymmetric.

Compared to the uniform raytracing model, the varying model allows for wide-ranging tunnel profiles. However, in it, the cross section is modeled as rectangular – a stipulation upon which the image reflection model is based – such that the reflected angle is the same as the incident angle in the (x,y)-plane. In practice, most tunnels have either a circular, oval, or arched cross section; nevertheless the uniform model has been shown to give accurate results in those cases. The image reflection model also stipulates that the reflected angle be the same as in the incident angle in the (x,z)- and (y,z)-planes, i.e. that the profiles are uniform – a condition which is obviously violated in the varying model. To minimize any violation, the orientation of the coordinate system should be chosen such that its axes lie as parallel as possible to the walls. So long as the local variation in the profiles is gradual – which is typical in most tunnels – the varying model can be applied. Results to substantiate this are presented in Section V.

B. Computing the reflection points

We now turn our attention to a method for finding the reflection points. This method is illustrated by considering the simple example for $(p=1, q=0)$ shown in Figure 3.

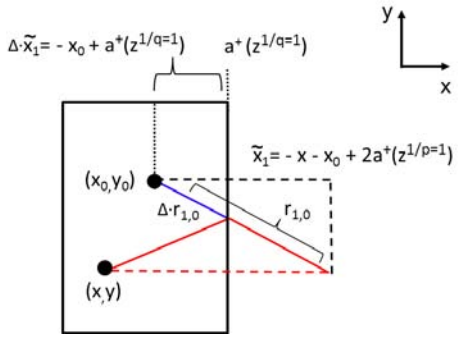


Fig. 3. The reflection point $z^{1|q=1}$ for $(p=0, q=1)$. The marginal distance $\Delta \cdot r_{1,0}$ traveled from the transmitter to the reflection point can be expressed as a fraction Δ of the total distance $r_{1,0}$ traveled to the receiver.

The objective is to compute the value of $z^{1|p=1}$ at which the single reflection occurs. This value then maps to $a^+(z^{1|p=1})$. The marginal distance traveled from the transmitter to the reflection point can be expressed as a fraction Δ of the total distance $r_{1,0}$, or

$$\begin{aligned} \Delta \cdot r_{1,0} &= \Delta \cdot \sqrt{\tilde{x}_1^2 + \tilde{y}_0^2 + z^2} \\ &= \sqrt{(\Delta \cdot \tilde{x}_1)^2 + (\Delta \cdot \tilde{y}_0)^2 + (\Delta \cdot z)^2}. \quad (12) \end{aligned}$$

Of course the marginal distance traveled is along the same direction as the total distance traveled, i.e. along the direction of image $(p=1, q=0)$. Moving an amount Δ along this direction implies scaling each of the three vector components proportionally as in (12). In addition, it can be seen from the figure that the x-displacement of the image to the reflection point is $\Delta \cdot \tilde{x}_1 = -x_0 + a^+(z^{1|p=1})$. From it, the value of Δ can be recovered as

$$\Delta = \frac{-x_0 + a^+(z^{1|p=1})}{\tilde{x}_1}. \quad (13)$$

By definition the z-coordinate of the image to the reflection point is given as $\Delta \cdot z = z^{1|p=1}$. Finally, by substituting (13) into the latter, we arrive at

$$z^{1|p=1} = \left[\frac{-x_0 + a^+(z^{1|p=1})}{\tilde{x}_1} \right] \cdot z. \quad (14)$$

Through the same analysis, the z-coordinate can be indexed to any vertical or horizontal reflection point as

$$z^{\tilde{p}|p} = \left[\frac{-x_0 + \text{sgn}(p) \cdot \left(a^{\pm}(z^{\tilde{p}|p}) + 2 \sum_{\varrho=\text{sgn}(\tilde{p})}^{\tilde{p}-\text{sgn}(\tilde{p})} a^{\pm}(z^{\varrho|p}) \right)}{\tilde{x}_p} \right] \cdot z \quad , \quad \tilde{p} = \text{sgn}(p) \ldots p. \quad (15)$$

and

$$z^{\tilde{q}|q} = \left[\frac{-y_0 + \text{sgn}(q) \cdot \left(b^{\pm}(z^{\tilde{q}|q}) + 2 \sum_{\varrho=\text{sgn}(\tilde{q})}^{\tilde{q}-\text{sgn}(\tilde{q})} b^{\pm}(z^{\varrho|q}) \right)}{\tilde{y}_q} \right] \cdot z, \quad \tilde{q} = \text{sgn}(q) \ldots q. \quad (16)$$

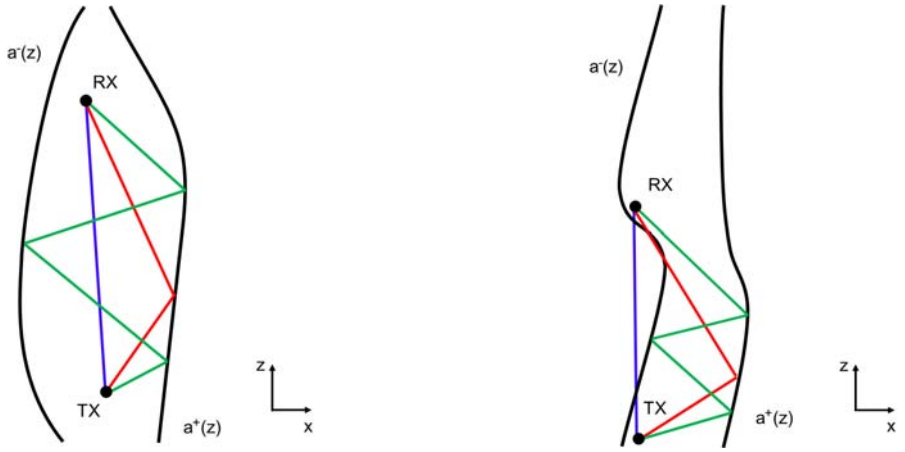
Since the numerator in (15) involves all the vertical reflection points, the $|p|$ equations must be solved simultaneously; likewise, the $|q|$ equations for the horizontal reflection points in (16) must also be solved simultaneously.

