IEEE
COMMUNICATIONS
MAGAZINE

February 2016, Vol. 54, No. 2

- Underwater Wireless Communications
- Optical Communications
- Automotive Networking

IEEE
IEEE COMMUNICATIONS SOCIETY
A Publication of the IEEE Communications Society
www.comsoc.org

IEEE Communications MAGAZINE

THANKS OUR CORPORATE SUPPORTERS

SAMSUNG

Anritsu

Test and Measurement Solutions

BEEcube

CISCO™

R&S ROHDE&SCHWARZ

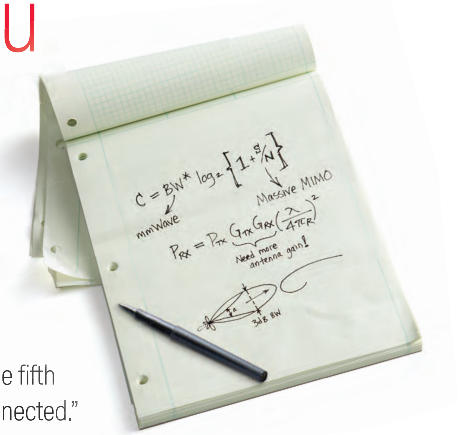
KEYSIGHT TECHNOLOGIES

IEEE
COMMUNICATIONS
MAGAZINE

February 2016, Vol. 54, No. 2

- Underwater Wireless Communications
- Optical Communications
- Automotive Networking

IEEE
IEEE COMMUNICATIONS SOCIETY
A Publication of the IEEE Communications Society
www.comsoc.org



$C = BW^* \log_2 \left\{ \left[1 + S/N \right] \right\}$
mmWave
Massive MIMO
$P_{RX} = P_{TX} \underbrace{G_{TX} G_{RX}}_{\text{Need more antenna gain!}} \left(\frac{\lambda}{4\pi R} \right)^2$
a
3dB BW

IEEE COMMUNICATIONS MAGAZINE

FEBRUARY 2016, vol. 54, no. 2
www.comsoc.org/commag

UNDERWATER WIRELESS COMMUNICATIONS AND NETWORKS – THEORY AND APPLICATION: PART 2

GUEST EDITORS: XI ZHANG, JUN-HONG CUI, SANTANU DAS, MARIO GERLA, AND MANDAR CHITRE

ADVANCES IN OPTICAL COMMUNICATIONS NETWORKS

SERIES EDITORS: OSMAN GEBIZLIOGLU AND VIJAY JAIN

AUTOMOTIVE NETWORKING AND APPLICATIONS

SERIES EDITORS: WAI CHEN, LUCA DELGROSSI, TIMO KOSCH, AND TADAO SAITO

IEEE COMMUNICATIONS MAGAZINE (ISSN 0163-6804) is published monthly by The Institute of Electrical and Electronics Engineers, Inc. Headquarters address: IEEE, 3 Park Avenue, 17th Floor, New York, NY 10016-5997, USA; tel: +1 (212) 705-8900; http://www.comsoc.org/commag. Responsibility for the contents rests upon authors of signed articles and not the IEEE or its members. Unless otherwise specified, the IEEE neither endorses nor sanctions any positions or actions espoused in *IEEE Communications Magazine*.

ANNUAL SUBSCRIPTION: $27 per year print subscription. $16 per year digital subscription. Non-member print subscription: $400. Single copy price is $25.

EDITORIAL CORRESPONDENCE: Address to: Editor-in-Chief, Osman S. Gebizlioglu, Huawei Technologies, 400 Crossing Blvd., 2nd Floor, Bridgewater, NJ 08807, USA; tel: +1 (908) 541-3591, e-mail: Osman.Gebizlioglu@huawei.com.

COPYRIGHT AND REPRINT PERMISSIONS: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. Copyright law for private use of patrons: those post-1977 articles that carry a code on the bottom of the first page provided the per copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint, or republication permission, write to Director, Publishing Services, at IEEE Headquarters. All rights reserved. Copyright © 2016 by The Institute of Electrical and Electronics Engineers, Inc.

POSTMASTER: Send address changes to *IEEE Communications Magazine*, IEEE, 445 Hoes Lane, Piscataway, NJ 08855-1331. GST Registration No. 125634188. Printed in USA. Periodicals postage paid at New York, NY and at additional mailing offices. Canadian Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40030962. Return undeliverable Canadian addresses to: Frontier, PO Box 1051, 1031 Helena Street, Fort Eire, ON L2A 6C7.

SUBSCRIPTIONS: Orders, address changes — IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08855-1331, USA; tel: +1 (732) 981-0060; e-mail: address.change@ieee.org.

ADVERTISING: Advertising is accepted at the discretion of the publisher. Address correspondence to: Advertising Manager, *IEEE Communications Magazine*, 3 Park Avenue, 17th Floor, New York, NY 10016.

SUBMISSIONS: The magazine welcomes tutorial or survey articles that span the breadth of communications. Submissions will normally be approximately 4500 words, with few mathematical formulas, accompanied by up to six figures and/or tables, with up to 10 carefully selected references. Electronic submissions are preferred, and should be sumitted through Manuscript Central: http://mc.manuscriptcentral.com/commag-ieee. Submission instructions can be found at the following: http://www.comsoc.org/commag/paper-submission-guidelines. For further information contact Zoran Zvonar, Associate Editor-in-Chief (zoran.zvonar@mediatek.com). All submissions will be peer reviewed.

ACCEPTED FROM OPEN CALL

CURRENTLY SCHEDULED TOPICS

TOPIC	ISSUE DATE	MANUSCRIPT DUE DATE
ENABLING MOBILE AND WIRELESS TECHNOLOGIES FOR SMART CITIES	DECEMBER 2016	FEBRUARY 29, 2016
NEXT GENERATION 911	NOVEMBER 2016	MARCH 15, 2016
INTEGRATED COMMUNICATIONS, CONTROL, AND COMPUTING TECHNOLOGIES FOR ENABLING AUTONOMOUS SMART GRID	DECEMBER 2016	APRIL 1, 2016
NEW WAVEFORMS AND MULTIPLE ACCESS METHODS FOR 5G NETWORKS	NOVEMBER 2016	APRIL 2, 2016
IMPACT OF NEXT-GENERATION MOBILE TECHNOLOGIES ON IOT-CLOUD CONVERGENCE	JANUARY 2017	APRIL 15, 2016
PRACTICAL PERSPECTIVES ON IOT IN 5G NETWORKS: FROM THEORY TO INDUSTRIAL CHALLENGES AND BUSINESS OPPORTUNITIES	FEBRUARY 2017	MAY 1, 2016

www.comsoc.org/commag/call-for-papers

TOPICS PLANNED FOR THE MARCH ISSUE

CRITICAL COMMUNICATIONS AND PUBLIC SAFETY NETWORKS

RADIO COMMUNICATIONS

NETWORK TESTING

COMMUNICATIONS STANDARDS SUPPLEMENT
SEMANTICS FOR ANYTHING-AS-A-SERVICE

FROM THE OPEN CALL QUEUE

DEVICE-TO-DEVICE BROADCASTING COMMUNICATIONS IN BEYOND 4G CELLULAR NETWORKS

STANDARDS FOR MEDIA SYNCHRONIZATION

NETWORK FUNCTIONS VIRTUALIZATION IN 5G

WHY TO DECOUPLE THE UPLINK AND DOWNLINK IN CELLULAR NETWORKS AND HOW TO DO IT

There's a Better Way to Design 5G Wireless Systems

It starts with an integrated design approach for rapid prototyping.

In the race to 5G, researchers need to accelerate design cycles. With the LabVIEW Communications System Design Suite and NI software defined radio hardware, they can build 5G prototypes fast to decrease the time between an idea and a real-world application.

>> See how at ni.com/5g

800 891 8841

NATIONAL INSTRUMENTS™

ComSoc Preparing for the Future

Harvey Freeman

J. Roberto de Marca

In this month's column, I would like to focus on ComSoc's strategy and goals for the next few years. IEEE and the Communications Society hold retreats every January to identify issues we need to understand and resolve in the next year or several years. The Board then tasks the respective Strategic Planning Committees to develop strategies and actionable plans to address these issues. During the ComSoc Management Retreat, the first stage of this process was completed and turned over to the 2016-17 ComSoc Strategic Planning Committee headed by Roberto de Marca.

J. Roberto de Marca was a Fulbright Scholar at the University of Southern California, where he earned a Ph.D. in Electrical Engineering. He was the 2014 IEEE President and CEO. He was also the 2000-2001 President of the IEEE Communications Society and the founding President of the Brazilian Telecommunications Society. Roberto is an IEEE Fellow and a full member of both the Brazilian Academy of Sciences and Brazilian National Academy of Engineering. He has extensive international experience, having held visiting appointments in several organizations, including AT&T Bell Laboratories, NEC Research Labs Europe, Politectnico di Torino, Italy, and Hong Kong University of Science and Technology. Roberto is currently the Chair of the Communications Society's Strategic Planning Committee and has prepared this column to share his insights.

The IEEE Communications Society has been the premier technical organization serving the communication professional community for several decades. As with any other successful enterprise, it faces big challenges to continue ahead of the competition in the future pursuing its vision of being the organization of choice for communications and related professionals throughout the world. Our world has been changing at a very fast pace, and so has technology. These changes in global society behavior have been mostly due to the successes of the communications engineering community, with the continued evolution of Internet services associated with and leveraged by the constant development of wireless technology, resulting in the availability of a true powerful personal computer (cellular terminal) available to the majority of the world population. However, the enormous success of the work of our community also creates many challenges for ComSoc. In 2012, the ComSoc leadership developed a report titled "ComSoc 2020" that outlined these challenges that still remain valid for the next five, 10, and 15 years, but if anything, have become more demanding. Here is a quote from that report. "As we face the next decade we need to understand the shift that will progressively empower communities of individuals. ComSoc has to become a community of communities whose existence "half-life" will vary considerably and will relate more and more to their effectiveness in meeting community needs. As the paradigm shifts from technology to market to social/society, we are adding value like a pyramid. The value of technology that entails is not lost because the perception and need has moved to market. … They will simply be taken for granted and ComSoc shall not abandon its core values but rather it should add on to these." This report went on to propose Goals for 2020, and it is worthwhile for the sequel to list some of these goals:

- Launch a neutral peer-reviewed yearly document on technology evolution and its potential implications on the market and human society.
- Promote through its conferences and publications a view of technology within the broader context of market forces/services and human society.
- Become the organization that policy makers will turn to for an understanding of technology and its evolution.
- Move quickly toward supporting a generation of freelance professionals.
- Feature a flexible structure able to respond quickly, and even anticipate the dynamics of markets and societies as the pace of evolution will quicken.

It turns out these goals are very much aligned with the results of the strategic effort developed in 2014-2015 by our parent organization, IEEE, and which resulted in the following goals and actions for the period 2015–2020:

- Expand and enable nimble, flexible, disband-able communities to help individuals from around the world to share, collaborate, network, debate, and engage with one another.
- Provide technically vital forums for the discussion, development, and dissemination of authoritative knowledge related to traditional technology while focusing on better serving the professionals working on emerging and disruptive technologies.
- Provide more opportunities, products, and services aimed at increasing IEEE's value to professionals working in industry, particularly young professionals and entrepreneurs.
- Leverage IEEE's technology related insight to provide governments, NGOs, and other organizations and the public with innovative, practical recommendations to address public policy issues.

- Lead humanitarian efforts around the world to use technology to solve the world's most challenging problems.

Combining these sets of goals, there is a clear view how the Society has to evolve in the next 10 years. It has to develop activities, products, and services that are considered relevant by communities of professionals that will have broad scope and dynamically changing interests, and that will have available a plethora of social media platforms constantly available for information digging and professional collaboration. This overarching goal is, of course, easier said than done. There are many difficult hurdles and uncertainties in every direction and in every piece of the overall challenge. If we consider Publications, for example, there are several questions that need to be answered. Will it be possible to continue to package and sell information in the way it is done today? If not, ComSoc's business model will need a dramatic change. How will the information be vetted? Will peer review (partially) change its nature? Is there a role for crowd sourcing? Can new results correspond to new versions of an "evolving article?" Are commenters becoming "authors" of a new (improved) version of the article? Will ComSoc be able to offer industry professionals the knowledge they need (and demand) to solve problems? How can we produce and disseminate this information?

Conferences also seem to require changes in format to remain attractive in the future. Actually, there are new styles of convening professionals being used very successfully. One example is the South by Southwest conference held each year in Austin, Texas, USA. The program of this event includes a broad spectrum of topics and speakers, and the content seems to be less attractive than who is giving the talk. For young professionals, social media appears to drive the face-to-face networking. Can we leverage this trend? What will be the role of on-line conferences and what format will they take.

The ComSoc leadership met on January 16-17 to wrestle in a retreat setting with these and other topics related to the future of ComSoc. One of the current acute problems for ComSoc is the significant decrease in its membership in the past few years despite the growth of professionals worldwide that develop activities in some way related to communications. It seems clear that the Society must provide more value to its members. The key question is what to offer. There is a perception, which is also valid for all of IEEE, that ComSoc serves reasonably well the members that work in academia, but does not do a similar good job for industry professionals ("practitioners"), in particular the young generation working in the services industry. It is important to understand from the employers what kind of products/services they will find of value to their employees, and in some cases would even allow them to spend time helping to develop their content. Answers to this question are not easily obtained because often industry leaders themselves do not know the answer, but it is essential to work closely with them to implement successful products. Often the notion of a corporate package that could include bulk membership is mentioned as a desirable offering, particularly outside the USA, but again the definition of the services included in the package is far from being clear.

Some offerings discussed as very good candidates for implementation came from another internal ComSoc document, "SPC Business Plans", developed by the 2014-2015 ComSoc Strategic Planning Committee, chaired by Past President Byeong Gi Li. This document states: "To date, ComSoc has not done enough to capture the shift of the market from the classic telecommunications industry to Internet companies and processing infrastructures. Also, we have not paid enough attention to the pervasiveness of telecommunications that has entered into many vertical sectors. Many non-communications people in different sectors have joined the enlarged ICT world". Then it went on to propose the creation of ComSoc's One-stop ICT Service System that would contain a Knowledge Base in ICT that could be used for both industry professionals and academics. Another proposal coming from this report is the creation of a user friendly Education Portal that would provide access to a world class Training and Professional Education program, addressed primarily to ComSoc members, which provides high quality instruction, at reasonable cost, and with easy access, to address the career needs of industry professionals in communications and related fields. Education has always been a challenging area for ComSoc, but it is also felt that successful products in this area can go a long way to attract new members, and also provide much needed services to young and mid-career industry professionals.

One most important factor for ComSoc staying relevant is to make sure that it continues to attract the thought leaders in the emerging trends and technologies related to its field. Accordingly, ComSoc is now leading activities within IEEE in 5G wireless technology, Internet of Things, Software Defined Networks, and Green ICT. These efforts should lead to new publications, conferences, and portals that should attract more members and volunteers and, as a consequence, bring additional revenue to sustain the Society's activities.

Last month's Management Retreat also dealt with changes in the ComSoc organization that would facilitate successfully addressing the challenges described. These changes should be implemented as soon as possible, most likely this June.

As can be seen, there is much work to be done in 2016 and 2017 to guarantee a bright future for the Society. One essential component of this success is to continue to attract top-notch volunteers and keep them engaged in developing high quality new products and services. Developing a structure that will keep these volunteers motivated, in particular the young volunteers, becomes an important strategic element for the Society.

We welcome suggestions and comments from our readers on how ComSoc should look in 2025.

UPDATED ON THE COMMUNICATIONS SOCIETY'S WEB SITE

www.comsoc.org/conferences

2016

FEBRUARY

IEEE BHI 2016 — IEEE Int'l. Conference on Biomedical and Health Informatics, 24–27 Feb.
Las Vegas, NV
http://bhi.embs.org/2016/

MARCH

DRCN 2016 — 12th Int'l. Workshop on Design of Reliable Communication Networks, 14–17 Mar.
Paris, France
https://drcn2016.lip6.fr/

ICBDSC 2016 — 3rd MEC Int'l. Conference on Big Data and Smart City, 15–16 Mar.
Muscat, Oman
http://www.mec.edu.om/conf2016/index.html

OFC 2016 — Optical Fiber Conference, 20–24 Mar.
Anaheim, CA
http://www.ofcconference.org/en-us/home/

IEEE ISPLC 2016 — 2016 IEEE Int'l.Symposium on Power Line Communications and Its Applications, 21–23 Mar.
Bottrop, Germany
http://www.ieee-isplc.org/

IEEE CogSIMA 2016 — IEEE Int'l. Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support, 21–25 Mar.
San Diego, CA
http://www.cogsima2016.org/

WD 2016 — Wireless Days 2016, 23–25 Mar.
Toulouse, France
http://wd2015.sciencesconf.org/

APRIL

IEEE WCNC 2016 — IEEE Wireless Communications and Networking Conference, 3–6 Apr.
Doha, Qatar
http://wcnc2016.ieee-wcnc.org/

IEEE INFOCOM 2016 — IEEE Int'l. Conference on Computer Communications, 10–15 Apr.
San Francisco, CA
http://infocom2016.ieee-infocom.org/

WTS 2016 — Wireless Telecommunications Symposium, 18–20 Apr.
London, U.K.
http://www.cpp.edu/~wtsi/

IEEE/IFIP NOMS 2016 — IEEE/IFIP Network Operations and Management Symposium, 25–29 Apr.
Istanbul, Turkey
http://noms2016.ieee-noms.org/

MAY

IEEE CQR 2016 — IEEE Int'l. Communications Quality and Reliability Workshop, 9–12 May
Stevenson, WA
http://www.ieee-cqr.org/

ONDM 2016 — Int'l. Conference on Optical Network Design and Modeling, 9–12 May
Cartagena, Spain
http://ondm2016.upct.es/index.php

IEEE CTW 2016 — IEEE Communication Theory Workshop, 15–18 May
Nafplio, Greece
http://www.ieee-ctw.org/

ICT 2016 — Int'l. Conference on Telecommunications, 16–18 May
Thessaloniki, Greece
http://ict-2016.org/

IEEE ICC 2016 — IEEE International Conference on Communications, 23–27 May
Kuala Lampur, Malaysia
http://icc2016.ieee-icc.org/

JUNE

IEEE BlackSeaCom 2016 — 4th Int'l. Black Sea Conference on Communications and Networking, 6–9 June
Varna, Bulgaria
http://www.ieee-blackseacom.org/

IEEE NETSOFT — IEEE Conference on Network Softwarization, 6–10 June
Seoul, Korea
http://sites.ieee.org/netsoft/

IEEE LANMAN 2016 — 22nd IEEE Workshop on Local & Metropolitan Area Networks, 13–15 June
Rome, Italy
http://www.ieee-lanman.org/

IEEE HPSR 2016 — IEEE 17th Int'l. Conference on High Performance Switching and Routing, 14–17 June
Yokohama, Japan
http://www.ieee-hpsr.org/

IEEE IWQOS — IEEE Int'l. Symposium on Quality and Service, 20–21 June
Beijing, China
http://www.dongliangxie.com/

MED-HOC-NET — Mediterranean Ad Hoc Networking Workshop, 20–22 June
Vilanova I la Geltru, Spain
http://craax.upc.edu/medhocnet2016/

EUCNC 2016 — European Conference on Networks and Communications, 27–30 June
Athens, Greece
http://eucnc.eu/

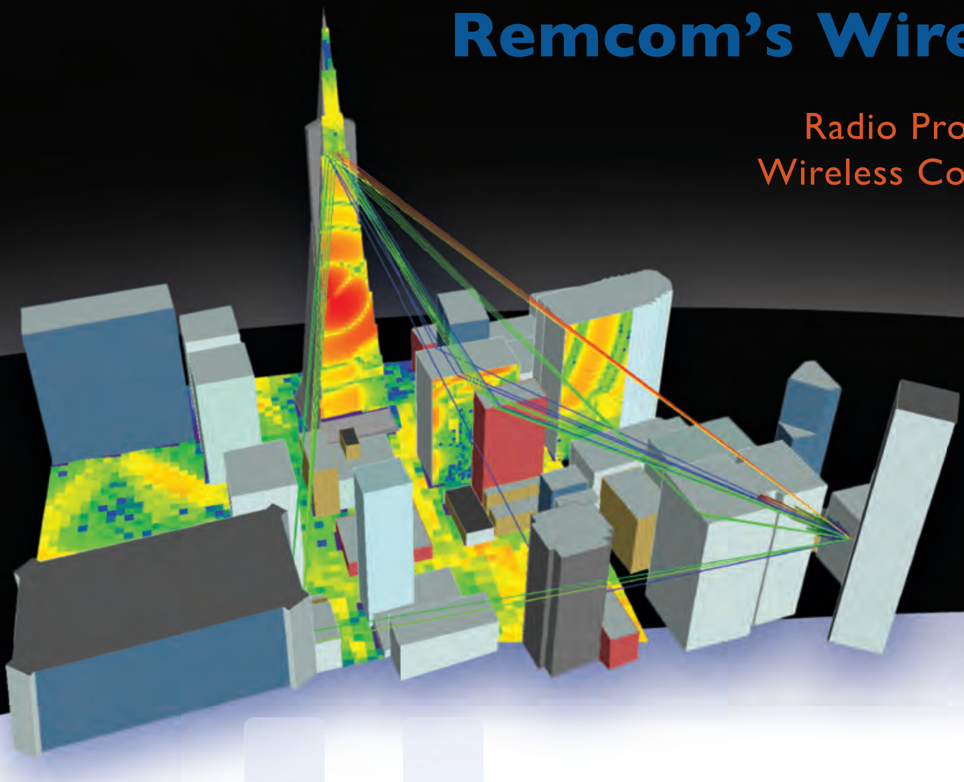
IEEE ISCC — Int'l. Symposium on Computers and Communications, 26–30 June
Messina, Italy
http://iscc2016.unime.it/

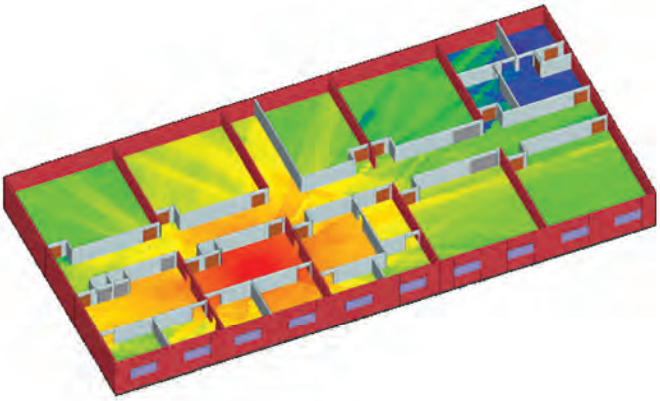
JULY

ICUFN 2016 — Int'l. Conference on Ubiquitous and Future Networks, 5–8 July
Vienna, Austria
http://www.icufn.org/main/

IEEE ICME 2016 — IEEE Int'l. Conference on Multimedia and Expo, 11–15 July
Seattle, WA
http://www.icme2016.org/

–Communications Society portfolio events appear in bold colored print.
–Communications Society technically co-sponsored conferences appear in black italic print.
–Individuals with information about upcoming conferences, Calls for Papers, meeting announcements, and meeting reports should send this information to: IEEE Communications Society, 3 Park Avenue, 17th Floor, New York, NY 10016; e-mail: p.oneill@comsoc.org; fax: + (212) 705-8996. Items submitted for publication will be included on a space-available basis.





GCN GLOBAL COMMUNICATIONS NEWSLETTER

February 2016
ISSN 2374-1082

REGIONAL REPORT

“Start-up Scrum” Spain: EU-XCEL European Virtual Accelerator

By Pilar Manzanares-López, Josemaría Malgosa-Sanahuja, and Juan Pedro Muñoz Gea, Spain

The EU-XCEL European Virtual Accelarator is an EU-Horizon 2020 funded project and part of the Startup Europe Innitiative. The project is coordinated by University College Cork (UCC), Ireland, with project partners including the Strascheg Center for Entrepreneurship from Munich University of Applied Sciences (MUAS), InQbator from Poznan Science & Technology Park (PSTP), DTU Skylab from Technical University of Denmark (DTU), Athens University of Economics and Business (AUEB), and the Cloud Incubator Hub from Universidad Politecnica de Cartagena (UPCT).

On April 15th the EU-XCEL European Virtual Accelerator launched a European-wide recruitment campaign to seek out talented aspiring entrepreneurs in the field of information and communication technology (ICT). The targets were young tech entrepreneurs, between the ages of 18 and 25, who graduated between 2010 and 2014, primarily from technology backgrounds with an entrepreneurial idea ready for development. The main areas of interest are Internet of Things, Health Informatics, Big Data, ICT for Development, Predictive Analysis, and E-/M-Commerce.

Out of a total of 600 applicants, 300 successful applicants from across Europe have the opportunity of participating in the project alongside some of their most promising and talented competitors in intensive, specially designed entrepreneurship training and mentoring programs over four months.

The EU-XCEL Virtual Accelerator includes a one week free intensive training program named “Start-up Scrum” in one of the six participating countries between May and July, 2015. When the applicants applied for a place in the project, they did not choose the Start-up Scrum they wanted to join. They only indicated their preferences according to the dates, but without knowing the country. Applicants who were chosen in the selection phase were assigned to a Start-up Scrum by the organization. One of the guidelines that was taken into account to allocate participants to scrums has been the internationalization. As a result, each scrum was attended by entrepreneurs from about nine different nationalities.

In Spain, the EU-XCEL summer scrum took place the first week of July at the Universidad Politécnica de Cartagena. During this week different activities, masterclasses, and events were scheduled as part of the intensive training of the entrepreneurs.

The first day was focused on ice-breaker and open-minded activities. Participants attended an interesting talk about the Internet of Things, and participated in a creative activity (creating innovative desserts) to apply cooperative design techniques. Another talk, “How to pitch yourself in a couple of minutes,” was scheduled to help entrepreneurs present themselves and their ideas.

On the second day, after exposing the criteria for team formation, entrepreneurs were ready to establish their teams and start work on their ideas. Market research and prototyping are essential to building a successful start-up. Throughout the day different activities were scheduled for training in these areas.

The third day was focused on the presentation of the business model canvas, a strategic management and entrepreneurial tool. Participants used it develop business models during the Scrum.

Although each group was created during the Scrum, the groups did not vanish at the end of the week. Each group will continue working remotely during a four-month period. EU-XCEL offers access to online technical and business development support to help teams further develop and refine their start-up idea. The online platform includes a community forum, training content and resources, virtual technical support and mentoring, and virtual business development support and mentoring.

The fifth start-up scrum for 2015 took place in Cartagena, Spain July 6–10. Forty young entrepreneurs from all over Europe were hosted at the Universidad Politécnica de Cartagena (UPCT).

For that reason, on the fourth day special attention was paid to training on international teams and virtual workspaces. However, the use of an appropriate online platform is not enough. When people from different countries work together, different aspects must be considered, some of them related to communication among participants (different communication styles between cultures, different time zones, etc.), but also other related to legal aspects when an enterprise is going to be created, for example, in which country is it better to register the enterprise. Teams were advised and trained about how to

(Continued on Newsletter page 4)

CONFERENCE REPORT

IM 2015 in Ottawa Focused on Integrated Management in the Age of Big Data

By Wahab Almuhtadi, Chair of the IEEE Ottawa ComSoc/CESoc/BTS Joint Chapter, Canada

The IEEE Communication Society, the IEEE Ottawa Section, and the IEEE Ottawa ComSoc/CESoc/BTS Joint Chapter joined forces to sponsor and organize the IFIP/IEEE International Symposium on Integrated Network Management 2015 (IM 2015) on 11-15 May at the Shaw Centre, located on the scenic banks of the Rideau Canal in the cultural center of Ottawa, Canada's beautiful capital city. As a core national high-technology and R&D center, Ottawa is an ideal setting for this conference. There are more than 1,800 companies in the area working in software, photonics, defense & security, and more. It also offers many historical and cultural sites, including Canadian Parliament buildings, the Canadian Museum of History, and the National Gallery of Canada, among others.

IM 2015 attracted and brought together many of the world's industry leaders, scientists, academics, engineering professionals, and students from around the world. The theme of the IM 2015 conference, "Integrated Management in the Age of Big Data," aimed at capturing the new management challenges and opportunities offered by Big Data, and examined the potential of Big Data to improve the quality of management in a number of areas, including network management, service management, and business management, as well as in management approaches and methodologies. The emergence of network virtualized function also creates a great opportunity for traditional network management, and the instrumentation when the hardware is generic. In this perspective, the conference program featured:

•Six keynote presentations presented by top industry and academic leaders who shared their visions and experiences, and challenged us in exciting new ways. The engaging keynote talks were: "Practical meets Transformational Aspirations" by Ibrahim Gedeon (TELUS, Canada); "Big Data in Science: The Good, The Bad, and The Ugly" by Joseph L. Hellerstein (University of Washington, USA); "Data Science: The Analytics of Big Data" by José M. F. Moura (Carnegie Mellon University, USA); "Harnessing a Petabyte: Opportunities and Challenges for Next Generation Ultra-Scale Data Platforms" by Richard J. Friedrich (Hewlett-Packard Laboratories, USA); "From Software-Defined Infrastructures to Smart City Platforms" by Alberto Leon-Garcia (University of Toronto, Canada); and "Micro Cloud: Moving Computing to Data to Deal with Data Management and Security Issues for Enterprise Clouds" by Dinesh C. Verman (IBM T.J. Watson, USA).

•Four panel sessions, featuring recognized experts sharing exciting and often controversial viewpoints on hot new topics of importance to our community. The topics covered included: big data analytics, cloud infrastructures, data centers, software-defined networking, virtualization, Internet-of-Things, smart cities, future Internet, content delivery, self-management, and security.

•Fourteen technical paper sessions and eight mini-conference sessions. The selected papers presented the latest research advances in the field. The conference had 206 technical paper submissions from all over the world, including Africa, Asia, Europe, North America, and South America. All submitted papers underwent a rigorous review process with at least three reviews for every paper and a rebuttal phase. The review process concluded with 56 papers accepted for the main technical track, with a competitive acceptance rate of 27 percent. Due to the high quality of the submitted papers, many good papers could not be selected for this track. The 36 best papers were selected for the eight mini-conference tracks.

•The program of IM 2015 also offered four experience paper sessions, 20 selected papers presenting lessons learned from developing and deploying management solutions.

•Four poster sessions (52 posters) that enabled one-on-one interactions between authors and attendees.

•Three demonstration sessions with 18 demos showcased research prototypes and research demonstration systems.

•Two dissertation sessions (eight papers), recognizing the best doctoral theses from among the best and brightest of our next generation.

•Workshops, with a specialized focus on the latest breakthroughs in information and communications technology management in an environment that encourages discussion and debate. There were 67 workshop paper submissions.

•Tutorials, offering educational material to keep up with new and emerging topics essential to today's engineering and technology environment.

•Industry exhibits, where major vendors and service providers displayed their latest products and services, including conference patrons.

•A young professional session, providing a forum for the IEEE Graduates of the Last Decade (GOLD) program to address career and other professional development needs for recent graduates.

In addition to the technical program, several Committee and Board meetings were held during the IM 2015, including the TNSM Editorial Board meeting; the CNOM/IFIP WG6.6 committee meeting; the IJNM Editorial Board meeting; the NOMS/IM Steering Committee meeting; the NOMS 2016 TPC/OC committee meeting; and the IRTF NMRG committee meeting.

The IM'15 Organizing Committee of dedicated and outstanding volunteers achieved their goal to make IM 2015 a successful event, and made the attendees' participation in IM 2015 one of their most valuable and memorable experiences from both professional and personal perspectives.

Our valued patrons also made the conference a success due to their contributions: TELUS (Diamond Patron), Algonquin College, Juniper Networks, and Huawei (Gold Patrons); the University of Ottawa, Carleton University, Ciena, Nakina Systems, IEEE SDN, and IEEE Big Data (Silver Patrons); openNMS (Bronze Patron); and Ottawa Tourism and CENGN (Supporters).

For more details on IM 2015, visit: http://im2015.ieee-im.org/.

IM 2015
IFIP/IEEE INTERNATIONAL SYMPOSIUM ON INTEGRATED NETWORK MANAGEMENT
OTTAWA, CANADA • MAY 11-15 2015
"Integrated Management in the Age of Big Data"

CONFERENCE REPORT

Highlights of the International Conference on Advanced Technologies for Communication (ATC 2015)

Ho Chi Minh City, Vietnam, October 2015

By Ha Hoang Kha, HCMUT, Viet Nam

The 2015 International Conference on Advanced Technologies for Communications (ATC2015) was held in Ho Chi Minh City October 14-16, 2015. The conference was jointly organized by the IEEE Communications Society, the Radio-Electronics Association of Vietnam (REV), and Ho Chi Minh City University of Technology. The main theme of the conference was Technology Integrating Antennas and Radio Frequency Integrated Circuits. The Technical Program Committee (TPC) of the Conference accepted 129 papers, including 100 oral presentations and 29 poster presentation. All papers will be indexed by IEEEXplore.

ATC2015 attracted more than 250 scientists and researchers from 30 different countries around the world to discuss the latest advanced technologies in electronics and communications. The technical sessions of the conference were focused on various topics, including antennas and propagation, microwave engineering, communications, signal processing, biomedical engineering, networks, IC and electronics design.

One of the important successful events at ATC2015 was the presence of the world leading keynote speakers. Three keynote speakers presented innovative technologies in wireless communications.

First, at the opening day ceremony, Prof. Constantine A. Balanis, Regent's Professor at Arizona State University, USA, delivered the speech, entitled "Smart Antenna: Technology Integrating Antennas, DSP, Communications and Networks." Prof. Balanis's speech garnered special attention from more than 250 delegates.

Conferring the best paper awards.

ATC 2015
The 2015 International Conference on
ADVANCED TECHNOLOGIES FOR COMMUNICATIONS
October 14-16, 2015, Hochiminh City, Vietnam
mobifone
NICT
UNIVERSITY OF TECHNOL

Prof. Constantine A. Balanis and Prof. Cam Nguyen talking during a keynote speech.

A technical session.

During the beginning sessions of the second day, Prof. Cam Nguyen, Texas Instruments Endowed Professor at Texas A&M University, USA, presented the hot topics related to "Radio Frequency Integrated Circuits: the Backbone of Modern Wireless Communications, Radar and Sensing."

It was followed by the keynote speech, "Wireless Powered Communication Systems: Overview, Recent Results and Challenges," by Prof. Robert Schober, Chair of Digital Communications, Alexander von Humboldt Professor at Friedrich-Alexander University of Erlangen-Nuremberg, Germany.

In addition, sessions on various topics were presented during the three days of the conference. A total of 129 papers were accepted to be presented at the conference out of 220 submitted papers, yielding an acceptance rate of 58 percent. More importantly, the Technical Committee chose two excellent papers to receive the best paper awards.

In addition to regular sessions, ATC2015 also featured a special session on "Computational Science and Computational Intelligence," and two tutorial sessions: "Small Cells for 5G: Fundamentals and Recent Theory" by Prof. Tony Q. S. Quek, Singapore University of Technology and Design, Singapore; and "Radar System Engineering" by Dr. Lee Kar Heng, Chief, TBSS Group, Singapore. During the conference, the conference organizer cooperated with the National Institute of Information and Communications Technology (NICT), Japan, to hold aworkshop on Revolution and Evolution of Photonic Technologies.

ATC2015 presented great opportunities for long-term and sustainability collaborations between scientists in Vietnam and other countries. The next conference is scheduled for 2016 in Ha Noi, Viet Nam.

Opening ceremony.