C. Convex versus concave profiles

Because the set of equations in (15) is non-linear, it must be solved through numerical methods. The numerical methods for finding the p reflection points involves searching along the vertical profiles. We first consider the case for which the profiles are convex. Figure 4(a) illustrates an example of convex positive / negative vertical profiles in the (x, z)-plane. The equations can be solved efficiently in the $|p|$-dimensional convex space through a gradient descent algorithm such as the Bisection method [6]. For uniform vertical profiles, the reflection points will occur at equal intervals along the z-axis; conversely, for varying vertical profiles, adjacent points occurring at consecutive positive and negative profile values below the respective profile averages will appear closer together – since the ray travels a shorter distance between them – while those occuring at values above the average will appear farther apart. At initialization, the points are ordered with equal spacing between each other; the gradient descent algorithm effectively scales the spacing according to the profile values at which they occur. While the range of any point is limited to $0 < z^{\tilde{p}|p} < z$, additional constraints limit the search space. Specifically, e.g. for p positive, knowing that the $\tilde{p}^{th}$ reflection will occur before the $(\tilde{p}+1)^{th}$ reflection along the shaft implies $z^{\tilde{p}|p} < z^{\tilde{p}+1|p}$, $\tilde{p} = 1 \ldots p-1$.

Fig. 4(b) illustrates an example of concave positive / negative vertical profiles in the (x, z)-plane. In tunnels with concave profiles, solving for the set of equations in turn entails a search in a concave space, which in general is much less efficient than in its counterpart. However, since the search is just between adjacent points – so long as the associated profiles are convex in between – finding the global optimum requires no additional complexity. This is often the case because, as mentioned earlier, in most tunnels the variation in the the wall profiles is gradual. Notwithstanding, for $|p| = 1$ the point could fall at any point in the range, indeed requiring a concave search. While in a single dimension a concave search is still quite efficient, it becomes more and more complex as the number of reflection points increases; however, the increased point density tightens the spacing between adjacent points, limiting the search space further and compensating to some degree for the greater complexity.

There is actually a more important issue in concave tunnels than the efficiency of the numerical method. In convex



(a) Convex (b) Concave

Fig. 4. Positive / negative vertical profiles in the (x, z)-plane. When the profiles are convex, the direct path and all the reflected paths are detected by the receiver, however, when the profiles are concave, they may interfere with some walls.

tunnels the transmitter and receiver are always in line-of-sight of each other. It follows that the direct path and all the reflected paths will pass unobstructed by the tunnel walls, as shown in Fig. 4(a). In concave tunnels, however, this is not always the case. When the direct path is obstructed – because the typical materials and the thickness of the tunnel walls (e.g. concrete, coal) permit virtually no penetration – it will not be detected by the receiver. It is easy to determine through simple geometry whether this is the case; if so, the associated term $(p = 0, q = 0)$ should be eliminated from the sum in (9). In general, once path (p, q) is computed, it should be checked through the same geometry and eliminated if obstructed. Note that as the reflection order increases, so does the incident angle, allowing the signal to more readily reflect around concave segments, as shown in Fig. 4(b). The paths for q = 0 (blue) and q = -1 (red) are obstructed, however not for q = -3 (green). For gradual profiles, unobstructed paths with lower reflections imply that the higher-order paths are also unobstructed and hence need not be checked.

IV. Surface roughness

Surface roughness, which may be viewed as small-scale or local variation in the dimensions of the cross section, can account for significant attenuation of the signal upon incidence with a wall, especially at higher frequencies. At higher frequencies, it can be seen from (9) that the phase of an individual path varies more as a function of path length. As the paths scattered from different parts of the rough surface combine with correspondingly different lengths, the phases add destructively to create interference, attenuating the signal [7].

The conventional metric for surface roughness is the quantity H, which is the root mean square of the deviation in height of the profile from its large-scale value. Dependent upon the incident angle, the effective height is reduced to $H \sin\phi$. The surface roughness coefficients of the signal upon incidence can be expressed as [7]

$$S_v^{\pm} = \exp\left(-2\left(kH_v^{\pm}\sin\phi_v\right)^2\right) \tag{17}$$

TABLE I

Tunnel parameters

	Train tunnel	Coal mine
a (m)	2.44	1.22 - 1.52
b (m)	3.12	0.84 - 1.14
$\epsilon\left(\frac{F}{m}\right)$	$7 \cdot \epsilon_0$	$4 \cdot \epsilon_0$
$\sigma\left(\frac{S}{m}\right)$	0.0150	0.0007
H (cm)	0	20

and

$$S_h^{\pm} = \exp\left(-2\left(kH_h^{\pm}\sin\phi_h\right)^2\right), \tag{18}$$

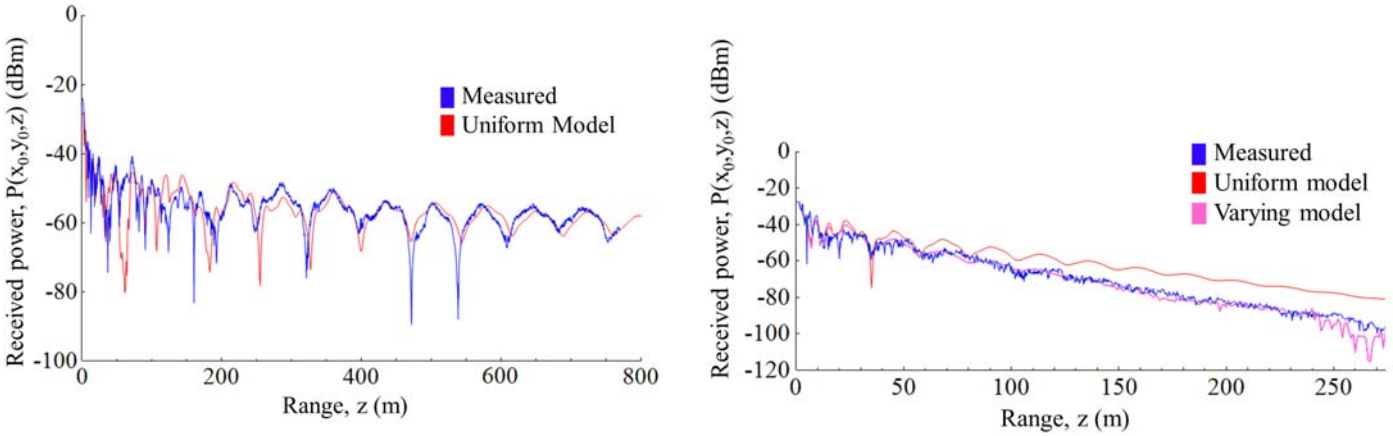
where ($H_v^{\pm}$, $H_h^{\pm}$) are the respective roughness metrics for the positive / negative vertical and horizontal walls. These metrics are depicted in Figure 2(a). Notice, as explained earlier, that the attenuation coefficients increase with center frequency. However, in the seminal paper [8] on wireless propagation in tunnels, Emslie et al. indicate exactly the opposite. This is because in their development they consider only the fundamental propagation modes in the waveguide model while here they are all considered.

By modeling the effect of surface roughness, we arrive at a new expression for the attenuation at the receiver:

$$\hat{\alpha}(x, y, z) = \sum_{p=-\infty}^{p=\infty}\sum_{q=-\infty}^{q=\infty}\left[\frac{\exp(-jkr_{p,q})}{r_{p,q}}\right]\cdot \tag{19}$$
$$(R_v^+ S_v^+)^{\left\|\frac{p+1}{2}\right\|}(R_v^- S_v^-)^{\left\|\frac{p}{2}\right\|}(R_h^+ S_h^+)^{\left\|\frac{q+1}{2}\right\|}(R_h^- S_h^-)^{\left\|\frac{q}{2}\right\|}.$$

V. Experimental results

In this section, we compare the uniform and varying models against field measurements collected in two different types of tunnel. In our experiments, the transmitter and receiver antennas were mounted on tripods and both positioned at the midpoint of the tunnel cross section. The receiver was initially placed at z = 1 m. Then the range from the transmitter was increased progressively while maintaining the same



(a) Train tunnel

(b) Coal mine

Fig. 5. Experimental results at center frequency f_0 = 915 MHz. In the train tunnel with uniform cross section, the uniform model reliably tracks the measured received power; however, in the coal mine with varying cross section, it falters while the varying model performs well.

position in the (x, y)-plane, hence for $x = x_0$ and $y = y_0$. At each point, the received power was measured across the frequency band of interest using a spectrum analyzer. Both the transmitter and the receiver were equipped with vertically polarized omni-directional antennas and the transmit power was set to P^{TX} = 1.5 dBm. Knowing the transmit power, the received power predicted from the attenuation model is given through Friis formula [4] as

$$P(x, y, z) = P^{TX} \cdot \left(\frac{c}{4\pi f_0}\right)^2 |\alpha(x, y, z)|^2. \qquad (20)$$

The measured received power and predicted received power, $P(x_0, y_0, z)$, were then compared.

The first tunnel was the Sorrento Tunnel in Plummer, Idaho. The train tunnel is constructed from concrete, thus cast with negligible surface roughness; most importantly, the structure has a uniform profile throughout. Since the ceiling is arched, the vertical profile parameter, b, was computed such that its value corresponds to the rectangular cross section which best approximates the actual one in the least-squares sense. Also, the material properties were the same for all four walls. Table I shows the parameters for both tunnels. The second tunnel was an experimental coal mine in Pittsburgh, Pennsylvania. There is a number of significant differences between the two tunnels. First of all the dimensions of the mine range from a fourth to half the size of the train tunnel. Secondly, the profile parameters vary up to 35% through the mine shaft while in the train tunnel they are constant. Finally, the surface roughness of the coal is 20 cm.

Fig. 5(a) displays the measured versus the predicted received power as a function of z in the train tunnel. Observe that, despite the rectangular approximation, the uniform model predicts very reliably – both in the near field of the transmitter as well as in the far field – tracking the peaks and valleys where the paths respectively combine constructively and destructively throughout the shaft. In view of the varying profiles coupled with the surface roughness of the coal mine, Fig. 5(b) shows that the uniform model falters. However, the varying model still predicts reliably.

VI. Conclusions

In this paper, we have proposed a novel analytical model for the propagation of electromagnetic waves in tunnels with varying cross section. Tunnels with both convex and concave profiles have been considered. The model has application, in particular, to the design of wireless communications in mines. The major underlying assumption is that the variance in the cross section rolls out gradually along the shaft. The proposed model has been validated with in-house field measurements and has been shown to predict reliably in a coal mine. Further work includes refining the model for more complex tunnel structures, including cross cuts.

References

[1] Mahmoud, S.F., "Modal Propagation of High Frequency Electromagnetic Waves in Straight and Curved Tunnels within the Earth," *Journal of Electromagnetic Waves and Applications*, vol. 19, no. 12, pp. 1611-1627, 2005.

[2] Dudley, D.G.; Lineard, M.; Mahmoud, S.F.; Degauque, P., "Wireless Propagation in Tunnels," *IEEE Antennas and Propagation Magazine*, vol. 49, no. 2, April 2007.

[3] Sun, Z.; Akyildiz, I.F., "Channel Modeling and Analysis for Wireless Networks in Underground Mines and Road Tunnels, *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1758-1768, June 2010.

[4] Frankl, D.R., *Electromagnetic Theory,* Prentice-Hall, Englewood Cliffs, NJ, 1986.

[5] Valenzuela, R.A., "A ray tracing approach to predicting indoor wireless transmission," *IEEE Vehicular Technology Conf.*, pp. 214-218, May 1993.

[6] Burton, R.L.; Fairies, D.J., *Numerical Analysis, Seventh Edition*, Brooks/Cole, Pacific Grove, CA, 2001.

[7] Beckmann, P. and Spizzichino, A., *The Scattering of Electromagnetic Waves from Rough Surfaces*, Macmillan, New York, 1963.

[8] Emslie, A.G.; Lagace, R.L.; Strong, P.F., "Theory of the Propagation of UHF Radio Waves in Coal Mine Tunnels," *IEEE Trans. on Antennas and Propagation*, vol. 23, no. 2, March 1975.