CHAPTER REPORT

2015 IEEE Harbin Communications Society Chapter Annual Meeting

By Weixiao Meng, Chair of Harbin ComSoc Chapter, China

The Harbin ComSoc Chapter held their annual meeting at the Communication Research Center at the Harbin Institute of Technology on Sept. 28, 2015. The meeting was chaired by Prof. Weixiao Meng, the Chair of the Harbin ComSoc Chapter, with the main purpose to promote the IEEE Communications Society to the faculty members and students.

Prof. Meng gave an opening speech with a brief introduction of the Harbin ComSoc Chapter and an overview of the organization of IEEE. Prof. Meng also highlighted the technical lectures of visiting scholars and several interesting self-designed logos and souvenirs. He also introduced the membership of IEEE and the flagship conferences sponsored by the IEEE Communications Society to raise awareness among the young students, who are keen on doing research and development.

Following the opening speech, Dr. Shuai Han, the Secretary of the Harbin ComSoc Chapter, gave a presentation on "Conference Paper versus Journal Paper." This topic described the newly published statement of IEEE, which helped the students to have a better understanding of how to use the reference materials that were already published via other conference(s) as a reference for their journal paper submission. This topic has captured the interests of the students when preparing their research papers.

After the presentation, the participants celebrated the Mid-Autumn Festival, which is a traditional festival celebrated by the Chinese. Romantically speaking, the festival is to commemorate Chang E, who in order to protect her beloved husband's elixir, ate it herself and flew to the moon. There are many other legends and stories related to this grand festival.

The Chapter organized several interesting activities during the celebration, such as guessing lantern riddles. Participants enjoyed sharing and eating the mooncakes, and the meeting was ended with the sound of joy.

09/28/2015
Monday IEEE ComSoc Harbin Chapter Annual Meeting & Membership Development

Participants at the 2015 IEEE Harbin Communications Society Chapter annual meeting.

START-UP SCRUM/*Continued from page 1*

identify and solve all these details. Finally, each group had to develop an organization plan and action plan that covers the following 12 weeks.

The last day was "Pitching Day." Forty-five entrepreneurs presented their start-ups, covering not only the "idea" but also all the sides trained during the Scrum: the team, the market, and of course, the competition.

The selection of the best participating teams will have the opportunity to compete in the EU-XCEL Ultimate Challenge Final on 2–3 November 2015 in Cork, Ireland, where they will pitch to and connect with some of Europe's leading venture capitalists, angel investors, and successful tech entrepreneurs.

OMBUDSMAN

ComSoc Bylaws Article 3.8.10
The Ombudsman shall be the first point of contact for
reporting a dispute or complaint related to Society activities
and/or volunteers. The Ombudsman will investigate, provide direction
to the appropriate IEEE resources if necessary,
and/or otherwise help settle these disputes at an appropriate level
within the Society....."
IEEE Communications Society Ombudsman
c/o Executive Director
3 Park Avenue
17 Floor
New York, NY10017, USA
ombudsman@comsoc.org
www@comsoc.org "About Us" (bottom of page)

GCN GLOBAL COMMUNICATIONS NEWSLETTER

STEFANO BREGNI
Editor
Politecnico di Milano — Dept. of Electronics and Information
Piazza Leonardo da Vinci 32, 20133 MILANO MI, Italy
Tel: +39-02-2399.3503 — Fax: +39-02-2399.3413
Email: bregni@elet.polimi.it, s.bregni@ieee.org

IEEE COMMUNICATIONS SOCIETY

STEFANO BREGNI, VICE-PRESIDENT MEMBER RELATIONS
PEDRO AGUILERA, DIRECTOR OF LA REGION
MERRILY HARTMANN, DIRECTOR OF NA REGION
HANNA BOGUCKA, DIRECTOR OF EAME REGION
WANJIUN LIAO, DIRECTOR OF AP REGION
CURTIS SILLER, DIRECTOR OF SISTER AND RELATED SOCIETIES

REGIONAL CORRESPONDENTS WHO CONTRIBUTED TO THIS ISSUE

THOMAS MICHAEL BOHNERT (THOMAS.BOHNERT@ZHAW.CH)
EWELL TAN, SINGAPORE (EWELL.TAN@IEEE.ORG)

A publication of the
IEEE Communications Society

www.comsoc.org/gcn
ISSN 2374-1082

NEWLY APPROVED AMENDMENTS TO THE IEEE COMSOC CONSTITUTION

In December 2015, the ComSoc Board of Governors approved a revision of ComSoc Constitution. The IEEE approved the revision in December 2015, and ComSoc Membership approval is currently pending.

The main changes in this revision can be summarized as follows:

- Alignment with IEEE governing documents.
- Alignment with ComSoc Bylaws, which were also revised in December 2015 (see this issue page 17).
- Various clarifications and language improvements.

The revised Constitution and the previous one can be found here:

http://www.comsoc.org/about/documents/constitution

Objections to proposed changes must be emailed to Susan Brooks, ComSoc Executive Director, at s.m.brooks@comsoc.org by 15 April 2016.

IEEE COMMUNICATIONS SOCIETY CONSTITUTION
(IEEE APPROVAL: DECEMBER 2015)
(COMSOC MEMBERSHIP APPROVAL: PENDING)

Table of Contents – Articles

Article 1 - Name, Purposes, and Scope

1.1 Name – The name of this organization is the IEEE Communications Society, hereinafter referred to as "the Society." It is organized within the Institute of Electrical and Electronics Engineers, Inc. hereinafter called "the IEEE."

1.2 Purposes – The purposes of the Society are:

- Scientific and educational – directed toward the advancement of the theory, practice and application of communications engineering and related arts and sciences;
- Professional – directed toward promotion of high professional standards, development of competency and advancement of the standing of members of the profession it serves.

1.3 The Society promotes cooperation and exchange of information among its members and those of other organized bodies within and outside the IEEE. Means to these ends may include, but are not limited to, the holding of meetings for the presentation and discussion of papers, the publication of journals, sponsorship of tutorial seminars and workshops, stimulation of research, the education of members, establishment of standards, and providing for the technical and professional needs of its members via organized efforts.

1.4 Scope – The IEEE Communications Society embraces the science, technology, applications and standards for information organization, collection and transfer using electronic, optical and wireless channels and networks, including but not limited to:

- Systems and network architecture, control and management;
- Protocols, software and middleware;
- Quality of service, reliability and security;
- Modulation, detection, coding, and signaling;
- Switching and routing;
- Mobile and portable communications;
- Terminals and other end devices;
- Networks for content distribution and distributed computing; and
- Communications-based distributed resources control.

1.5 Authority – Society organization and operations are in accordance with the IEEE Constitution and Bylaws, IEEE Policies, IEEE Operations Manuals of Major Boards reporting to the IEEE Board of Directors, and Society governing documents that are not in conflict with any of the aforementioned documents.

Article 2 - Bylaws

2.1 Bylaws are rules and regulations adopted by the Society for governing its members and for the overall management of its affairs. They provide guidance to govern all phases of the organization, management and activities, as outlined in the Constitution. Bylaws may not be in conflict with the Constitution. Bylaws are approved and amended by the Board of Governors and may be changed as Society interests evolve.

Article 3 - Policies and Procedures

3.1 Policies and Procedures provide more detailed statements about specific policies, objectives, and procedures than are contained in the Constitution or Bylaws. Policies and Procedures shall be amended as specified in the Bylaws.

Article 4 - Membership

4.1 IEEE members of any grade shall become Society members upon application and payment of the Society membership dues. The membership dues, the cost of publications, and other considerations members receive for the membership dues are to be set as part of the Society's annual budget.

4.2 Grades of membership for the Society shall be as specified in the Bylaws.

4.3 Individuals who are not members of the IEEE may become Affiliate members of the Society upon:

- Meeting the requirements established in the IEEE Bylaws for Affiliate membership;
- Making proper application for Affiliate membership;
- Making appropriate payment for Affiliate membership.

Any other requirements for Affiliate membership shall be as established in the IEEE Bylaws.

Article 5 - Organization

5.1 Board of Governors (BoG) – The Society shall be governed by the BoG. The BoG shall have primary fiduciary responsibility for the Society and shall set Society policy.

5.2 Society Officers.

5.2.1 Elected Officers – The Society Officers elected by the Society membership shall be:

- President-Elect – The President-Elect shall serve a one-year term the year following his/her election. Following the term of one year as President-Elect, the holder of that office shall serve as President in the subsequent two years and shall serve as Past President in the year after that. The President is the Chief Executive Officer of the Society and chairs the BoG.
- Vice Presidents – Each shall chair a Council responsible for a key area of interest to the Society.
- Members-at-Large – They shall be elected for staggered multiyear terms. The operations of the Society shall be assessed periodically, and the number of Members-at-Large shall be adjusted in accordance with membership needs and growth.
- IEEE Division III Delegate(s)/Director(s) elected by Society membership. By virtue of such election, the holder of that office shall also be a Society Officer.

All elected Officers, including the President and the Past President, shall be voting members of the BoG.

5.2.2 Appointed Officers – The President shall appoint Officers to assist in managing Society activities. As specified in the Bylaws, such appointments may be made upon recommendation of the appropriate Vice President may require BoG approval, and the initial appointments for a President's term shall be proposed for BoG approval in an odd-numbered year by the President-Elect. These Officers serve concurrently with the nominal term of the President. The Society appointed Officers, subject to BoG approval, shall be:

- Treasurer, who shall be responsible for the financial affairs of the Society.
- Directors, each of whom shall chair a Board.
- Other appointed Officers as may be provided for in the Bylaws.

Appointed Officers may be non-voting members of the BoG, as specified in the Bylaws. Appointed Officers serving on the BoG may participate in discussions and make or second motions. If appointed Officers are serving on the BoG concurrently as elected Officers, then they shall be voting members of the BoG.

5.3 Operating Committee – Management of Society affairs between regular and special meetings of the BoG is delegated to an Operating Committee (OpCom). Actions of OpCom shall be ratified at the next BoG meeting, except for those actions taken in areas already delegated.

5.4 Councils – Councils report to the BoG. They are responsible for the policies of their Boards/Standing Committees, and they oversee the operations of their Boards/Standing Committees, address issues common to all their Boards/Standing Committees, address issues that could not be resolved at the Board/Standing Committee level, and escalate such issues to the BoG if they cannot be resolved at the Council level. Councils are established and dissolved through resolutions of the BoG, as specified in the Bylaws.

5.5 Boards – Boards are the major operational entities of the Society and report to Councils, as specified in the Bylaws. Boards have scopes aligned with the scope of their Councils and, within their scopes, they may decide policies and make operational decisions as allowed by their Councils. Boards are established and dissolved through resolutions approved by the BoG, as specified in the Bylaws.

5.6 Standing Committees – Standing Committees report to the BoG or Councils, as specified in the Bylaws. They are established and dissolved through resolutions approved by the BoG, as specified in the Bylaws.

5.7 Ad Hoc Committees – Ad Hoc Committees may be established or dissolved by the President, and report to the BoG.

5.8 Professional Staff – The staff consists of paid professional employees of IEEE who support the activities of the Society. The staff is managed by the Society Executive Director who is a Society Officer, an ex-officio non-voting member of the BoG, and also serves as the BoG Secretary.

Article 6 - Finances

6.1 Assets – All funds and property held by or for the Society are vested in the IEEE.

6.2 Revenues – Basic revenues consist of dues or assessments that are levied on members of the Society, covering publications supplied and services rendered to all members. Other revenues may be raised from other sources, consistent with IEEE regulations. Proposed new income sources require the approval of IEEE.

6.3 Budget – An annual budget shall be prepared and approved by the BoG and IEEE in advance of each fiscal year. Any changes to the budget, or expenditures in excess of budgeted amounts or for unbudgeted items, require advance approval by the BoG before commitment and/or payment.

6.4 Debts – Neither the Society nor any Officer or representative thereof has any authorization to contract debts for, pledge the credit of, or in any way bind the IEEE without prior approval or authorization by persons, organizations or documents as specified by IEEE.

Article 7 - Member Services

7.1 Meetings and Conferences – Principal Society meetings are conferences, workshops, symposia and conventions, held either alone or in cooperation with other IEEE units and/or other professional or technical organizations.

7.2 Publications – The Society, subject to the editorial and fiscal policies of the IEEE, publishes magazines, transactions, journals and other technical materials, such as leading-edge technical articles, tutorials, conference papers, etc. Subscriptions charged for such publications may be higher for non-member subscribers and purchasers than for Society members.

7.3 Education – Principal educational activities include basic and continuing education and training programs.

7.4 Standards – The Society sponsors standards development in accordance with the process defined and approved by IEEE Standards Associations. It also organizes standards-related activities that comply with applicable IEEE/ComSoc and/or IEEE-SA policies.

Article 8 - Amendments

8.1 Constitution

8.1.1 Amendments to this Constitution may be initiated with a:

- Proposal approved by the BoG.
- Petition submitted to the President by a minimum of 100 Members.

8.1.2 Procedure on Proposals – Proposed amendments to the Constitution require two-thirds majority vote of all the voting members of the BoG. Amendments are subject to the approval of the IEEE Vice President, Technical Activities, in accordance with the guidelines set forth in the TAB Operations Manual. After such approval, the proposed amendment shall be published in the Society magazine, or sent to the membership via email or eNewsletter. The amendment becomes effective unless one percent or more of the membership objects in writing to the designated IEEE office within 60 days of publication.

8.1.3 Procedure on Petitions – When a petition for a proposed amendment is submitted, the BoG shall prepare a summary statement and a recommendation for or against adopting the amendment. Summary statements and recommendations require a two-thirds majority vote of all the voting members of the BoG. The petition, summary statement, and recommendation shall be subject to approval by the IEEE Vice President, Technical Activities, in accordance with the guidelines set forth in the TAB Operations Manual. After such approval, the proposed amendment shall be published in the Society magazine, or sent to the membership via email or eNewsletter. The amendment becomes effective unless one percent or more of the membership objects in writing to the designated IEEE office within 60 days of publication.

8.1.4 Objections – If one percent or more objects, a ballot with the proposed amendment shall be mailed or emailed to all voting members of the Society. A return date of at least 60 days shall be allowed. Proposed amendments require a two-thirds majority of the returned ballots for approval.

8.1.5 An amendment shall become effective 60 days after all necessary approvals and notifications.

8.2 Bylaws

8.2.1 Approval of amendments to the Bylaws at a BoG meeting shall require a two-thirds vote of BoG members in attendance, provided a quorum is present.

8.2.2 Approval of amendments to the Bylaws without a meeting shall require a two-thirds majority vote of all the voting members of the BoG.

8.2.3 Bylaws amendments are subject to the approval of the IEEE Vice President, Technical Activities, in accordance with the guidelines set forth in the TAB Operations Manual. After such approval, the amendment shall be published in the Society magazine or sent to the membership via email or eNewsletter.

NEWLY APPROVED AMENDMENTS TO THE IEEE COMSOC BYLAWS

In December 2015, the ComSoc Board of Governors approved a major revision of ComSoc Bylaws. The IEEE approved the revision in December 2015, and the revised Bylaws are now in force.

The main changes in this revision can be summarized as follows:

- Alignment with IEEE governing documents.
- Alignment with the ComSoc Constitution, which was also revised in December 2015 (see this issue page 14).
- Various clarifications and language improvements.

The revised Bylaws now in force and the previous ones can be found here:

http://www.comsoc.org/about/documents/bylaws

IEEE COMMUNICATIONS SOCIETY BYLAWS (DECEMBER 2015)

Table of Contents – Articles

ARTICLE 1 – OBJECTIVES

1.1 Objectives – The objectives of the Society are to provide to its members and the global community of communications professionals the services outlined in Clauses 1.2, 1.3, and 1.4.

1.2 Technical Information

- Creation by research and innovation by the Communications Society community
- Identification and promotion of hot topics
- Dissemination worldwide by publications, presentations, and electronic media
- Exchange by Chapter activities, workshops, discussions, mutual assessments, general networking on technical subjects, and other means of professional communication
- Facilitation of standards activities

1.3 Education (basic and continuing)

- Tutorials, short courses, lecture programs
- Chapter support and other delivery mechanisms

1.4 Professional Services

- Personal career growth by providing technical and personal development information
- Job opportunity benefits through inter-personal networking and facilitation of interactions among members
- IEEE programs

ARTICLE 2 – MEMBERSHIP

2.1 Availability of membership in the Society is specified in the Constitution.

2.2 Categories of Society membership shall be in accordance with IEEE Bylaws. A member's grade within the Society shall be the same as that member's IEEE grade.

2.3 Members who hold the grade of Graduate Student Member, Member, Senior Member, or Fellow in the IEEE shall have all the rights and privileges of membership within the Society unless otherwise specified in these Bylaws.

2.4 Student Members, Associate Members, and Affiliates shall have all the rights and privileges of membership within the Society with the exception of the right to vote on matters presented to the Society membership and the right to hold office.

2.5 A Society member who is delinquent in paying Society dues shall be dropped from membership according to IEEE procedures. A former member may reinstate membership upon payment of current dues.

ARTICLE 3 – OFFICERS AND OPERATIONS

3.1 All officers (except the Society Executive Director) who are members of the Board of Governors (BoG), Councils, Boards, and Standing and Ad Hoc Committees or are Technical Committee Chairs and Society Representatives shall be Members of the Society. The President-Elect and Vice Presidents (VPs) shall be Senior Members or Fellows of the IEEE.

3.2 Elected Officers

3.2.1 President-Elect, Vice President-Technical and Educational Activities (VP-TEA), Vice President-Publications (VP-PUB), Vice President-Conferences (VP-CON), Vice President-Member and Global Activities (VP-MGA), Vice President-Industry and Standards Activities (VP-ISA), IEEE Division III Delegate(s)/Director(s)-Elect, and Members-at-Large of the BoG are elected by direct vote of the voting Members of the Society.

- President-Elect shall be elected in even-numbered years, and Vice Presidents in odd-numbered years.
- One-third of the total (12) Members-at-Large shall be elected annually.
- When an elected officer is elected to another Society position, except President-Elect, during his or her term, he/she shall resign from the former position upon taking office.
- When an elected officer is elected to the position of President-Elect during his/her term, he/she may continue holding the earlier position through the conclusion of its term or upon entering the position of President, whichever comes first.