Novel Scheduling for a Mixture of Real-time and Non-real-time Traffic

Huseyin Haci, Huiling Zhu and Jiangzhou Wang
School of Engineering and Digital Arts, University of Kent, UK

***Abstract*— This paper presents a novel scheduling policy for wireless communications where users contain mixed real-time (RT) and non-real-time (nRT) traffic. The proposed policy introduces a novel tradeoff concept in scheduling, where a value of quality of service (QoS) satisfaction can be traded off with the achievable system throughput. Two weighting factors, corresponding to RT and nRT traffic, are introduced in the novel scheduling policy. The proposed policy is evaluated by simulation results. The simulation results show an optimal range of the weights to meet QoS satisfactions.**

I. INTRODUCTION

The performance of a wireless network can be severely affected by the time varying property of wireless channels and the co-channel interference, which are caused by dynamic changes in users' locations, multipath propagation and frequency reuse. Conventionally researchers exploit link adaptation and diversity techniques to mitigate disturbances.

Link adaptation techniques are generally developed through radio resource management (RRM) [1], which adjusts radio resources, such as transmit power and modulation level, depending on instantaneous wireless channel quality.

Besides RRM, scheduling mechanisms take advantage of an inherent property of wireless networks, and multiuser diversity to improve the network performance. It is shown in [2] that the system performance can be optimized by scheduling appropriate users (e.g. users with good channel conditions [3]-[5]) to be served in each RRM period.Scheduling and RRM play complementary roles in providing efficient network function and quality of service (QoS) in wireless networks. Scheduling determines the order of users to be served in time domain and RRM allocates resources among users chosen to be served.

Two popular scheduling approaches are Round Robin (RR) and opportunistic (OP) scheduling [6]. RR aims to provide complete fairness between users and do not care about system performance optimization (e.g. pay no attention to channel conditions). This, as a result, significantly degrades system throughput. On the contrary, OP aims to maximize the system performance without considering fairness. OP exploits the time-varying channel conditions of users and provides access right to the user with the best channel quality. The optimum system capacity can be achieved with OP. However, due to severe unfairness (i.e. less probability of serving users with bad channels), it is not very practical.

To find an efficient yet practical solution these two approaches were later integrated into a scheme called proportional fair (PF) scheduling [7]. PF incorporates both fairness and throughput and offers fairness without sacrificing too much throughput. The scheduling metric of a user is calculated based on the ratio of instantaneous condition of the user (e.g. instantaneous data rate) to the averaged condition the user perceived till that moment (e.g. average data rate achieved). The PF scheduler selects a user with maximum priority by

$$j = \arg \max_{1 \le k \le K} D_k(t) / R_k(t-1), \quad (1)$$

where arg max denotes the argument of maximum, k is the index of the user, j is the selected user and K is the total number of users. $D_k(t)$ represents the instantaneous data rate that can be achieved by user k at time t and $R_k(t-1)$ represents the average data rate user k perceived till time $t-1$.

PF scheme favors users which has good channel conditions and high instantaneous data rate, in order to keep system throughput high. However, in the meantime it also considers users with bad channels since these users' low average rate $R_k(t-1)$ will increase their chance to be selected for next scheduling slot. This interesting tradeoff between fairness and performance results in wide acceptance and study of this scheme [1] and [6].

Above and beyond addressing system throughput and fairness, another vital target of scheduling schemes is to guarantee QoS requirements. Fundamental user QoS metrics are delay, bit error rate (BER) and data rate. These metrics are used to guarantee an agreed service level to the user. Depending on applications, QoS metrics to be guaranteed may be different.

In case of non-real time (nRT) traffic (e.g. ftp data, web browsing), BER and data rate have significant impact on user's QoS. BER guaranteeing is handled at the later stage of RRM, by adaptive power allocation. For the data rate, there are a quite large number of researches that study and advance scheduling to guarantee/maximize data rate provided to users in a fair way [1].

For real time (RT) traffic, (e.g. VoIP, live video streaming) in addition to the QoS metrics for nRT traffic, delay is another important metric that need to be taken into account. Special attention should be given to the delay to assure good multimedia performance to users, since packets delayed for long time are disregarded in recovering the RT data stream at the receiver.

Therefore, when a network contains both RT and nRT traffic at the same time, delay constraint must be considered when carrying out throughput maximization.

Only considering scheduling parameters used in PF mechanisms, i.e. $D_k(t)$ and $R_k(t-1)$, to address such extended problem will not be adequate. The reason is that:

these parameters are not taking into account the delay-constraint for RT traffic, e.g. residual time of a packet, which indicates the time offset of a packet to reach its delay threshold. In other words, the time left for a packet to be still useful for a RT application at the receiver. Failure of taking such an important aspect into account will result in poor multimedia performance.

Accordingly, a common practice in literature is to introduce another scheduling policy (i.e. delay-related scheduling policy). Currently, these two scheduling policies are independently used to handle RT (e.g. Delay-related policy) and nRT traffic (e.g. PF policy), as in [8] - [10], and work in isolation to each other as shown in Fig. 1.

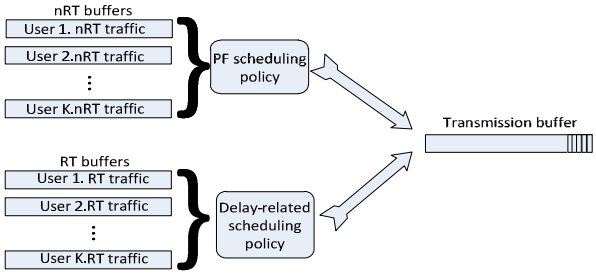


Figure 1. Isolated scheduling policies for nRT and RT traffic.

The scheduling policy for RT traffic mainly aims to guarantee delay QoS satisfaction and that for nRT traffic mainly aims to provide highest system throughput in a fair way. Employing policies in isolation as shown in Fig. 1 or employing only one of the policies to handle both kind of traffic has the following drawbacks:

1) While simply considering guaranteeing delay QoS satisfaction, the policy of RT traffic will weaken the target of throughput maximization. As an example, instead of scheduling a user with highest channel quality – adopting the multiuser diversity in the system – the policy will focus on satisfying delay constraints on RT traffic and can schedule a user with not so good channel but whose packet need to be transmitted to comply with RT traffic delay threshold. That is, the concept of multiuser diversity is undermined. This significantly degrades performance of the system.

2) While simply considering maximizing the system throughput in a fair way, the policy of nRT traffic will ignore delay constraint on RT packets. This can result in high probability of RT packets exceeding their delay threshold. The event that the delay threshold of a RT packet is exceeded is called a QoS violation in this paper. In case of high probability of QoS violation, enjoying multimedia services will not be possible. Furthermore, under the isolated scheduling policies, although the system throughput seems to be maximized, for a RT application, received packets delay-threshold of which is violated become unusable in recovering the whole data stream and considered to be lost. In case of large number of packets experiencing QoS violation, although the system throughput is high, the quality experienced by the user will be greatly degraded and it will be improper to evaluate the system performance by corresponding achieved system throughput.

Another inefficiency in general approach of scheduling mixed nRT and RT traffic [1][6][8]–[10] is to aim achieving highest throughput with no QoS violation. The tight constraint on QoS violation lets the system act conservative and accordingly underutilizes the resources.

Seeing the paradigm shift of Internet users from demanding single application at a time (e.g. solely watching video or web browsing) to the demand of multiple (mixture of RT and nRT) applications simultaneously (e.g. using online social websites and skype chat at the same time or downloading data and watching online video concurrently) the shortcomings mentioned above can be considered as emerging problems to be solved for future high data rate communications.

The aim of this paper is to propose a new scheduling policy to overcome these shortcomings for a mixture of RT and nRT traffic. A new approach of combining the two individual, isolated scheduling policies into a single joint policy with appropriate weights is proposed in this paper. An illustration of the proposed policy is given in Fig. 2. This joint policy will be applied to rank both nRT and RT traffic from all users.

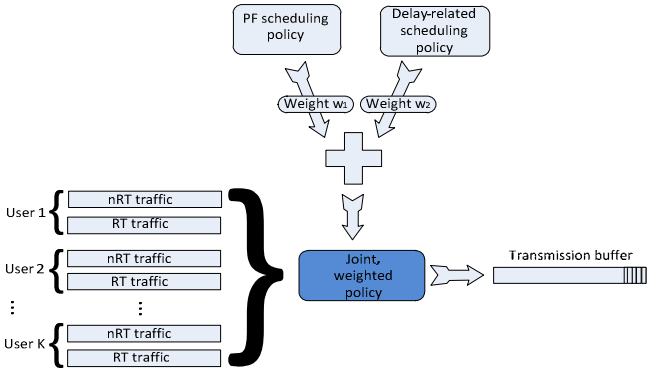


Figure 2. Joint scheduling policy for nRT and RT traffic.

In Fig. 2, the weights w_1 and w_2 are used to trade between severity of QoS violation and achievable throughput. This novel tradeoff can remove drawbacks of isolated approaches and the inefficiency of tight QoS violation constraint. Here, the combination of w_1 and w_2, i.e. (w_1, w_2), is called the scaling vector. With proper scaling vector the consideration of throughput maximization, delay QoS satisfaction and fairness can be enabled. On the other hand, by trading the severity of QoS violation to achievable throughput, the constraint on QoS violation can be relaxed and the system can act in a more opportunistic way to increase its resources' utilization and performance. Such tradeoff have never been introduced and studied in scheduling.

II. System Model

The considered system model is illustrated in Fig. 3. A wireless local area network (WLAN) with single access point

(AP) and K users is considered. Each user is assumed to have two data streams – a RT multimedia stream and an nRT data stream. Accordingly there are $2K$ streams in the system.

To serve user's streams, the AP applies three main functionalities – admission control (AC), scheduling and RRM. AC is the first point of contact for any request from streams. Depending on available resources and security posture it can decide to admit or reject a request. It is assumed that AC is already designed and functionalized in the system.

After AC, the next step is scheduling, where admitted streams are ranked and elected. For proper function, a scheduling mechanism should consider the network's medium access technology. The medium access technology considered in this paper is orthogonal frequency division multiple access (OFDMA) [11]-[13]. OFDMA enables simultaneous transmissions for multiple users by partitioning a broadband channel into a number of orthogonal narrowband subchannels. Each subchannel is allocated to a user to perform simultaneous communications. Here, M subchannels are assumed in the system. OFDMA based scheduling is performed per subchannel basis. That is, for each scheduling attempt, the scheduler ranks all the requests from streams and selects the highest ranked one to occupy a subchannel. The ranking will depend on: 1) delay QoS satisfaction parameter; 2) throughput maximization parameter; and 3) fairness parameter.

The upper limit on the number of streams scheduled at a time – the scheduling group size – depends on the number of subchannels available at the system. Since: 1) the scheduling is performed per subchannel basis and 2) it is assumed the aggregated capacity requirement of all streams do not exceed the total capacity of the system. However, the typical group size is very likely to be less than the number of subcarriers, as more than one subcarrier might be needed to fulfill the data rate requirement of a single stream.

Please notice that although for simplicity an OFDMA based WLAN system is considered as an example in this paper, the proposed policy can be applied to OFDMA based cellular system.

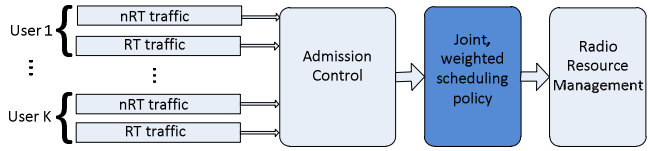


Figure 3. An illustration of system model

After scheduling, the last step to conclude AP's main functionalities is the RRM. The detailed description of similar RRM (i.e. water filling in frequency domain) is presented in [14].

In time domain, the scheduling is slotted into frames of τ ms, called the scheduling period. At the beginning of each period, users' channel state information is made available at the AP and AC then scheduling and RRM are performed. Once three fundamental functions of AP are completed, the resultant resource allocation is feedbacked to the users. In the rest of the period time, scheduled users transmit their traffic simultaneously.

In channel model, all subchannels are considered to be Rayleigh fading channels that introduce an additive white Gaussian noise (AWGN).