3.2.2 Terms of Office

- President-Elect shall serve a one-year term the year following his/her election (odd-numbered) and begin a two-year term as President the following year (even-numbered), and then continue for a one-year term (even-numbered) as Past President.
- Vice Presidents shall serve two-year terms beginning in the even-numbered year following their election.
- Members-at-Large shall serve a three year term beginning the year following their election.

3.2.3 Eligibility for Re-election

- The President-Elect shall not be re-elected President-Elect for more than one term, consecutive or otherwise.
- Vice Presidents may be re-elected to the same office for a second consecutive two-year term, but are further ineligible for that office until the lapse of one year.
- A member shall be ineligible for a Vice President-level position after being elected for a total of any five vice-presidential terms, consecutive or otherwise.
- Members-at-Large may be re-elected as Members-at-Large for a second consecutive term, but are further ineligible for that office until the lapse of one year.

3.2.4 Absence or Incapacity of:

- President – Duties shall be performed by the President-Elect (odd-numbered years)/Past President (even-numbered years) and then by the Vice President-Technical and Educational Activities, Vice President-Publications, Vice President-Conferences, Vice President-Member and Global Activities, and Vice President-Industry and Standards Activities, in that order.
- President-Elect – The term shall be filled by the Past President, who shall continue in that capacity until a special election is held and a new President-Elect is chosen.
- Vice President – Individuals shall be identified from the appropriate candidate group slate, in the sequence of the number of votes received, and the individual receiving the most number of votes shall be automatically appointed to serve the remainder of the elected term.
- Member-at-Large – Individuals shall be identified from the same regional slate as the candidate being replaced, in the sequence of the number of votes received, and the individual receiving the most number of votes shall be automatically appointed to serve the remainder of the elected term. If none of these individuals can serve, the vacancy shall be filled by action of the BoG upon proposal by the President; a person filling a position in this manner shall serve for the remainder of the elected term.

3.2.5 Removal from Office. A Society Officer elected by the voting members of the Society may be removed from office, with or without cause, as follows:

- By a two-thirds majority vote of the BoG, or
- By a vote of the voting members of the Society within thirty days following the receipt by IEEE of a petition signed by at least 10% of the total number of voting members of the Society moving for the removal of such individual. A ballot on such motion shall be submitted to the voting members of the Society. If a majority of the ballots cast by the voting members for or against such motion are to remove such individual, the individual shall be removed from such positions.

3.2.6 Vacancies. In the event that an elected position is vacated before the full term is completed, the position shall be filled as defined in the governing documents of the Society.

3.3 Appointed Officers

3.3.1 The Society appointed Officers shall be:

- Treasurer – Is responsible for assuring sound financial practices, establishing prudent budgetary policies, overseeing preparation and presentation of the Society's budget and working with IEEE on financial matters.
- Chief Information Officer (CIO) – Oversees cost-effective planning, acquisition, maintenance and use of the Society's information systems and networking, databases and telecommunications services.
- Directors – Each chairs a Board and serves in the Council to which the Board reports.
    - A Regional Director shall be appointed as specified in Clause 3.3.2 from a list containing at least two candidates from each region submitted by the respective Regional Board before December 15 of odd-numbered years. If the respective Board does not submit its list by this deadline, the President-Elect shall propose the appointment in consultation with the VP-MGA Elect, as specified in Clause 3.3.2.
- Parliamentarian – Advises the President on rules of order and proper procedures during BoG meetings. The President may, in case of conflict, request a ruling on procedures from the Parliamentarian.
- Standing and Ad Hoc Committee Chairs

3.3.2 The appointment of Treasurer, CIO, Parliamentarian, Directors, and Standing Committee Chairs shall be proposed for BoG approval by the President-Elect in consultation with the VP-Elect (if any) with whom the position is associated, except for those positions which are held ex-officio as specified in these Bylaws. Such appointment proposal shall be approved by the outgoing BoG at a meeting (regular or special) in an odd-numbered year.

3.3.3 The appointment of Chairs of Ad Hoc Committees rests with the President.

3.3.4 All Society appointed Officers shall serve for the nominal term of the President.

3.3.5 Removal from Office – An appointed Society Officer or Board/Committee member may be removed from office, with or without cause, by (i) a 2/3 majority vote of the appointing assembly or (ii) the individual(s) who hold(s) the office that made the appointment.

3.3.6 Vacancies – In the event that an appointed position is vacated before the full term is completed, the position shall be filled as defined in the governing documents of the Society.

3.4 President, President-Elect, and Past President Responsibilities

3.4.1 In the odd-numbered years, the President-Elect shall assist the President in discharging the responsibilities of that office. In the even-numbered years, the Past President shall assist the President in taking on the responsibilities of that office.

3.4.2 During his/her term, the President-Elect shall start the selection process for Society appointed Officers as specified in Clause 3.3.

3.4.3 During his/her term, the President shall appoint Officers as needed in consultation with the appropriate VP and with BoG approval, as required by the Society governing documents.

3.4.4 The President shall inform the BoG of the roster of all Boards and Standing Committees as soon as they are finalized and by the first BoG meeting of his term.

3.4.5 The President is the highest ranking volunteer officer of the Society. He/she is responsible for leading the implementation of strategic actions and directions set by the ComSoc BoG. The President or his/her delegate represents the Society in negotiations with Sister Societies and other similar organizations.

3.4.6 The President shall oversee and coordinate handling of ethics and conduct issues involving members, including author misconduct, at the Society level. The President shall be assisted by the Society Executive Director and by volunteers with experience in such matters, as needed.

3.4.7 The President shall keep the BoG informed of openings in IEEE leadership positions, solicit from the BoG recommendations for candidates which may also be conveyed directly to IEEE, and encourage ComSoc members to volunteer for such IEEE positions.

3.5 Vice Presidents – Responsibilities

3.5.1 Vice Presidents are accountable for activities in their areas of responsibility. Each chairs a Council and represents that Council in the BoG.

3.5.2 Vice President – Technical and Educational Activities is responsible for all technical activities and educational services within the Society. The following Boards and Standing Committees report to the Council chaired by this Vice President:

- Educational Services Board
- Technical Services Board
- Awards Standing Committee
- Communications History Standing Committee
- Emerging Technologies Standing Committee
- Fellow Evaluation Standing Committee
- Technical Committees Recertification Standing Committee

3.5.3 Vice President – Publications is responsible for all activities of the Society related to print and electronic products, such as journals, magazines and online offerings. The following Boards report to the Council chaired by this Vice President:

- Journals Board
- Magazines Board
- Online Content Board

3.5.4 Vice President – Conferences is responsible for all aspects of technical conferences, workshops, and professional meetings, including conference publications. The following Boards and Standing Committees report to the Council chaired by this Vice President:

- Conference Development Board
- Conference Operations Board
- GLOBECOM/ICC Management & Strategy Standing Committee
- GLOBECOM/ICC Technical Content Standing Committee

3.5.5 Vice President – Member and Global Activities is responsible for: a) all activities, services and programs associated with members and chapters, and oriented to membership retention, development and marketing in the four regions; b) all activities related to the organization and management of chapters; and c) relations with other IEEE and professional societies worldwide.

The following Boards and Standing Committee report to the Council chaired by this Vice President:

- Member Services Board
- Sister & Related Societies Board
- Asia/Pacific Region Board
- Europe, Middle-East & Africa Region
- Latin America Region Board
- North America Region Board
- Women in Communications Engineering Standing Committee

3.5.6 Vice President – Industry and Standards Activities shall be responsible for overseeing all Society activities and programs related to all standards activities and industry services within the Society, including: (i) fostering technical activities related to relevant current standards development and industry services; (ii) identifying opportunities and fostering ComSoc's engagement in new and/or existing standards development projects that are under development by different standards development organizations worldwide; (iii) increasing the visibility of ComSoc industry and standards initiatives within IEEE, the wider international standards community, and the broad international community of communications technologists; (iv) using ComSoc industry and standards activities to forge closer ties with ComSoc's other departments and activities; (v) maintaining a close and informed relationship with the IEEE-SA; (vi) management within ComSoc, according to IEEE governing documents, ComSoc-sponsored IEEE Standards Association (IEEE-SA) projects, and (vii) fostering and implementing activities that are of interest to industry and government, including practitioners, managers, executives, young professionals and other industry professionals. The Vice President – Industry and Standards Activities shall be the official ComSoc liaison to the IEEE Standards Association Board of Governors.

The following Boards and Standing Committee report to the Council chaired by this Vice President:

- Standards Development Board
- Standardization Programs Development Board
- Industry Outreach Board
- Industry Content and Exhibition Standing Committee (ICEC)

The Vice President – Industry and Standards Activities shall be an ex-officio voting member of the Standards Development Board, the Standardization Programs Development Board, the Industry Outreach Board, and the ICEC.

3.6 Board of Governors (BoG)

3.6.1 Officers on the BoG:

- Elected Officers (Voting):
  - President (BoG Chair)
  - President-Elect (odd-numbered years only)
  - Immediate Past President (even-numbered years only)
  - Vice Presidents
  - Members-at-Large
  - IEEE Division III Delegate(s)/Director(s)
  - IEEE Division III Delegate(s)/Director(s)-Elect (odd-numbered years only)

- Ex-Officio Officers (Non-voting):
  - Directors
  - Chief Information Officer
  - Parliamentarian
  - Treasurer
  - Society Executive Director (BoG Secretary)

3.6.2 The BoG shall hold at least two regular in-person meetings annually.

3.6.3 Special BoG meetings may be held at the request of the President or any four members of the BoG. Notice of such special meetings, giving the meeting type (in-person or teleconference), attendance details (time and place or dial-in information), the purpose of the meeting, and the names of the BoG members calling the meeting, shall be sent to the BoG not less than 21 days (for in-person meetings) or 2 days (for teleconference meetings) before the date set for the start of the special meeting.

3.6.4 Each year, the President, Vice Presidents, and Society Executive Director shall submit the coming year's Operating Plans to the BoG. Progress on these plans shall be reviewed throughout the year by the BoG.

3.7 Operating Committee (OpCom)

3.7.1 In between regular BoG meetings, OpCom conducts business on behalf of the BoG and is comprised of a subset of the BoG members. Actions from a duly called OpCom meeting shall be submitted to the BoG for ratification in a consent agenda or further consideration in its next meetings.

3.7.2 OpCom members are:

- President
- President-Elect (odd-numbered years), immediate Past President (even-numbered years)
- Vice Presidents
- Members-at-Large (three – one from each annually elected group and appointed by the President)
- IEEE Division III Delegate(s)/Director(s)
- IEEE Division III Delegate(s)/Director(s)-Elect
- Directors
- Chief Information Officer
- Treasurer
- Parliamentarian
- Society Executive Director

3.7.3 The OpCom members with voting rights are those who are voting members of the BoG.

3.7.4 OpCom shall hold two regular annual meetings, in person or by other means.

3.7.5 Special OpCom meetings may be held at the request of the President or any four OpCom members. Notice of such special meetings, giving the meeting type (in-person or teleconference), attendance details (time and place or dial-in information), the purpose of the meeting, and the names of the OpCom members calling the meeting, shall be sent to OpCom not less than 21 days (for in-person meetings) or 2 days (for teleconference meetings) before the date set for the start of the special meeting.

3.7.6 All OpCom members are expected to attend OpCom meetings, except for the Directors. The President shall determine which of the Directors shall be invited to a particular OpCom meeting.

3.8 Operations in all Society assemblies (BoG/Councils/Boards/Any Committee)

3.8.1 A majority of the voting members of a Society assembly constitutes a quorum.

3.8.2 The vote of a majority of the votes of the members present at a meeting and entitled to vote at the time of voting, shall be the act of the Society assembly provided a quorum is present.

3.8.3 The Chair of a Society assembly shall have no vote except if the vote is by secret ballot or unless the Chair's vote can change the outcome of the vote.

3.8.4 A Society assembly may meet and act upon the vote of its members in person, by any means of telecommunications, or by a combination thereof. The normal voting requirements shall apply when action is taken whereby all persons participating in the meeting can hear each other and speak to each other at the same time.

3.8.5 For meetings with in-person and remote participants, remote participants who either cannot hear other participants or are not heard by other meeting participants do not meet the requirements for meeting attendance and, therefore, are not included in quorum calculations or allowed to vote.

3.8.6 Business may be conducted by means other than formally held meetings when the matter can be adequately handled via letter, electronic ballot, electronic mail interchange, etc., and procedures for transacting business in such a manner shall be specified in the Society Policies and Procedures (P&Ps). When transacting business without a meeting, a majority vote (simple or higher, depending on the motion type) of all assembly members eligible to vote is required for actions so taken. Approved motions shall be confirmed promptly in writing or by electronic transmission and recorded in the minutes of the next meeting.

3.8.7 Minutes of a BoG and OpCom meeting shall be distributed to the BoG within 30 days of the meeting. For executive sessions, only motions passed shall be included in the BoG and OpCom minutes. Executive session minutes shall be kept on file in the office of the Society Executive Director.

3.8.8 Members of BoG and OpCom shall receive notice of their regular meetings no fewer than 21 days prior to the scheduled meeting start date.

3.8.9 If a quorum is not present at a duly called Society assembly meeting, the only business that can be transacted and concluded within the meeting is to take measures to obtain a quorum, fix the time to which to adjourn, to adjourn, or to take a recess. Any other business shall be limited to informal discussion upon which no action shall be taken.

3.8.10 Individuals holding more than one voting position on any Society assembly shall be limited to one vote on each matter being considered by the assembly.

3.8.11 Business transacted by Society assemblies at a meeting shall be conducted according to Robert's Rules of Order (latest revision) unless other rules and procedures are specified in the Not-for-Profit Corporation Law of the State of New York, the IEEE Certificate of Incorporation, and applicable IEEE governing documents.

3.8.12 Business transacted by Society assemblies without a meeting shall follow the procedures specified in the Society P&Ps.

3.8.13 Proxy voting is not allowed.

3.8.14 Society assemblies other than the BoG and OpCom (Councils, Boards, and any Committee) shall hold regular meetings with sufficient frequency to transact Society business with reasonable dispatch.

3.8.15 Society assemblies other than the BoG and OpCom (Councils, Boards, and any Committee) may hold regular or special meetings at the request of the assembly Chair or of a number of voting members equal to the maximum between two voting members and 20% of the assembly voting members. Notice requirements for regular meetings shall be specified in the Society P&Ps. Notice for special meetings, giving the meeting type (in-person or teleconference), attendance details (time and place or dial-in information), the purpose of the meeting, and the names of the assembly members calling the meeting, shall be sent to the assembly not less than 21 days (for in-person meetings) or 2 days (for teleconference meetings) before the date set for the start of the special meeting.

3.8.16 A Management Retreat may be held annually at the discretion of the President.

3.8.17 Assembly members shall adhere to assembly decisions, unless such decisions violate IEEE or Society Constitutions, Bylaws or Policies.

3.8.18 The Ombudsman shall be the first point of contact for reporting a dispute or complaint related to Society activities and/or volunteers. The Ombudsman shall investigate, provide direction to the appropriate IEEE resources if necessary, and/or otherwise help settle these disputes at an appropriate level within the Society. The Nominations & Elections Committee shall nominate two candidates for the position of Ombudsman who are not currently on the BoG and have not been on the BoG for at least two years. The BoG shall then select one of the two candidates to serve for a two-year term beginning the second year of the President's term. The ombudsman shall report to the BoG.

3.8.19 Constitution, Bylaws, and P&Ps of the Society shall be in accordance with the IEEE Governing documents.

3.8.20 Proposed amendments to Society Governance documents should be reviewed by the Governance Committee prior to approval.

3.9 Professional Staff

3.9.1 Subject to compliance with all applicable IEEE Bylaws and Policies, the Society may create an Executive Office supported by IEEE staff. The Society's Executive Office functions to coordinate and carry-out the day-to-day operations, policies and procedures concerning all aspects of the Society's business. The Office also maintains corporate memory and provides ongoing and ad hoc management reports/documents. In addition, the Society's Executive Office serves as one of the Society's primary points of contact for both members and IEEE staff.

3.9.2 Subject to compliance with all applicable IEEE Bylaws and Policies, the Society may determine the budget for the Executive Office. The staff is hired by the IEEE and all conditions of employment shall be based upon IEEE Bylaws, staff policies and practices and all applicable laws and regulations. Office organization, job descriptions, IEEE staff policies and employment practices are available from the IEEE Human Resources Department.

3.9.3 The Society Executive Director is the most senior position on the IEEE staff that supports the Society, and as such, he/she manages and develops, personally and through subordinate management staff, the paid IEEE staff members that support the Society's operations and activities. The Society Executive Director supports the Society President, officers and volunteer leadership to achieve the Society goals. This Society Executive Director reports through the Managing Director, Technical Activities, to the IEEE Executive Director.

3.9.4 The Society Executive Director serves as BoG/OpCom secretary, assisted by staff members as needed.

ARTICLE 4 – COUNCILS

4.1 Councils are chaired by Vice Presidents to address Technical and Educational Activities, Publications, Conferences, Member and Global Activities, and Industry and Standards Activities. Directors of Boards and Chairs of Standing Committees reporting to a Council serve on the Council and all are voting members together with the Council Chair. Vice Presidents should appoint a Council Vice Chair chosen among the Council voting members and may also appoint additional non-voting members as needed. Councils may approve voting rights for these additional members, with the approval of the BoG.

4.2 Council P&Ps are developed by the Council and approved by the BoG.

4.3 Councils may be established or dissolved by a two-thirds majority vote of the BoG. The scope, responsibilities, and P&Ps of a new Council shall be defined before incorporating it into the Bylaws.

4.4 The following Councils shall be formed:

4.4.1 Technical and Educational Activities Council (TEA-C) – This Council is responsible for the educational and technical interests of the Society, encompassing the broad range of communications and communications-related technical areas.