III. Joint, Weighted Scheduling Policy

This section will study the proposed scheduling policy in details. Referring back to the main aim of removing shortcomings of applying scheduling policies in isolation, the proposed policy unites isolated policies – PF policy and delay-related policy – into a single joint policy. For the mathematical formulation of the former policy, simple yet effective expression, expression (1), will be used. And for the latter policy, the popular delay-related QoS metric – residual time of a packet – will be employed. To define residual time metric, the ratio of a packet's time in the system to the delay threshold of the packet is given by

$$\textit{residual time metric} = d_{k,j} \big/ d_{k,j}^{th} \tag{2}$$

where $d_{k,j}$ is the delay encountered by a packet at the head of the jth stream of the kth user. And $d_{k,j}^{th}$ represents the delay threshold for packets at the jth stream of the kth user.

The joint scheduling is determined to be summation of individual policies with associated weights w_1 and w_2, as follows

$$j = \max \underset{\substack{1\le k\le K \\ 0\le j\le 1}}{\arg} \left\{ w_1 \left(\frac{d_{k,j}}{d_{k,j}^{th}} \right) + w_2 \left(\frac{D_{k,j}(t)}{R_{k,j}(t-1)} \right) \right\} \tag{3}$$

where $j = 0$ represents the RT stream and $j = 1$ represents the nRT stream. w_1 is the weight for first policy and w_2 is the weight for the second policy. $D_{k,j}(t)$ represents the instantaneous data rate that can be achieved by stream j of user k at time t and $R_{k,j}(t-1)$ represents the average data rate that stream j of user k perceived till time $t-1$.

Note that, when determining the scheduling metric of a stream, this formula considers; the delay encountered by corresponding packet of the stream, the delay threshold need to be satisfied by the packet, the instantaneous data rate that can be achieved by that stream and the average data rate the stream have perceived till that moment. Such formulation can target requirements of RT and nRT traffic.

Furthermore, the concept of joint weighted scheduling introduces a novel performance-enhancing tradeoff to scheduling. The basic idea is to allow a desired, tolerable severity of QoS violation to let the system be more opportunistic in resource allocation and accordingly improve its resources' utilization. The higher the severity of QoS violation, the more opportunistic the system will be to utilize

its resources and increase its performance.

Weight parameters w_1 and w_2 have been exploited for introducing the QoS violation to scheduling. Actually the weight parameters stand for the impact of an individual policies on the joint policy. For example, larger w_1 means larger impact of delay-related scheduling policy on the joint policy. That is, joint policy will satisfy the delay constraints with little focus on throughput maximization (i.e. the system will perform conservatively). A system with larger w_1 will most probably have lower throughput.

Similarly, larger w_2 means larger impact of PF scheduling policy on the joint policy. That is, the joint policy will maximize throughput with less focus on delay constraints of packets. Accordingly, higher achievable data rate as well as severity of delay QoS violation will result from larger w_2. It is very likely that, a system with larger w_2 will not be able to well transmit multimedia traffic.

So the values of weight parameters must be carefully set. Comprehensive simulation results will be shown in the next section to reveal the accurate scaling vector that can make the proposed policy truly serve required fairness, throughput maximization and QoS satisfaction purposes.

IV. Simulation Results

A. Simulation Performance Metrics

In the simulation of the proposed scheduling policy, four performance metrics – total system throughput, delay QoS violation, data rate fairness index and QoS violation fairness index – will be considered.

The *total system throughput* metric characterizes how efficient the policy is at scheduling traffic that maximizes the system throughput, defined as

$$\textit{Total system throughput} = \sum_{k=1}^{K}\sum_{q=0}^{1}\tilde{b}_{k,q} \tag{4}$$

where $\tilde{b}_{k,q}$ [14] is the average number of bits achieved by the qth stream of the kth user.

On the other hand, the *delay QoS violation ratio* metric represents how effective the policy is at ranking traffic to avoid delay threshold violations, given by

$$\textit{Delay QoS violation ratio} = \sum_{k=1}^{K}\eta_k \Big/ U \tag{5}$$

where η_k represents the total number of packets lost from RT stream of user k and U is the total number of packets generated from all RT streams. The *Delay QoS violation ratio* is used as the index of the severity of QoS violation.

The investigation for fairness of the proposed policy will be done from two perspectives: 1) fairness to the data rate achieved per user and 2) fairness to the QoS violation experienced per user. The same expression will be used to measure both of the fairness indexes. However, the variables of the expression will be different. The normalized variance of users' achieved data rates will provide the *Data rate fairness index* and the normalized variance of users' experienced QoS violations will provide the *QoS violation fairness index*, given by

$$F(x_1, x_2, ..., x_i) = \frac{E[(X - E(X))^2]}{[E(X)^2]} = \frac{\frac{1}{K}\sum_{i=1}^{K}\left(X_i - \frac{1}{K}\sum_{k=1}^{K}X_k\right)^2}{\left(\frac{1}{K}\sum_{k=1}^{K}X_k\right)^2} \tag{6}$$

where x_i is the data rate achieved by user i or QoS violation user i experienced, respectively for the first and second fairness index. The values of fairness indexes are lower bounded by zero and upper bounded by one, where zero indicates that all users are served equally.

B. Simulation Results

The simulated WLAN environment assumes that the wireless broadband channel is divided into M = 12 orthogonal narrowband subchannels. The average signal to noise ratio (SNR) on each subcarrier is 20 dB.

There is one AP in the environment which serves K = 8 users. Each user has one nRT and one RT stream, i.e. there are in total 16 streams (8 nRT + 8 RT streams) in the system. The packet inter-arrival from nRT and RT streams are modeled as exponentially distributed arrival processes and the mean values are setup to heavily load the network. This is done to investigate the functioning and performance of the proposed scheduling policy in congested network scenario. The benefit of the policy to satisfy QoS requirements and performance gain should be investigated at congested network. Also, the load of nRT and RT traffic is set to be equal to have same conditions while measuring fairness of the policy from two perspectives.

For RT traffic the delay threshold is set to be three times the scheduling period. For example, if the scheduling period of the system is 10ms, then the delay threshold is set to be 30ms. That is, a packet that is not scheduled at three consecutive scheduling intervals will be dropped from the corresponding stream's queue. This is because in RT communications, packets arriving too late will be disregarded (e.g. to display video) at the receiver. In case of the nRT traffic the delay threshold is not used, and it is set to be infinite.