4.4.2 Publications Council (PUB-C) – This Council is responsible for the needs of the Society and Society Members related to print and electronic projects, such as journals, magazines, TC-edited Newsletters and similar publications, and online offerings, not including conference publications.

4.4.3 Conferences Council (CON-C) – This Council is responsible for the needs of the Society and Society Members related to technical conferences, workshops, and professional meetings. Additional voting members of the CON-C shall be the voting members of the Boards reporting to this Council.

4.4.4 Member and Global Activities Council (MGA-C) – This Council is responsible for all Society activities and programs related to members, chapters, membership development, sister and related societies, and Society regions.

4.4.5 Industry and Standards Activities Council (ISA-C) – This Council is responsible for the needs of the Society and Society members related to industry and standards. Additional (non-voting) members of the Council include the Chairs of any Standards Committee reporting to the Standards Development Board.

ARTICLE 5 – TECHNICAL COMMITTEES AND SPECIAL INTEREST GROUPS

5.1 Technical Committees (TCs)

5.1.1 Technical Committees are established to promote and achieve the technical objectives of the Society and report to the Technical Services Board.

5.1.2 Technical Committees may be created, merged, modified, or dissolved by resolution of the BoG, as necessary to ensure the continued relevance and effectiveness of the Society TCs. Proposals to create/merge/modify/dissolve TCs may be made by the Technical Services Board. In the case of proposals for creating new TCs, a petition to the Technical Services Board by 25 Society members can also be made. Proposals shall include the name, scope, tentative program for the first year, and approximate numbers of interested and potential members. Proposals shall be forwarded to the BoG after having been evaluated by the Technical Services Board and the TEA-C.

5.1.3 Technical Committees shall have P&Ps which shall include officer positions and election procedures and they shall conform to the template specified in the Society P&Ps. P&Ps shall be developed by the Technical Committee and approved by the Technical Services Board.

5.1.4 A change of the scope of a Technical Committees shall require approval of the TEA-C and the BoG.

5.1.5 The Chair of a new Technical Committee is appointed for two years by the Director - Technical Services with the approval of the VP-TEA. During this period, a mentor is assigned to the Technical Committee by the Director of Technical Services. Subsequently, the Chair shall be elected by the members of the Technical Committee.

5.1.6 Elections for Technical Committee Chairs are held every two years for a two-year term. A Chair cannot serve more than two consecutive terms of office.

5.2 Special Interest Groups (SIGs)

5.2.1 Special Interest Groups are established to cover substantial and diverse topical areas of current industry interests and report to the Industry Content and Exhibition Committee (ICEC).

5.2.2 Special Interest Groups may be created, merged, or dissolved by the ICEC, as specified in the ICEC's P&Ps.

5.2.3 Special Interest Groups shall develop P&Ps which shall include scope and responsibilities, officer positions, and election procedures. SIG P&Ps shall conform to the template specified in the Society P&Ps and shall be approved by the ICEC.

ARTICLE 6 – BOARDS

6.1 Boards are the operational and strategic entities of their respective Councils and are chaired by Directors. Boards shall report to Councils as specified below:

- Conferences Council (CON-C)
  - Conference Development Board
  - Conference Operations Board
- Member and Global Activities Council (MGA-C)
  - Member Services Board
  - Sister & Related Societies Board
  - AP Region Board
  - EMEA Region Board
  - LA Region Board
  - NA Region Board
- Publications Council (PUB-C)
  - Journals Board
  - Magazines Board
  - Online Content Board

- Industry and Standards Activities Council (ISA-C)
  - Industry Outreach Board
  - Standards Development Board
  - Standardization Programs Development Board
- Technical and Educational Activities Council (TEA-C)
  - Educational Services Board
  - Technical Services Board

6.2 Directors are responsible for appointing members to their Boards, with the approval of the appropriate Vice President.

Unless otherwise specified in the "Board Descriptions," all Board members (ex-officio or not), including the Director, shall be voting members and shall serve two-year terms concurrent with the nominal duration of the presidential term.

In addition to the members specified in the "Board Descriptions" Clause, Directors may appoint additional non-voting members as needed. Boards may approve voting rights for these additional members, with the approval of the VP under which the Board is aligned. Directors should appoint a Board Vice Chair chosen among the Board voting members.

6.3 P&Ps for each Board shall be developed by the Board and approved by the Council to which the Board reports. In the case that the Council does not approve the P&Ps and a compromise cannot be found, the Board may request the BoG to resolve the matter and approve the P&Ps. An exception exists for the Standards Development Board P&Ps and those of the Standards Committees reporting to the Standards Development Board which shall be approved by the IEEE-SA Standards Board.

6.4 Boards may be established or dissolved by a two-thirds majority vote of the BoG. The scope, responsibilities, and P&Ps of a new Board shall be defined before incorporating it into the Bylaws.

6.5 The following Boards shall be formed:

6.5.1 Conferences Development – This Board is responsible for the strategic planning, technical scope, and growth of all ComSoc financially-sponsored conferences (defined as portfolio conferences).

Members include representatives from the TEA-C and at least four Members-at-Large with at least one having served as the Technical Program Committee chair and at least one as the General Chair of a major conference.

6.5.2 Conferences Operations – This Board is responsible for the oversight and management of the operational, publications, and financial aspects of all ComSoc conferences.

Members include the ComSoc Treasurer and at least five Members-at-Large with at least two having served as the General Chair and at least one as the Technical Program Chair of a major ComSoc conference.

6.5.3 Educational Services – This Board is responsible for the oversight of all Society education and training activities, including administration of the Society's programs on continuing education, incorporating tutorials, short courses, lectures, etc. In particular, this Board is responsible for developing and maintaining continuing professional education and training programs, while striking a balance between generating revenue, supporting outreach efforts, and providing services to members.

Members include representatives from the Technical Services, Conference Development, Conference Operations, Industry Outreach and Member Services Boards, ICEC, and at least two Members-at-Large.

6.5.4 Industry Outreach – This Board is responsible for assuring a comprehensive and cost-effective outreach program of Society products and services to industry and governmental communities. It is also responsible for developing liaisons with communications and networking related enterprises to promote ComSoc products and services and to attract industry and government leaders into ComSoc's volunteer community.

Members include a representative from each of the following: the Technical Services Board, the Educational Services Board, the PUB-C, the CON-C, the Standardization Programs Development Board, the MGA-C, and the ICEC. In addition, up to four Members-at-Large may be appointed by the Director to represent external industrial and governmental interests.

6.5.5 Journals – This Board is responsible for the oversight of Society journals. Board members are the Editors-in-Chief of Society journals for which ComSoc is the Managing Partner and/or has a majority financial stake, and two Members-at-Large, in addition to the Director. Additional members may be appointed, including Liaison Editors to other IEEE journals.

6.5.6 Magazines – This Board is responsible for the oversight of Society magazines. Board members consist of the Editors-in-Chief of Society magazines and two Members-at-Large, in addition to the Director. Additional members may be appointed, including Liaison Editors to other IEEE magazines.

6.5.7 Member Services – This Board is responsible for the oversight of all services and programs addressed to members and chapters, and oriented to membership retention and development in the four regions. In particular, this Board is responsible for developing and providing individual-level membership services globally.

Members include the four Regional Directors plus one member per Region selected by the Director from a list of candidates, consisting of at least two names per region, submitted by each Regional Director.

6.5.8 Online Content – This Board is responsible for initiating, assessing and overseeing Society online content. It supports Technical Committee activities; online services; as well as publications, conferences, and education products and services.

Members include representatives from the Conference Development, Conference Operations, Educational Services, Journals, and Magazines Boards; the TEA, MGA, and ISA Councils; the CIO; and up to three additional Members-at-Large.

6.5.9 Regional Boards – These Boards are responsible for stimulating, coordinating and promoting the activities of ComSoc members and chapters throughout the IEEE regions. The four regions, each with its own Board, are:

- Asia/Pacific (AP)
- Europe, Middle-East & Africa (EMEA)
- Latin America (LA)
- North America (NA)

Each Board shall have a minimum of five members, in addition to the Director.

6.5.10 Sister & Related Societies – This Board is responsible for enhancing Society activities with our sister and related societies (SRS) by developing new programs with SRS and cooperating with SRS in offering Society/SRS products and services globally, and strengthening the Society's global and professional reach. This includes establishing and maintaining Society relationships on an international, regional, national or local scale with SRS.

Where appropriate, enhancing Society SRS activities shall be accomplished through collaboration with IEEE sections/chapters, including Society and non-Society chapters.

Membership includes up to three representatives from the MGA-C, up to three members from selected Sister and Related Societies, and up to three Members-at-Large.

6.5.11 Standards Development – This Board is responsible for the promotion and advancement of communications standards.

It consists of eight members in addition to the Director. The Director shall select the members in accordance with the priorities listed in the next paragraph.

The Director shall give priority to serve on the Standards Development Board to: Standards Committee Chairs, ComSoc appointed Chairs or Co-Chairs for joint Standards Committees, Working Group Chairs who are directly sponsored by the Standards Development Board, ComSoc-appointed Working Group Chairs or Co-Chairs for jointly sponsored Working Groups and volunteers in ComSoc Technical Committees.

The Director shall be the official ComSoc liaison to the IEEE Standards Association Standards Board (SASB).

6.5.12 Standardization Programs Development – This Board is responsible for launching pre-and post-standardization technical activities, not restricted to those standards being developed by the IEEE. These would include, but not be limited to Research Groups that lead to the discovery of standardization opportunities and, for completed standards, creation of follow-up programs, such as compliance testing, standards education, workshops, conferences, and publications on technical issues that are relevant to standards.

The Board shall consist of up to eight members in addition to the Director. The Director of the Standardization Programs Development Board serves as the liaison to the IEEE-SA Industry Connections Program.

6.5.13 Technical Services – This Board is responsible for the oversight and promotion of the technical communities of the Society and their activities including the promotion of technical content and development of educational content. In particular, this Board is responsible for developing and providing one-stop ICT (information-communications technology) services.

The Technical Services Board shall also evaluate proposals to create/merge/modify/dissolve TCs. Such proposals shall be forwarded by the Technical Services Board to the TEA-C, with a recommendation. TEA-C shall make a recommendation to the BoG, who shall have final authority for approving the proposal.

Members include Technical Committee Chairs, Chair of the Emerging Technologies Committee, Chair of the Technical Committees Recertification Committee, and a representative of the Educational Services Board. The GITC Chair shall be a non-voting member of this Board.

ARTICLE 7 – STANDING COMMITTEES

7.1 Standing Committees shall report to the BoG or a Council as specified below:

• Finance	BoG
• Governance	BoG
• Nominations& Elections	BoG
• Operations & Facilities	BoG
• Strategic Planning	BoG
• GLOBECOM/ICC Management & Strategy	CON-C
• GLOBECOM/ICC Technical Content	CON-C
• Women in Communications Engineering	MGA-C
• Awards	TEA-C
• Communications History	TEA-C
• Distinguished Lecturers' Selection	TEA-C
• Emerging Technologies	TEA-C
• Fellow Evaluation	TEA-C
• Technical Committees Recertification	TEA-C
• Industry Content & Exhibition	ISA-C

7.2 Standing Committee Chairs shall be responsible for appointing Standing Committee members with the approval of the Society Officer chairing the BoG/Council to which the Standing Committee reports.

Unless otherwise specified in the "Standing Committee descriptions" Clause, all Standing Committee members (ex-officio or not), including the Chair, shall be voting members and shall serve a two-year term concurrent with the nominal duration of the presidential term.

In addition to the voting members specified in the "Standing Committee Descriptions" Clause, Standing Committee Chairs may appoint additional non-voting members as needed. Standing Committees may approve voting rights for these additional members, with the approval of the VP under which the Board is aligned. Standing Committee Chairs should appoint a Vice Chair chosen among the Standing Committee voting members. For Standing Committees reporting to the BoG, the Chair should also appoint a BoG-liaison to represent the Standing Committee on the BoG; the BoG-liaison should be chosen among the Standing Committee voting members who are also BoG members.

7.3 P&Ps of Standing Committees shall be developed by the Standing Committee and approved by the BoG/Council to which the Standing Committee reports. In the case that a Council does not approve the P&Ps of a Standing Committee and a compromise cannot be found, the Standing Committee may request the BoG to resolve the conflict and approve the P&Ps.

7.4 Standing Committees may be established and dissolved by a two-thirds majority vote of the BoG. The scope, responsibilities, and P&Ps shall be defined before incorporating the new Standing Committee into the Bylaws.

7.5 The following Standing Committees shall be formed:

7.5.1 Awards – This Committee is responsible for all major awards and recognitions made or proposed by the Society. It consists of not less than twelve (12) members who shall serve for a three-year term. One-third of the members are appointed each year. Committee members may not provide nominations or reference letters while in office, nor participate in deliberations on awards or recognitions for which they may be under consideration.

7.5.2 Communications History – This Committee is responsible for identifying, placing in electronic archives, and raising public awareness through all appropriate steps on the most important facts/person/achievements of communications history in particular, as well as telecommunication milestones in general. The Committee consists of three members who shall serve a three-year term, with one member appointed each year.

7.5.3 Distinguished Lecturers Selection – This Committee is responsible for establishing selection criteria and for the appointment of lecturers. The ex-officio Chair shall be the Vice Chair of the TEA-C. Members of this Committee shall be the VP-TEA, VP-MGA, the Director-Member Services, and the Chair of the Emerging Technologies Committee.

7.5.4 Emerging Technologies – This Committee is responsible for identifying, describing, and nurturing new technology directions, recommending new programs, and nurturing potential Technical Committees for formal proposal via the VP-TEA.

The Chair shall be chosen among the members of the Strategic Planning Committee with the recommendation of the VP-TEA Elect, as specified in Clause 3.3.2. Standing Committee members shall include at least one more member from the Strategic Planning Committee. The Committee shall have six members appointed for three years with one-third appointed each year. In addition, the Editor-in-Chief of IEEE Communications Magazine and the Editor-in-Chief of IEEE Journal of Selected Areas in Communications are ex-officio voting members of the Committee.

7.5.5 Fellow Evaluation – This Committee is responsible for the Society's evaluation of Fellow nominations being considered by the IEEE Fellow Committee. It consists of a Chair and nine members that shall serve a three-year term with one-third of the members being appointed each year. Chair and members shall be IEEE Fellows and Members of the Society.

7.5.6 Finance – This Committee is responsible for facilitating the Society's budget process and for managing and providing direction in all aspects of Society financial matters. The Committee meets twice a year at ICC/GLOBECOM. The ex-officio Chair shall be the Society Treasurer. Committee members shall be: the President, Past or President-Elect, Vice Presidents, CIO, and a representative from each Member-at-Large class. The Society Executive Director shall be an ex-officio non-voting member of the Committee.

7.5.7 GLOBECOM/ICC Management and Strategy (GIMS) – This Committee is responsible for the successful conduct, strategic evolution, and policies of the IEEE Global Communications Conference (GLOBECOM) and the IEEE International Conference on Communications (ICC). Members of the GIMS Committee shall be the Chair, three or four Members-at-Large, three past members of an ICC or GLOBECOM Organizing Committee, and the GITC Committee Chair.

7.5.8 GLOBECOM/ICC Technical Content (GITC) – This Committee is responsible for providing strategic vision and management of the technical content of GLOBECOM and ICC to guarantee timeliness and the highest level of quality.

The GITC Chair shall be appointed as in Clause 3.3.2, and in consultation with both the VP-CON Elect and the VP-TEA Elect. GITC members shall be appointed by the GITC Chair in consultation with the VP-CON and the VP-TEA. GITC members shall be: the Chair, two to four Members-at-Large, three past ICC or GLOBECOM Technical Program Chairs, and the GIMS Committee Chair. The Director of Technical Services shall be an ex-officio non-voting GITC member. The GITC Vice-Chair shall be appointed from among the voting members by the GITC Chair in consultation with the VP-CON and the VP-TEA.

7.5.9 Governance – This Committee is responsible for all matters related to Society Governance, including but not limited to: reviewing any proposed amendment to Society Governance documents (Constitution, Bylaws, P&Ps) prior to its discussion in the BoG; crafting amendments to Society Governance documents that result from actions of the BoG; establishing Society-wide Governance best practices and overseeing their application across all Councils, Boards, and Committees; upon request or when needed proposing changes to existing Society Governance documents with the goal of keeping them current and consistent; and serving as an interpretive Committee on Governance issues.

Committee membership: Chair, the Society Parliamentarian, and up to three additional members appointed by the President upon recommendation of the Committee Chair and shall include one previous ComSoc President and one sitting BoG Member at Large. The Committee Chair may appoint up to three additional voting Committee members. The Chair and Committee members (except the Parliamentarian) shall serve three-year terms with one reappointment allowed.

Terms of all members (except the Parliamentarian) shall be staggered so that no more than half of the members' terms expire every two years; when necessary, such staggering shall be created by appointing members to terms shorter than three years, as indicated at the time of their appointment.

7.5.10 Industry Content & Exhibition – This Committee is responsible for developing and promoting a strategic vision and oversight for organizing and promoting internal ComSoc communities that are attractive to members from industry, government, or other non-academic sectors. This includes processes to assure the quality and value of content in industry oriented conferences, events and education. It is within the overall objective and mission of the ICEC to increase industry participation in ComSoc events.

Members of this Committee are the Chair and 4-to-6 Members at Large.

7.5.11 Nominations & Elections – This Committee is responsible for identifying candidates to fill elected Society office positions, and for the development, implementation and supervision of election procedures. The Nominating Committee for the IEEE Division III Director shall be a separate Committee and shall operate as specified in the TAB Operations Manual. Meetings of this Committee shall always be held in Executive Session.

The Committee shall consist of the following members:

- Committee Chair (ex-officio) – Shall be the most recent former Past President, who shall serve for a two year term and shall take office as Chair immediately after his/her nominal term of Past President ends (the beginning of an odd-numbered year). If the Past President is unable to serve, the President shall appoint a Chair for a similar period, with the approval of the BoG.
- Nine Members-at-Large – Each shall be a voting member appointed by the BoG, upon recommendation of the President and the N&E Chair, and approved by the BoG for a three-year term with one-third of the members appointed each year. At least one annual appointee shall not be a member of the BoG.
- IEEE Division III Delegate(s)/Director(s) (ex-officio) – Shall be a non-voting member.
- The Past President (ex-officio), only during even-numbered years.
- President-Elect (ex-officio) – Shall be a non-voting member whose term shall start immediately after being elected and officially announced.

The Chair shall not be eligible to be elected to the BoG during his/her term of service. A Committee member may be nominated for a position only if (i) the nomination is not made by a member of the Committee, and (ii) the member resigns from the Committee prior to its first meeting of the year in which the nomination shall be made.

Individual voting members eligible to vote in an election may nominate candidates by written petition, provided such nominations are made at least 28 days before the date of the election. The number of signatures required for a petition candidate to appear on a ComSoc ballot shall be equal to what is set in the IEEE Bylaws as follows: For all positions where the electorate is less than 30,000 voting members, signatures shall be required from 2% of the eligible voters. For all positions where the electorate is more than 30,000 voting members, 600 signatures of eligible voters plus 1% of the difference between the number of eligible voters and 30,000 shall be required.

7.5.12 Operations & Facilities – This Committee is responsible for supporting the President in making recommendations to the BoG on operations, facilities and related capital expenses. The President is the ex-officio Chair. Members shall serve for the nominal duration of the President's term and shall include:

- CIO
- Treasurer
- Four members appointed by the President from among volunteer BoG officers
- One member appointed by the President from among volunteers who are not on the BoG

The Society Executive Director is an ex-officio non-voting member of this Committee. The Committee may approve the participation of invited experts as required by the agenda items.

7.5.13 Strategic Planning – This Committee is responsible for preparing a long-term strategic plan to guide the direction and future of the Society and for preparing short-term plans to direct specific areas, as appropriate.

Members of the Committee shall be the Chair, the Vice Presidents (or a representative named by the VP), and up to four Members-at-Large, all appointed by the President.

7.5.14 Technical Committees Recertification – This Committee recommends the establishment of new Technical Committees and reviews current Technical Committees to determine whether they are fulfilling their responsibilities.

The VP-TEA is the ex-officio Chair. Committee members shall include the TEA-C Vice Chair and six members appointed by the Chair for a three-year term. One-third of the members are appointed each year by the Chair from among Members-at-Large of the BoG. Committee members who are also officers of Technical Committees under review shall excuse themselves from deliberations related to their Technical Committee.

7.5.15 Women in Communications Engineering – This Committee is responsible for encouraging the participation and membership of women communications engineers in the Society. The Committee shall meet at least once a year at ICC or GLOBECOM, and shall provide an annual written report to the Society President, Vice Presidents, and Technical Committee Chairs prior to each ICC.

Committee membership is as follows: Vice-Chair, Publicity Chair, Secretary, IEEE Women in Engineering Committee (WIEC) Society Coordinator, Awards sub-committee Member-at-Large, and up to five Members-at-Large.

ARTICLE 8 – AD HOC COMMITTTEES

8.1 Ad Hoc Committees may be established by the President in consultation with the BoG to address broad technical or operational issues within the Society or IEEE. The scope, responsibilities, and P&Ps of an Ad Hoc Committee shall be defined upon establishment. Ad Hoc Committees report to the BoG.

8.2 Ad Hoc Committee Chairs and members shall be appointed by the President. Upon establishment, the President shall report to the BoG the composition, mission, and expiration date of the Ad Hoc Committee.

8.3 OpCom shall review all Ad Hoc Committees annually and recommend to the BoG whether they should continue, disband, or be elevated to Standing Committees.

8.4 Ad Hoc Committees shall automatically expire at the conclusion of their duration or at the end of each President's term or by resolution of the President or the BoG, whichever comes first. The President may reestablish an expired Ad Hoc Committee in consultation with the BoG.

ARTICLE 9 – SOCIETY REPRESENTATIVES

9.1 Society representatives to other IEEE Organizational Units or non-IEEE organizations are responsible for representing Society interests.. They are appointed by the President for terms as required by the other organizations, in consultation with the appropriate Vice President.

ARTICLE 10 – BUDGET AND FINANCE

10.1 Officers shall prepare budgets for the coming calendar year in the first half of each year to be approved by the BoG at its midyear meeting. Actuals shall be reviewed throughout the year, and a forecast reported at each meeting.

10.2 Dues and fees are set by the BoG in accordance with IEEE and Society guidelines and are based upon proposals by the Treasurer to the BoG. Billing and receipt of annual dues are part of the IEEE dues billing process.

10.3 Budget

10.3.1 Each year the Society produces a budget which shall be approved by the BoG.

10.3.2 The Treasurer is responsible for the development of the Society annual budget and submitting to IEEE Technical Activities for consolidation with other societies, and ultimately to the IEEE for their consolidated budget. The Treasurer monitors revenues and expenses, providing interim reports on budgets, forecast, actuals at each BoG and OpCom meeting. A complete financial report, including actual versus budget, net assets, and reserves is presented by the Treasurer annually.

10.4 Finance

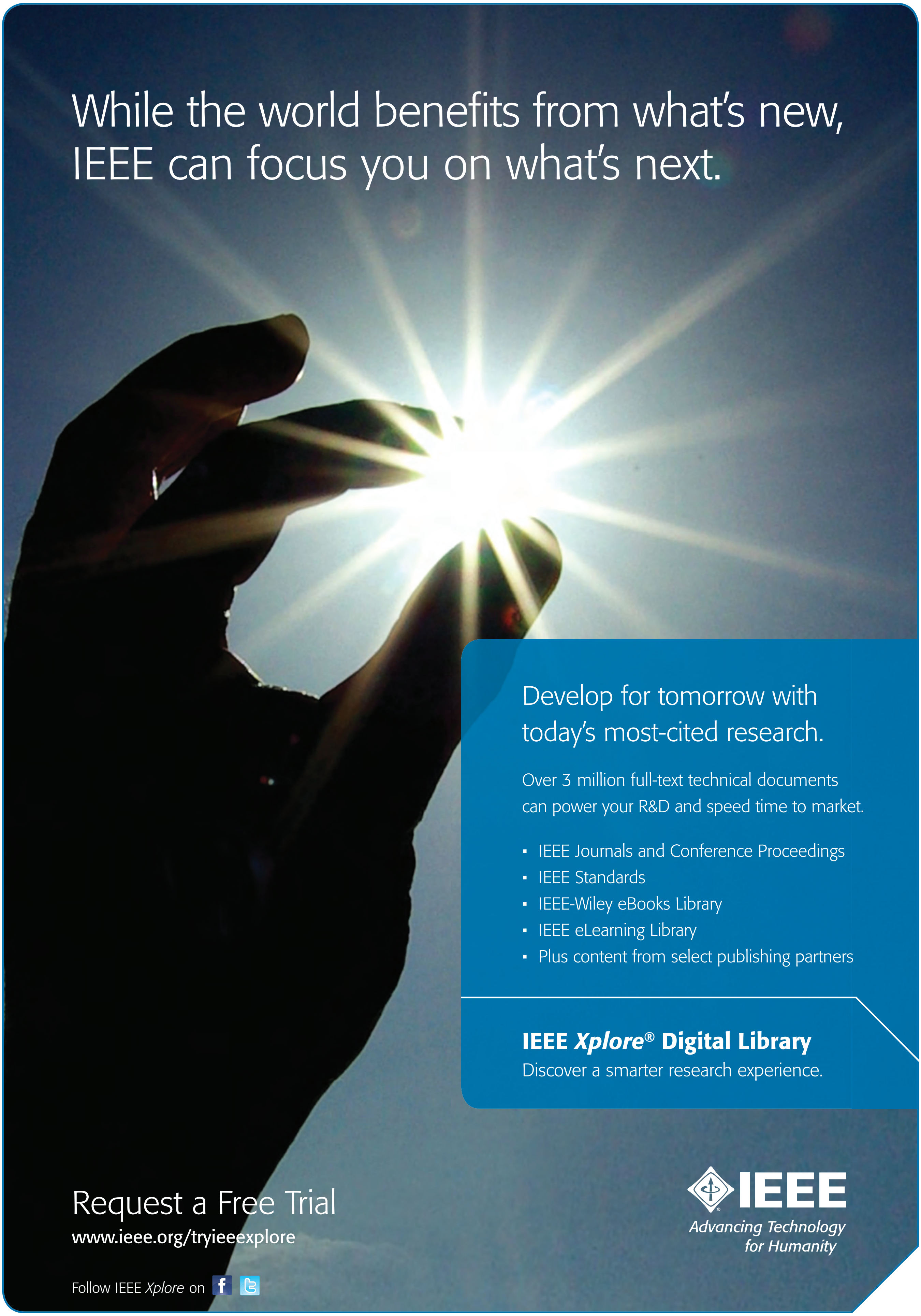
10.4.1 The Treasurer has oversight responsibility for all Society financial matters.

10.4.2 Funds shall be handled as designated by the Treasurer and shall be deposited with IEEE or with external financial institutions, as approved by the BoG and/or IEEE Board of Directors.

10.4.3 The Treasurer, or Society Executive Director, or their designee shall follow orderly procedures for disbursement of funds, providing sufficient checks and balances and appropriate record keeping. A budgeted expenditure requires no further approval beyond approval of the Treasurer.

10.4.4 The Treasurer shall periodically review the Society finances and recommend adjustments needed to insure financial stability of the Society.

10.4.5 The Treasurer shall cooperate with Society and IEEE officials to accomplish financial audits when requested. The results of these audits shall be presented to the BoG.

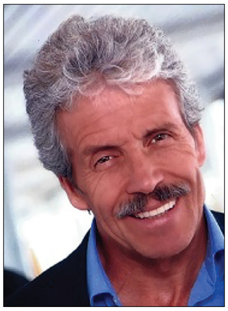


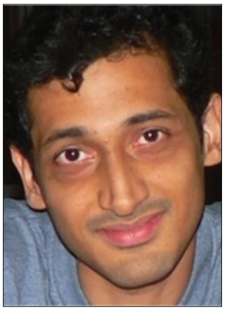
UNDERWATER WIRELESS COMMUNICATIONS AND NETWORKS: THEORY AND APPLICATION: PART 2

Xi Zhang

Jun-Hong Cui

Santanu Das


Mario Gerla


Mandar Chitre

The Earth is a water planet, two-thirds of which is covered by water. With the rapid developments in technology, underwater wireless communications and networks have become a fast growing field, with broad applications in commercial and military water-based systems. The need for underwater wireless communications exists in applications such as remote control in the off-shore oil industry, pollution monitoring in environmental systems, collection of scientific data from ocean-bottom stations, disaster detection and early warning, national security and defense (intrusion detection and underwater surveillance), as well as new resource discovery. Thus, the research into new underwater wireless communication techniques has played the most important role in the exploration of oceans and other aquatic environments. In contrast to terrestrial wireless radio communications, the underwater channel poses serious technical challenges depending on the communications modalities (e.g., acoustic, optical, or RF/magnetic) employed. These include, but are not limited to, ambient channel noise, severe attenuation, propagation delay, multipath, frequency dispersion, bio-fouling, lack of access to precise time synchronization (GPS), and constrained bandwidth and power resources. These challenges also provide an opportunity for design of hybrid and adaptive transmission, such as the underwater acoustic and optical communications and networks, which have somewhat complementary properties, with potential for longer range and higher bandwidth networked communications in size- and power-constrained modems and mobile unmanned systems.

Inspired by the attractive and unique features and potential benefits of advanced underwater communications, the topic of underwater wireless networks has attracted increasing attention from researchers not only in academia, but also in the military and industrial sectors. While a great deal of research efforts have been made in recent years on underwater wireless networks, the aforementioned challenges posed by underwater acoustic as well as optical wireless channel exploitation in future underwater wireless system developments still remain an open problem. As we present Part 2 of this Feature Topic of *IEEE Communications Magazine* focusing on underwater wireless communications and networking, we aim to address the urgent needs in both theory and application aspects of industry, military, and the research community in order to better understand the recent progress, explore the future potential research directions, and define new research paradigms in underwater wireless communications and networks. The response to our Call for Papers on this Feature Topic was overwhelming, with a total of 52 articles submitted from all around the world. Going through the rigorous two-round review process, Part 1 of this Feature Topic, which consisted of eight excellent articles addressing various aspects of underwater wireless networks, was published in the November 2015 issue of *IEEE Communications Magazine*. Part 2 of this Feature Topic presents the following four excellent articles focusing on the key issues and emerging concepts of contemporaneous underwater wireless networks and techniques.

The first article, "RSS-Based Secret Key Generation in Underwater Acoustic Networks: Advantages, Challenges, and Performance Improvements," overviews the advantages, explores the major challenges, and evaluates the performance improvements of received signal strength (RSS)-based key generation techniques in underwater acoustic wireless networks. The second article, "Design Guidelines for Opportunistic Routing in Underwater Networks," investigates the two main building blocks for the design of opportunistic routing protocols for underwater sensor networks — candidate set selection and candidate coordination procedures — and discusses how the resulting approaches are related to the opportunistic routing protocol designs for different scenarios in underwater sensor networks. The third article, "A Journey toward Modeling and Resolving Doppler in Underwater Acoustic Communications," surveys the evolution of Doppler modeling and resolution in underwater acoustic communications through five modeling stages: quasi-static model, uniform Doppler shift model, basis expansion model (BEM), plus path speed model and non-uniform path speed model, and characterizes their respective performance matrixes. The fourth article, "Impulse Response Modeling for General Underwater Wireless Optical MIMO Links," investigates underwater wireless optical communications (UWOC) multiple-input multiple-output (MIMO) systems with M light-sources and N detectors, focusing on the impulse response to characterize the temporal behavior of UWOC links and proposing an M-order weight Gamma function polynomial (WGFP) to model the impulse response of $M \times N$ UWOC MIMO links.

We would like to thank all the authors for their excellent contributions and all the reviewers for their valuable reviewing comments. We also appreciate strong support from Dr. Sean Moore, the former Editor-in-Chief, and Dr. Osman Gebizlioglu, the current Editor-in-Chief of *IEEE Communications Magazine*, and the IEEE Communications Society publishing team. Finally, we hope that the readership will find this Feature Topic interesting and stay tuned for new developments in this compelling research area.

BIOGRAPHIES

XI ZHANG [F'16] (xizhang@ece.tamu.edu) received his Ph.D. degree from The University of Michigan, Ann Arbor. He is a full professor at Texas A&M University. He has published more than 300 research papers, received the U.S. NSF CAREER Award, is an IEEE Distinguished Lecturer, and received four IEEE Best Paper awards. He is the author of an IEEE BEST READINGS journal paper. He has been an Editor for numerous IEEE transactions and journals, TPC Chair for IEEE GLOBECOM 2011, and TPC Vice Chair for IEEE INFOCOM 2010.

JUN-HONG CUI received her Ph.D. degree from the University of California Los Angeles (UCLA) in 2003. She is now a full professor at the University of Connecticut. Her recent research mainly focuses on underwater sensor networks, autonomous underwater vehicle networks, cyber-aquatic systems, smart ocean technology, and ocean computing. She co-founded the ACM International Conference on Underwater Networks and Systems, and is now serving as its Steering Committee Chair. She has received an NSF CAREER Award and an ONR Young Investigator Award.

SANTANU DAS is the program manager of Communications and Networking within the C4ISR Department of the Office of Naval Research, where he has broad responsibility for planning, executing, and providing leadership for integrated science and technology projects to develop new capabilities for naval communication networks. He received a Ph.D. in electrical engineering from the University of Alberta, Edmonton, Canada, and conducted research at AT&T Bell Labs, Whippany, New Jersey, in areas of 3G-wireless and fiber optic communications.

MARIO GERLA [F'02] received his Ph.D. degree from UCLA. He was part of the team that developed the early ARPANET protocols under the guidance of Prof. Leonard Kleinrock. He joined the UCLA Computer Science Department in 1976. He is leading several advanced wireless network projects under industry and government funding. His team is developing a vehicular testbed for safe navigation, content distribution, urban sensing, and intelligent transport.

MANDAR CHITRE received a Ph.D. degree in electrical engineering via research in underwater acoustic communications. He currently holds a joint appointment with the Department of Electrical and Computer Engineering at the National University of Singapore as an assistant professor and with the Tropical Marine Science Institute as head of the Acoustic Research Laboratory. His current research interests include underwater communications, autonomous underwater vehicles, and acoustic signal processing.

RSS-Based Secret Key Generation in Underwater Acoustic Networks: Advantages, Challenges, and Performance Improvements

Yu Luo, Lina Pu, Zheng Peng, and Zhijie Shi

How to secure acoustic communications in a UAN is becoming an important but challenging topic. Among different secret key approaches, received-signal-strength-based (RSS-based) key generation is particularly appealing, as it can eliminate the need to deploy an additional key distribution center, making it a more attractive method than conventional symmetric key cryptography in resource-constrained UANs.

ABSTRACT

Due to the broadcast nature of acoustic channel, underwater acoustic networks (UANs) face threats of eavesdropping and fake data injection. How to secure acoustic communications in a UAN is becoming an important but challenging topic. Among different secret key approaches, received-signal-strength (RSS)-based key generation is particularly appealing, as it can eliminate the need to deploy an additional key distribution center, making it a more attractive method than conventional symmetric key cryptography in resource-constrained UANs. A variety of RSS-based key generation approaches have been designed for wireless radio networks. However, no attempt has been made to evaluate their performance in underwater environments. In this article, we provide an overview of the advantages of RSS-based key generation and explore the major challenges from the unique features of underwater systems through experiment results of sea trails. Meanwhile, we discuss viable solutions to improve the performance of RSS-based key generation in oceans.

INTRODUCTION

As the source of life, oceans never stop attracting people's attention in both academia and industry. Underwater acoustic networks (UANs) enable scalable and distributed data acquisition in a wide spectrum of applications [1], including unmanned ocean exploration, ocean surveillance, and target detection. The possibility of secure message delivery may determine the success or failure of a mission. Therefore, how to protect the communications in a UAN is becoming an important topic.