To investigate a vital aspect of the proposed policy – determining the scaling vector between individual policies – a series of experiments were conducted. The goal is to analyze the system performance metrics with respect to values of w_1 and w_2 and find out an optimal scaling vector. Values of w_1 and w_2 are bounded to be $0 \le w_1 \le 1$ and $0 \le w_2 \le 1$ and

constrained by $w_1 + w_2 = 1$, (e.g. $w_1 = 0$ and $w_2 = 1$ represent conventional PF scheduling).

The main simulation results are illustrated in Figs. 4 – 7.

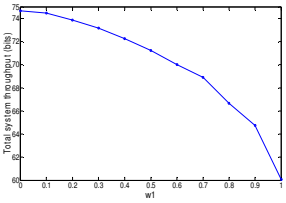


Figure 4. Total system throughput versus weight parameter w_1.

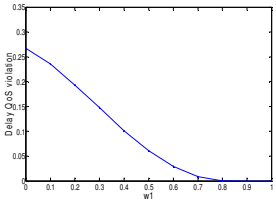


Figure 5. Delay QoS violation versus weight parameter w_1.

From these results, the following observations, explanations and decision can be concluded.

Observation 1. Clearly, the weight parameters w_1 and w_2 have severe impact on all chosen performance metrics.

Observation 2. As Fig. 4 demonstrates, the highest system throughput can be achieved when the value of w_1 is zero (i.e. value of w_2 is one). According to (3), the focus of joint policy is to maximize the throughput in a fair way. And it completely ignores the delay QoS satisfaction for RT traffic. The joint policy actually performs the same as the isolated/conventional PF scheduling policy. According to Fig. 5, it is clearly seen that there is more than 25% delay violation in RT traffic for such scheduling policy. No RT application can properly function at this severity of violation. And more, the system throughput shown in Fig. 4 is impractical, since the throughput of RT traffic will have no use now.

Explanation 1. The reason for observation 2 is that: when $w_1 = 0$ the scheduling metric of a stream now only depends on $D_{k,j}(t)$ and $R_{k,j}(t-1)$, where $d_{k,j}$ and $d_{k,j}^{th}$ are completely ignored. In this case, the higher levels of the ranking list will be occupied by streams which belong to users with higher SNR and/or that have achieved lower data rate so far, instead of stream's that has packets with commencing delay thresholds.

Observation 3. From the other extreme case of $w_1 = 1$ and $w_2 = 0$, the focus of the joint policy is to satisfy delay QoS requirement of RT traffic. This intention of the policy is the same as the intention of isolated Delay-related policy. Fig. 5 shows the success of the policy in achieving zero QoS violence. Yet, referring to Fig. 4 when $w_1 = 1$ and $w_2 = 0$, it is seen that there is around 20% loss in the total system throughput when compared to the PF scheduling policy. So being very conservative results in a significant loss in achievable system throughput.

Explanation 2. The reason for observation 3 can be explained in the similar way to that for observation 2: when $w_1 = 1$ the scheduling metric provided to a stream only depends on $d_{k,j}$ and $d_{k,j}^{th}$, while $D_{k,j}(t)$ and $R_{k,j}(t-1)$ are completely ignored. Now, the higher levels of the ranking list will be occupied by streams that have RT packets with commencing delay threshold.

Explanation 3. The result for such significant loss, 20%, in observation 3 lies under the extreme case of completely focusing on delay QoS satisfaction. This statement can be clarified by Fig.4, where it is seen that even if a little focus is given to throughput maximization, when $w_1 = 0.9$ and $w_2 = 0.1$, the loss in performance reduces to around 12% with no QoS violation presented at Fig. 5. The reason for this nearly twofold improvement is as follows: now even the dominating variables in streams' rank are $d_{k,j}$ and $d_{k,j}^{th}$, still $D_{k,j}(t)$ and $R_{k,j}(t-1)$ have some impact on the last outcome of the policy. Hence, although the streams at the high levels of the ranking list will still be mainly depending on delay-related concerns, among these high level streams the top rank will be given to the stream that have better SNR and/or achieved low data rate at the past.

Observation and Explanation 4. The fairness metric on user's achieved data rate is shown in Fig. 6. The behavior of this metric can be explained as follows: up to $w_1 = 0.4$ there is nearly complete fairness in the system, this is due to dominating impact of throughput related concerns $D_{k,j}(t)$ and $R_{k,j}(t-1)$ in the ranking formula. This shows that with appropriate scaling vector the policy can achieve nearly complete fairness with respect to data rate. For $0.4 \le w_1 \le 0.6$, there is a linear increase in the system fairness value by 0.08, and the actual fairness perceived by users is decreasing. The reason for this decrease in fairness is that: throughput related concerns are losing their domination and the impact of delay-related concerns $d_{k,j}$ and $d_{k,j}^{th}$ is getting more and more. In the range of $0.6 \le w_1 \le 0.9$ the dominant concern of the ranking formula becomes delay-related concerns now and there is less importance paid to fair throughput maximization, hence the fairness metric becomes steady in this range but at a relatively high value of 0.08. When $w_1 = 1$ the value of fairness has a sharp jump to 0.17. The reason for this sharp jump is the same mentioned as in comment 3; the extreme case of completely ignoring the throughput related concerns in the ranking formula.

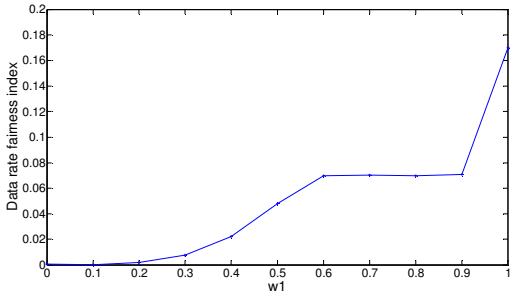


Figure 6. Data rate fairness versus weight parameter w_1.

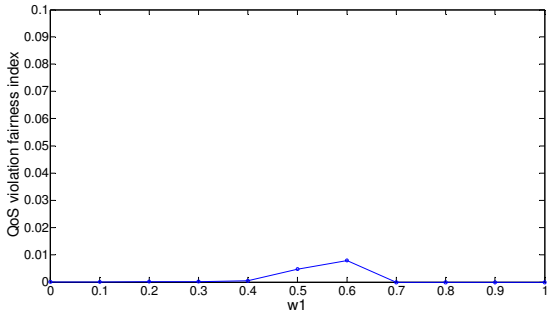


Figure 7. QoS violation fairness versus weight parameter w_1.

Observation and Explanation 5. The fairness metric on user's delay QoS violation is shown in Fig. 7. Up to $w_1 = 0.4$ (similar to Fig. 6) there seems to have almost complete fairness in the system with respect to delay QoS violation. This shows that the severity for delay violation is similar among users. Yet, checking Fig. 5, it can be seen that this similar severity of violation is at high level, which is due to domination of throughput related concerns. From $0.4 \le w_1 \le 0.7$ there is a slight decrease in the fairness, which is due to having more balanced scaling between throughput-related and delay-related concerns in the ranking formula. The metric of a user is now affected by $d_{k,j}$, $d_{k,j}^{th}$, $D_{k,j}(t)$ and $R_{k,j}(t-1)$, instead of being dominated by either first or last two parameters. This can slightly change the severity of delay violation experienced among users. When $w_1 \ge 0.7$ the system come back to complete fairness. This is due to domination of $d_{k,j}$ and $d_{k,j}^{th}$ in the ranking formula so that, referring to Fig. 5, there is almost no delay violation in the system.

Decision. According to the obtained results and presented tradeoff concept between QoS violation and achievable throughput, this study suggests the scaling vector to be set at $(w_1, w_2) = (0.7, 0.3)$. With this scaling vector the system can guarantee 1% delay QoS violation by only 6% difference from the highest achievable throughput. And from both perspectives of the fairness, the system is considered to be at acceptable level of fairness. The suggestion is based on the constraint of 1% packet loss at transmission of RT video. If compared to isolated Delay-related scheduling policy, the proposed policy can successfully handle RT traffic besides nRT traffic by sacrificing much less achievable performance.

V. Conclusion

A joint, weighted scheduling policy has been developed and a novel tradeoff concept is introduced to the scheduling. The inter-relationship between combined scheduling policies is revealed and optimal weight parameters have been examined through a series of simulation results. The proposed policy with introduced tradeoff can be used for efficient scheduling of mixed RT and nRT traffic in emerging wireless networks. Although OFDMA is assumed in the proposed policy, it can be applied to other multiple access techniques [15][16].

Acknowledgement

The research leading to these results has received funding from the EU IST Seventh Framework Programme ([FP7/2007-2013]) under grant agreement no 25741, project ULOOP (User-centric Wireless Local Loop) and NTT DOCOMO, JAPAN.

References

[1] E. Yaacoub and Z. Dawy, 'A Survey on Uplink Resource Allocation in OFDMA Wireless Networks', IEEE Communications Surveys & Tutorials, vol. PP, no. 99, 2011, pp. 1–16.

[2] S. H. Y. Wong, S. Lu, H. Yang, and V. Bharghavan, 'Robust rate adaptation for 802.11 wireless networks', in MobiCom '06, Los Angeles, CA, USA, 2006, p. 146.

[3] B. Xia and J. Wang, "Effect of channel estimation error on QAM systems with antenna diversity," IEEE Transactions on Communications, vol. 53, no. 3, pp. 481-488, March 2005.

[4] Y. Q. Zhou, J. Wang and M. Sawahashi, "Downlink transmission of broadband OFCDM systems Part I: hybrid detection," IEEE Transactions on Communications, vol. 53, no. 4, pp. 718-729, April 2005.

[5] J. Wang, H. Zhu, and N. Gomes, "Distributed antenna systems for mobile communications in high speed trains," IEEE Journal on Selected Areas in Communications, vol. 30, pp. 675-683, May 2012.

[6] A. Gyasi-Agyei and S. Kim, 'Comparison of opportunistic scheduling policies in time-slotted AMC wireless networks', in 1st International Symposium on Wireless Pervasive Computing, 2006.

[7] A. Jalali, R. Padovani, and R. Pankaj, 'Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system', in VTC 2000-Spring, 2000, vol. 3, pp. 1854–1858.

[8] J. He, K. Yang, K. Guild, H. Chen. "On Bandwidth Request Mechanism with Piggyback in Fixed IEEE 802.16 Networks", IEEE Transactions on Wireless Communications, Vol. 7, No. 12, December 2008. Pages: 5238-5243.

[9] K. Yang, S. Ou, K. Guild, HH Chen. "Convergence of Ethernet PON and IEEE 802.16 broadband access networks and its QoS-aware dynamic bandwidth allocation scheme", IEEE JSAC, Volume 27, Number 2, February 2009: 101-116.

[10] Q. Liu, X. Wang, and G. B. Giannakis, 'A cross-layer scheduling algorithm with QoS support in wireless networks', IEEE Transactions on Vehicular Technology, vol. 55, no. 3, pp. 839–847, May 2006.

[11] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - Part I: chunk allocation," IEEE Transactions on Communications, vol. 57, no. 9, pp. 2734-2744, Sept. 2009.

[12] H. Zhu and J. Wang, "Chunk-based resource allocation in OFDMA systems - Part II: joint chunk, power and bit allocation," IEEE Transactions on Communications, vol. 60, pp. 499-509, no. 2, Feb. 2012.

[13] H. Zhu, "Radio Resource Allocation for OFDMA Systems in High Speed Environments," IEEE Journal on Selected Areas in Communications, vol. 30, pp. 748-759, May 2012.

[14] H. Haci, H. Zhu, and J. Wang, 'Resource Allocation in User-Centric Wireless Networks', in VTC2012-Spring, Yokohama, Japan, 2012.

[15] J. Wang and L. B. Milstein, "CDMA overlay situations for microcellular mobile communications," IEEE Transactions on Communications, vol. 43, no. 2/3/4, pp. 603-614, Feb/March/April 1995.

[16] J. Wang and J. Chen, "Performance of wideband CDMA systems with complex spreading and imperfect channel estimation," IEEE Journal on Selected Areas in Communications, vol. 19, no. 1, pp. 152-163, Jan. 2001.