Like terrestrial sensor networks, UANs are susceptible to various attacks, which target different components in a UAN system. For example, attacks like wormholes target routing protocols, and jamming attacks can disrupt links among nodes. An adversary can also violate communication security by passively eavesdropping on private messages or actively injecting fake information to the network. Among the aforementioned security issues, communication security is one of the most fundamental and critical tasks in UANs, which use a broadcast channel for acoustic transmissions. Public key cryptography is nearly infeasible in networks with constrained energy and processing power. Alternatively, symmetric key ciphers are often used to provide confidentiality in underwater communications because of their performance advantages.

However, symmetric key cryptography requires a shared secret key between a sender and its intended receiver for both encryption and decryption. This requirement makes key generation and key exchange challenging, especially in resource-constrained UANs. It is difficult, if not impossible, to specify an online key distribution center (KDC) in oceans to allocate secret keys among devices. A more acceptable solution is to combine pseudorandom key generators and key predistribution. However, lack of randomness is a common problem in those key generators, leading to cryptanalytic breaks. Meanwhile, key predistribution has connectivity and resilience issues, as an isolated node possibly exists when it has no common key with its neighbors. All the methods that preinstall keys on nodes also have the risk that a single compromised node may make a number of parties unsafe, sharing common keys with the compromised entity.

Received signal strength (RSS)-based key generation allows each pair of nodes, after being deployed, to update secret keys easily at any time. In this scheme, the randomness of a key depends on the entropy naturally available in the environment. Specifically, the communicating parties on the two ends of a reciprocal link can produce a shared key through local RSS measurements [2]. An adversary that is monitoring the communication channel, however, can hardly guess the secret key if it is physically near neither of the communicating entities [3]. Security is consequently ensured with the spatial diversity of a wireless channel, as shown in Fig. 1, where Alice and Bob are two communicating parties, while Eve is an eavesdropper.

Recently, many RSS-based key generation approaches have been proposed for terrestrial radio networks, and extensive theoretical and experimental studies have been conducted [4]. However, the UAN is fundamentally different from any ground-based wireless networks due

The authors are with the University of Connecticut. Lina Pu is the corresponding author.

0163-6804/16/$25.00 © 2016 IEEE

to the unique features of the underwater channel and acoustic communication systems [5]. For instance, the long preamble sequence [6] in an acoustic communication greatly increases the length of a probe in RSS measurements, which results in a slow generation rate of secret keys; the high dynamics of an acoustic channel and the half-duplex feature of an acoustic modem may affect the robustness of a key generation method in terms of bit mismatch rate.

Currently, little is known about the actual performance of existing RSS-based key generation methods in oceans, and no attempt has been made in the literature to evaluate them with sea experiments. This motivates us to put effort into this thread. In this article, we first present an overview of RSS-based key generation methods. Then we explore the advantages of RSS-based key generation and the challenges from unique features of UANs along with experiment results. Finally, we discuss how to improve the performance of RSS-based key generation approaches in oceans.

RSS-Based Key Generation Techniques

An RSS-based key generation method usually consists of four stages, which are signal preprocessing, key extraction, information reconciliation, and privacy amplification. In particular, signal preprocessing helps communicating parities to generate an appropriate sequence based on the raw RSS measurements; key extraction quantizes the output of signal preprocessing into a secret bitstream; information reconciliation removes the erroneous bits from the keys of Alice and Bob; and privacy amplification reduces Eve's knowledge of the secret key to a small value.

Signal Preprocessing

Depending on the application scenario, signal preprocessing could include different processes to reduce the discrepancy or to improve the randomness of an RSS sequence.

Interpolation: Due the half-duplex feature of communication equipment, a pair of nodes may not be able to send their probes at exactly the same time for collision avoidance, which results in disagreements between their RSS sequences. When the time difference between the transmissions of a pair of probes is shorter than the correlation time of a wireless channel, an interpolation filter could be exploited to reduce the discrepancies.

Large-Scale Fading Elimination: When large-scale fading occurs, a wireless channel may have poor quality for a long time, and continuous RSS measurements have a high probability to have small values. This decreases the randomness of an RSS sequence and the security of a key. A simple approach to counter large-scale fading is to divide all of the RSS sequences into several subgroups and then subtract each RSS measurement with the average value in the corresponding subset. This process removes the large-scale fading and retains the random variation of a channel in the small-scale fading for key generation.

Beamforming: By using the beamforming technique, the communicating parties could leverage directional probe transmission and reception to improve channel diversity. The eavesdropper and communicating parties obtain RSS samples that are affected by distinct channel responses, which protects the information from an eavesdropper. However, beamforming techniques usually require each communicating entity to be equipped with at least two transducers, such as antennas, microphones, or hydrophones, which may not be available for acoustic nodes due to the size and cost constraints.

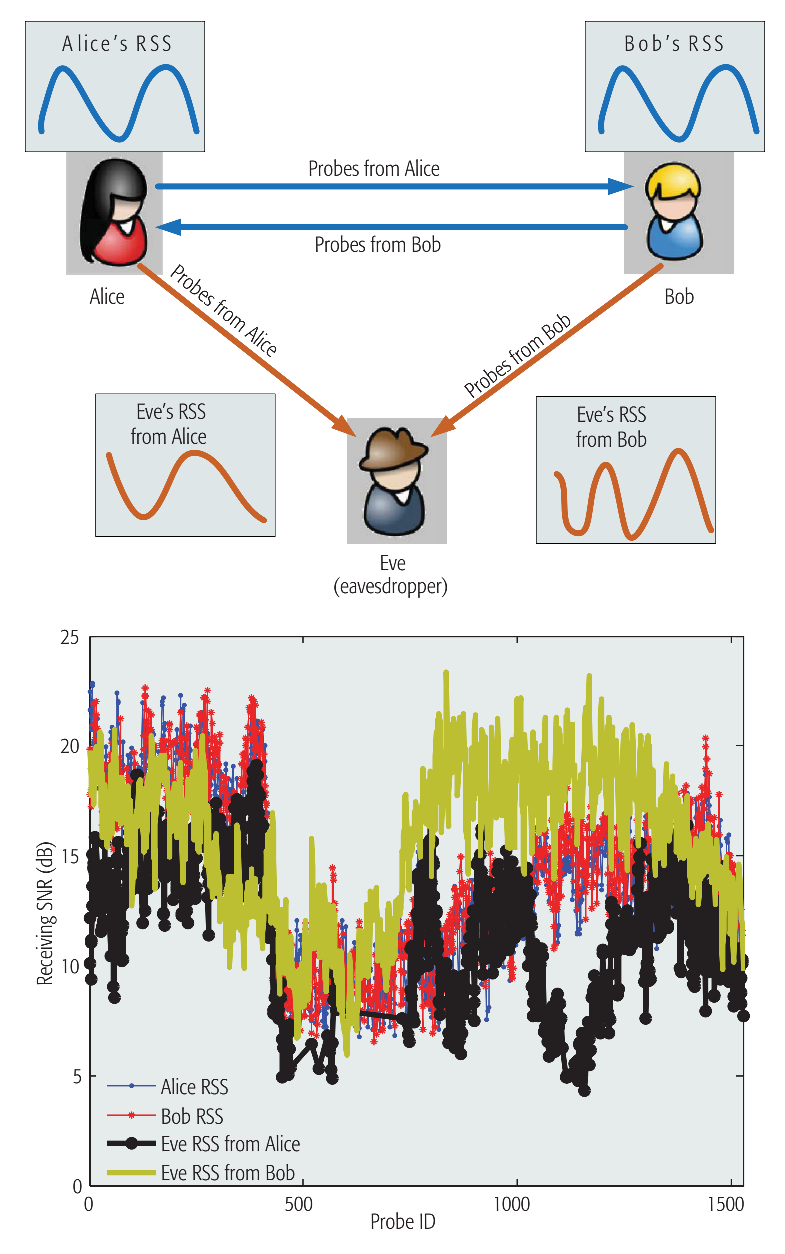


Figure 1. RSS measurements in a network: a) schematic diagram; b) experiment results from sea trials.

Decorrelation: If the time interval between successive probes is less than the correlation time of a wireless channel, there will be inherent correlation among RSS measurements. In this case, the produced key may be predictable to some extent, which must be avoided for security purposes. To address this problem, the key generation parties can use a decorrelation filter to reduce the autocorrelation within an RSS sequence.

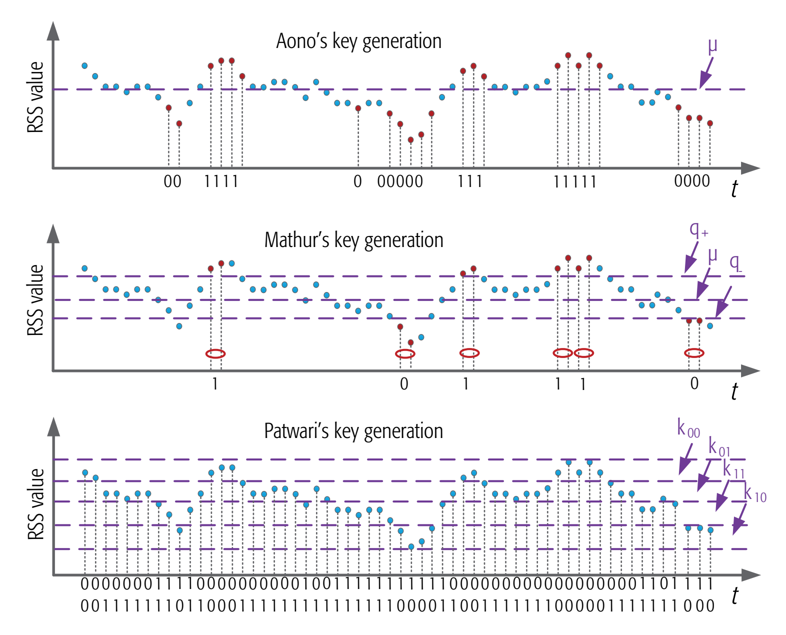


Figure 2. Key extraction algorithms.

Key Extraction

The key extraction, as the second stage of an RSS-based method, could generally be classified into two categories:

- Single-bit approaches, in which each RSS point is quantized into at most one bit
- Multi-bit approaches, which extract multiple secret bits from a single RSS measurement

In Fig. 2, we use the methods proposed by Aono [7], Mathur [8], and Patwari [9] as examples to briefly introduce how different key extraction algorithms work after signal preprocessing.

Aono's Approach: In this single-bit approach, two communicating parties first determine the length of the key, which is denoted as l ($l = 24$ in Aono's approach of Fig. 2). After that, they quantize their highest $l/2$ and the lowest $l/2$ RSS measurements into "1" and "0," respectively.

Mathur's Approach: In this single-bit approach, two communicating parties set up the thresholds q_+ and q_- based on their RSS sequences, where $q_\pm = \mu \pm \alpha \cdot \sigma$, α is a quantizer level coefficient, and μ and σ are the average and standard deviation of the RSS sequence, respectively. Then they quantize n ($n = 2$ in Mathur's approach of Fig. 2) successive RSS points above q_+ and below q_- to "1" and "0," respectively.

Patwari's Approach: In this multi-bit approach, two communicating nodes, Alice and Bob, divide the cumulative distribution of their RSS points into 4×2^k intervals, where k is the number of bits used to represent each RSS measurement ($k = 2$ in Patwari's approach of Fig. 2). The identification of interval i is denoted as m_i, $i = 1, 2, \ldots, 4 \times 2^k$. Then both nodes generate two sets of k-bit Gray codes, and each codeword is repeated four times. Let $d_j(i)$ be codeword i in set j, where $j = \{0, 1\}$ and d_0 is the circular shift of d_1. After that, Alice creates a binary vector $\mathbf{E} = [e_1, e_2, \ldots]$ based on the interval identification of her RSS points, and then sends $\mathbf{E}$ to Bob. Finally, two nodes encode their RSS measurements i with codeword d_1 if $e_i = 1$, or with codeword d_0 whenever $e_i = 0$.

Information Reconciliation

After signal preprocessing and key extraction, Alice and Bob have the shared keys with some disagreements caused by the imperfect symmetry between their RSS measurements. To agree on the same key, they need to use an information reconciliation protocol to correct the erroneous bits between their keys.

One of the most famous information reconciliation protocols is the cascade, which uses iterative processes. In a cascade, Alice divides her key into multiple blocks and sends parity information of each block to Bob. Bob divides his key and computes the parities in the same manner. After that, Bob compares his parities with those from Alice. If a difference in a parity is found, a binary search is performed to identify and correct the error bits. After all blocks have been corrected, Alice and Bob both permute the secret bits of their keys in the same random way, and start a new round of communication and correction. This process repeats a number of times to ensure that two communicating parties have identical keys with high probability.

Privacy Amplification

Information reconciliation leaks partial information during the error correction process. An adversary may utilize this information to guess the content in the key. Privacy amplification thus is advocated to solve the information leakage problem.

By applying a privacy amplification, Alice and Bob can produce a new key with fewer secret bits than the original one, while the eavesdropper, Eve, only has a little knowledge about the new key. This is achieved by letting two communicating entities use a universal hash function, which is randomly selected from a publicly known set, to produce a short key with high entropy from the original one. In order to reduce the probability that Eve has any knowledge of the new key, the length of the new key should be calculated based on how much information of the old key Eve has obtained.

Performance Metrics

When assessing the performance of RSS-based key generation approaches, the following three metrics are usually used:

- Key generation rate: The average amount of secret bits extracted from each RSS measurement.
- Bit mismatch rate: The ratio of the mismatched bits between the keys produced by Alice and Bob to the length of the key.
- *Randomness:* "A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself" [10].

The key generation rate and bit mismatch rate are the two most important metrics describing the capability of a key generation approach to produce valid bits from the raw RSS measurements, which in turn affects the minimum time to establish a secret connection between two communicating parties. Randomness is a crucial metric to evaluate the predictability of the secret key. More specifically, a key produced by two communicating parties should have enough randomness to support their secret communications. Otherwise, an eavesdropper could predict the content in the key easily once the pattern of the key is obtained. There are a number of methods to test the randomness of a sequence. For example, we could run an approximate entropy test and calculate the *P*-value of the key, which can help us evaluate the randomness of the secret bits.

Advantages of RSS-Based Key Generations

Compared to conventional cryptographic key generation schemes, the RSS-based methods have the following advantages, which make them promising techniques for UANs.

Feasibility: Any two parties that want to communicate secretly can simply use a point-to-point probe transmission protocol to generate a key without the participation of any key management entities. In addition, as a critical parameter in communication systems, the RSS could be measured by most commercial acoustic modems directly without any modification of the hardware or software.

Security: Unlike pseudorandom key generation, which has potential cryptanalytic breaks in large networks, the security of RSS-based methods is naturally preserved by the spatial diversity and random variation of an acoustic channel. Particularly, an attacker close to neither of the communicating entities measures an uncorrelated channel, and thus can hardly guess the key through overhearing. Furthermore, the high dynamics of an acoustic channel guarantees that the RSS sequences collected in different time periods are uncorrelated [11], which is a favorable feature allowing a pair of nodes to flexibly update their secret key at any time.

Resilience: RSS-based key generation schemes have high resilience, since the compromise of some good nodes will definitely not reveal the security information of other links in the network. The secret keys are essentially produced from local measurements on the channel response, which have significant diversity among different links. A pairwise key for two communication parties is unknown to any other entities. The high resilience of RSS-based key generation provides good quality of resistance against hacking attempts on the network.

Scalability: Apart from conventional pairwise key sharing schemes, which require large memory to store a considerable amount of preinstalled keys in large-scale networks, RSS-based key generation has no constraint on the memory space. Different and random keys are naturally created benefiting from the entropy feature of underwater environments. Therefore, an RSS-based scheme could operate efficiently in a large UAN or in networks with incremental deployments.

Key Connectivity: Key connectivity represents the probability that two neighboring nodes have common keys to establish a secure link for communications. High connectivity requires a large amount of shared keys on two nodes in conventional random key generators. The requirements for resilience, scalability, and connectivity, however, are conflicting in general symmetric key generation approaches. In RSS-based key generation schemes, any pair of nodes can produce shared keys as long as they are physically reachable through an acoustic channel. Upper-layer services like routing will get considerable benefit from the high connectivity of the key in RSS-based key generation schemes.

It is worth noting that RSS-based key generation approaches mainly focus on protecting the communication between authenticated parties against malicious adversaries. Like other symmetric key generators, an RSS-based scheme has no mechanism for device authentication, thereby requiring the establishment of an initial secure link between communication entities. There have been extensive authentication mechanisms in the literature [12], which could be used in conjunction with RSS-based key generation methods. For instance, similar to the Diffie-Hellman protocol, the nodes can use a public-key-based key exchange mechanism for device authentication. Another viable solution is to pre-distribute some temporary keys to authenticate the identities and exchange the initial shared secret [13].

Challenges from Underwater Environments

In oceans, electromagnetic waves suffer from heavy attenuation; as a result, the sound signal becomes the preferred information carrier for wireless communications. Underwater channels and acoustic communication systems, however, have some unique features, which in turn lead to grand challenges for RSS-based key generation approaches.

In oceans, electromagnetic waves suffer from heavy attenuation; as a result, the sound signal becomes the preferred information carrier for wireless communications. Underwater channels and acoustic communication systems, however, have some unique features, which in turn lead to grand challenges for RSS-based key generation approaches.

Long Transmission Time of a Probe Signal

Benefiting from wide bandwidth, the transmission time of a probe in the radio network is usually less than 1 ms. This allows the communicating entities to generate a key with a desired length very fast. On the contrary, due to the long preamble signal in UANs, the transmission time of a probe could be thousands of times longer than in terrestrial environments, which causes a slow generation rate of secret keys.

More specifically, for signal detection and automatic gain control (AGC) purposes, an acoustic modem needs to attach a preamble sequence before each packet. In UANs, the length of a preamble can be half a second or even longer [6], 1000 times larger than that in the radio network. This prevents communication parties from transmitting a sequence of probes in a short period as they do in terrestrial environments. Therefore, an RSS-based key generation approach in oceans needs a longer time to create a long enough secret key.

Using Mathur's single-bit approach, listed in Table 1 as an example, its key generation rate in sea tests was 0.09 b/probe. To produce a key of 128 secret bits, a node was required to send a total number of 1422 probes. In experiments, the minimum transmission time of a probe signal was 0.5 s. Therefore, two communicating entities

Approach	Signal preprocessing method	Key extraction	Parameters	Key generation rate	Bit mismatch rate (%)	Approximate Eentropy	*P*-value
Aono	Beamforming[1]	Single-bit	$l = 0.1 \times N_p$[2]	0.10	10.8	0.34	< 0.01
Mathur	Large-scale fading elimination	Single-bit	$\alpha = 0.1, n = 2$	0.09	38.5	0.68	0.43
Patwari	Interpolation and decorrelation	Multi-bit	$k = 3$	3.00	49.3	0.69	0.61

[1] Due to the hardware constraint of an acoustic modem, we excluded array processing from our experiments.
[2] N_p is the total number of RSS measurements.

Table 1. Performance comparison among three representative approaches.

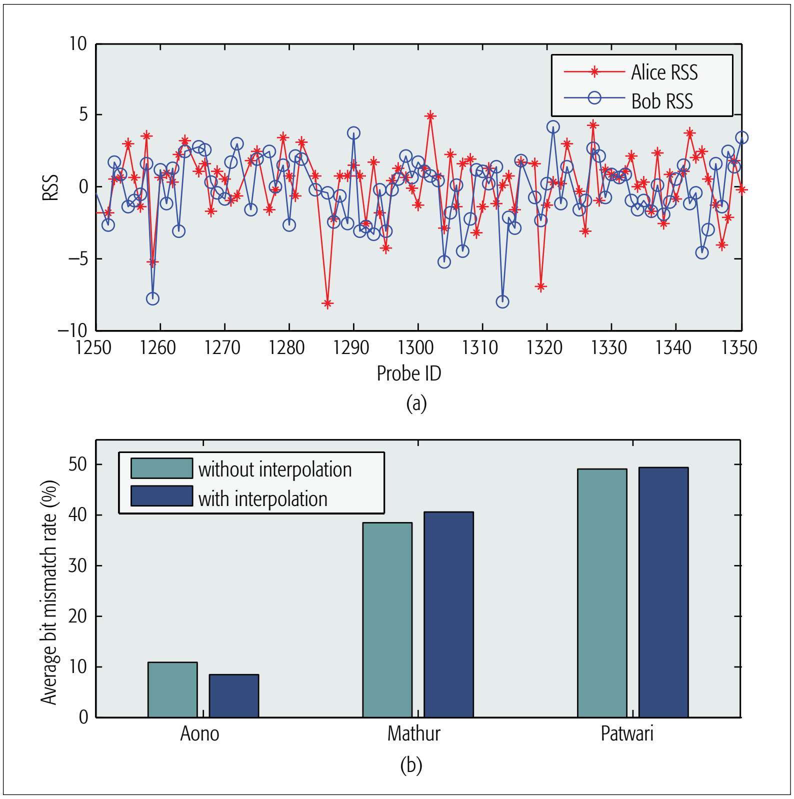


Figure 3. Large RSS discrepancies and high bit mismatch rates: a) RSS measurements on two parties; b) bit mismatch rates with and without the interpolation filter.

have to take about 13 min for probe transmissions, which makes single-bit approaches inefficient in UANs.

Given the long transmission time of probes in oceans, the efficiency of RSS-based key generation becomes a major challenge for single-bit key extraction approaches. Multi-bit key generation methods, on the other hand, can significantly improve the key generation rate, but at a cost of higher bit mismatch rate, as depicted in Table 1. How to balance the key generation rate and the bit mismatch rate in an RSS-based key generation is still an open issue in UANs.

Asymmetric RSS Measurements

RSS-based key generation relies considerably on the reciprocity of acoustic channel. However, due to the half-duplex feature of acoustic modems and the fast variation of an underwater channel, the RSS measurements of the communicating parties are not exactly symmetric, which may affect the robustness of a key generation method in terms of bit mismatch rate.

In particular, due to the size and cost constraints, most existing acoustic modems are only equipped with a single transducer for communications. These modems operate in half-duplex mode in the sense that they are capable of either transmitting or receiving. Therefore, if two communicating parties send probes simultaneously, there may be a collision between the transmission and reception, especially given the long transmission time of a probe signal in the underwater environment. To avoid collisions, two parties have to use non-concurrent probe transmissions in UANs.

In radios, due to the negligible propagation delay and the short transmission time of probes, the time difference between a pair of RSS measurements caused by non-concurrent probe transmissions is very short, usually less than the channel correlation time. In such a scenario, the effect of asymmetric RSS measurements could be mitigated by using an interpolation filter in the signal processing stage. Compared to radio networks, however, the transmission time of a probe is thousands times longer in UANs, and the propagation speed of an acoustic signal in water is five orders of magnitude lower than that of an electromagnetic wave in air. Hence, when sending probes non-concurrently, the time difference between the transmissions of a pair of probes in UANs is thousands of times longer than that in terrestrial environments. For this reason, the asymmetry of RSS measurements on a pair of nodes is significant in oceans.

Using a sea experiment as an example, we configured a node called Alice as the initiator for key generation. In each round of communication, she sent a probe signal to another node, Bob, who replied with the same signal after he received the probe from Alice successfully. This procedure was repeated multiple times until two parties got enough RSS measurements for key extraction. The distance between two communicating entities in the experiment was 556 m; therefore, the minimal time difference between an RSS pair measured by Alice and Bob was $t_p + t_s = 0.87$ s, where t_p is the propagation delay (0.37 s) and t_s is the probe transmission time (0.5 s).

Figure 3a demonstrates the asymmetric RSS measurements caused by the non-concurrent transmission of probe signals in oceans. The dif-

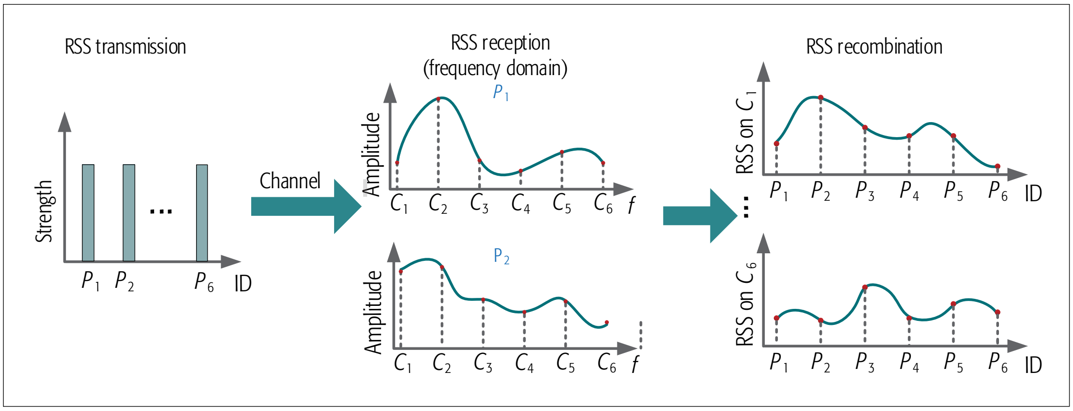

RSS transmission
Strength
P_1 P_2 ... P_6 ID
Channel
RSS reception (frequency domain)
P_1
Amplitude
C_1 C_2 C_3 C_4 C_5 C_6 f
P_2
Amplitude
C_1 C_2 C_3 C_4 C_5 C_6 f
RSS recombination
RSS on C_1
P_1 P_2 P_3 P_4 P_5 P_6 ID
RSS on C_6
P_1 P_2 P_3 P_4 P_5 P_6 ID

Figure 4. Multi-channel key generation scheme, where P_i is the probe ID and C_i is the ID of a subchannel.

According to the environmental conditions, we can select different smooth filters to achieve good performance for the key generation. For instance, if the probe signals are polluted by strong ambient noise, a Savitzky-Golay filter or a symmetric moving average filter is recommended to improve the reciprocity of RSS sequences.

ferences finally result in a high bit mismatch rate on the RSS-based key generation approaches, as listed in Table 1. In addition, this bit mismatch rate cannot be evidentially reduced by applying an interpolation filter, as shown in Fig. 3b.

Solutions for Performance Improvements

As discussed earlier, the long transmission time of probes leads to a low generation rate of secret keys, and the asymmetric RSS measurements result in much disagreement between the secret bitstreams. In this section, we provide some feasible solutions to tackle these two problems.

Improvement of the Key Generation Rate

Intuitively, we can increase the key generation rate by using a multi-bit approach in the key extraction stage. However, compared to single-bit methods, the conventional multi-bit RSS-based key generation scheme is susceptible to imperfect asymmetry between RSS measurements, thereby resulting in a high disagreement probability between the shared keys. As listed in Table 1, the bit mismatch rate of the multi-bit approach proposed by Patwari is near 50 percent in oceans, which requires an overhaul before applying it to UANs.

To keep the advantages of low bit-mismatch rate in single-bit methods and high key generation rate in multi-bit approaches, a viable solution is to use the scheme of multi-channel key generation. In this scheme, the nodes divide the communication bandwidth into multiple independent subchannels, and the probe signal, such as an orthogonal frequency-division multiplexing (OFDM) signal, should have at least one frequency component on each subchannel. After receiving the probes, the receiver transforms the received signal to the frequency domain.

With this scheme, the RSS measurements can be performed on each subchannel, producing independent RSS sequences. The communicating parties can harvest a total number of N_{sub} RSS measurements from each probe reception, where N_{sub} is the number of subchannels, as shown in Fig. 4.

Obviously, the more subchannels we use, the higher the key generation rate we can achieve. However, there is a possibility that the RSS measurements at neighboring subchannels are correlated if their frequency interval is small. In this case, the randomness of the secret bits will decrease. For this reason, we should choose the subchannels not close to each other in the frequency domain.

Improvement of Key Agreement Probability

Evidentially, the key disagreement in RSS-based key generation approaches can be reduced by improving the symmetry of RSS sequences between two communicating entities. However, due to the large time difference of non-concurrent RSS transmissions in oceans, there could be considerable discrepancies in RSS measurements. Therefore, the interpolation filter, which is usually adopted in radio networks, may fail in UANs. Here we advocate the use of a smooth filter in the signal preprocessing stage to improve key robustness.

The smooth filter has been widely applied in many areas, such as statistics and image precessing. It is an efficient way to capture the critical features in data, while removing the fast varying components like noise and interference. By using a smooth filter in the RSS-based key generation approach, the two parties can reduce the random fluctuations in their RSS measurements and thus decrease the bit mismatch rate of the key.

According to the environmental conditions, we can select different smooth filters to achieve good performance for key generation. For instance, if the probe signals are polluted by strong ambient noise, a Savitzky-Golay filter or a symmetric moving average filter is recommended to improve the reciprocity of RSS sequences. If the symmetry of RSS sequences is degraded by a burst interference, such as the signal from a sonar or a marine mammal, a robust local polynomial regression (LOESS) filter is preferred.

In Fig. 5, we use experiment results to verify the efficiency of a smooth filter in reducing the bit mismatch rates for the three representative RSS-based key generation approaches. As demonstrated in Fig. 5, the bit mismatch rates of all three approaches are significantly reduced with the increased size of a smooth window. More specifically, by using a symmetric moving average smooth filter in the experiments, the average bit mismatch rates of Aono, Mathur, and Patwari dramatically decrease by 100, 63, and 20 percent, respectively.

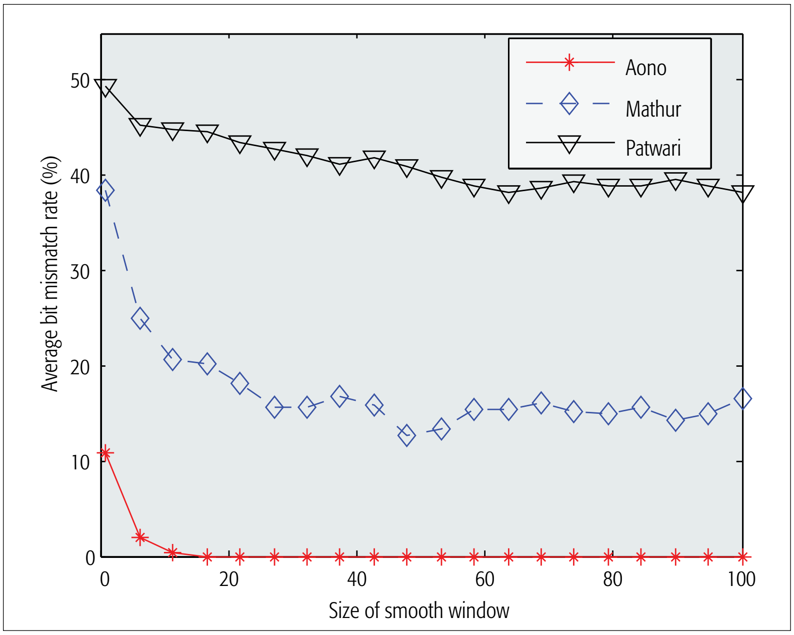


Figure 5. Bit mismatch rates with respect to the size of the smooth window.

It is worth noting that there is a trade-off between the mismatch rate and the randomness of the secret key when a smooth filter is applied. According to the experiment results, the P-value of the approximate entropy test would be less than 0.01 for Mathur's and Patwari's key generation approaches if the size of the smooth window is over 15.

Conclusion

In this article, we have presented a tutorial on RSS-based key generation approaches in underwater environments for secret acoustic communications. While these approaches have been well studied in terrestrial networks, they face many new challenges that have yet to be addressed due to the unique features of acoustic systems.

From the experiment results we observe that:

- The transmission time of a probe signal in UANs is much longer than that in radio networks and thus results in a low key generation rate.
- Due to the long propagation delay and large transmission time, the asymmetry of RSS measurements between two communicating parties is more significant in UANs than in radio networks, which causes a high bit mismatch rate on the shared key.

Finally, we have introduced two solutions to improve the performance of RSS-based key generation approaches in terms of key generation rate and bit mismatch rate. The multi-channel key generation scheme enables communicating parties to extract secret bits on multiple subchannels, thereby significantly improving the efficiency of key generation. A smooth filter can improve the symmetry of RSS measurements, thereby reducing the discrepancies between the shared keys. According to experiment results from sea trials, we have seen that by using a symmetric moving average smooth filter, the average bit mismatch rates of the approaches proposed by Aono, Mathur, and Patwari were decreased by 100, 63, and 20 percent, respectively.

References

[1] J. Heidemann, M. Stojanovic, and M. Zorzi, "Underwater Sensor Networks: Applications, Advances and Challenges," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 370, 2012, pp. 158–75.

[2] M. Guillaud, D. T. Slock, and R. Knopp, "A Practical Method for Wireless Channel Reciprocity Exploitation Through Relative Calibration," *Proc. Int'l. Symp. Signal Processing and Its Applications*, 2005, pp. 403–06.

[3] G. D. Durgin, *Space-Time Wireless Channels*, Prentice Hall Professional, 2003.

[4] K. Ren, H. Su, and Q. Wang, "Secret Key Generation Exploiting Channel Characteristics in Wireless Communications," *IEEE Wireless Commun.*, vol. 18, no. 4, 2011, pp. 6–12.

[5] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater Acoustic Sensor Networks: Research Challenges," *Ad Hoc Networks*, vol. 3, no. 3, 2005, pp. 257–79.

[6] L. Pu *et al.*, "Impact of Real Modem Characteristics on Practical Underwater MAC Design," *Proc. IEEE OCEANS-Yeosu*, May 2012.

[7] T. Aono *et al.*, "Wireless Secret Key Generation Exploiting Reactance-Domain Scalar Response of Multipath Fading Channels," *IEEE Trans. Antennas and Propagation*, vol. 53, no. 11, 2005, pp. 3776–84.

[8] S. Mathur *et al.*, "Radiotelepathy: Extracting a Secret Key from an Unauthenticated Wireless Channel," *Proc. ACM MobiCom*, 2008, pp. 128–39.

[9] N. Patwari *et al.*, "High-Rate Uncorrelated Bit Extraction for Shared Secret Key Generation from Channel Measurements," *IEEE Trans. Mobile Computing*, vol. 9, no. 1, 2010, pp. 17–30.

[10] A. N. Kolmogorov, "Three Approaches to the Quantitative Definition of Information," *Problems of Information Transmission*, vol. 1, no. 1, 1965, pp. 1–7.

[11] X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications*, Springer, 2002.

[12] Q. Huang *et al.*, "Fast Authenticated Key Establishment Protocols for Self-Organizing Sensor Networks," *Proc. Wireless Sensor Networks and Applications*, ACM, 2003, pp. 141–50.

[13] D. S. Wong and A. H. Chan, "Mutual Authentication and Key Exchange for Low Power Wireless Communications," *Proc. IEEE MILCOM*, 2001, pp. 39–43.

Biographies

Yu Luo (yu.luo@engr.uconn.edu) received his B.S. and M.S. degrees in electrical engineering from Northwestern Polytechnical University, China, in 2009 and 2012, respectively. Currently, he is pursuing a Ph.D. degree in computer science and engineering at the University of Connecticut, Storrs. His major research focus is on cross-layer design for cognitive acoustic networks and underwater acoustic networks. He was a co-recipient of the IFIP Networking 2013 best paper award.

Lina Pu (lina.pu@engr.uconn.edu) received her B.S. degree in electrical engineering from Northwestern Polytechnical University in 2009. She is currently working toward a Ph.D. degree in the Department of Computer Science and Engineering at the University of Connecticut, Storrs. Her research interests lie in the area of MAC design, performance evaluation, and experimental study for underwater acoustic networks. She was a co-recipient of the IFIP Networking 2013 best paper award.

Zheng Peng (zhengpeng@engr.uconn.edu) received Bachelor's degrees in control science and computer science from Zhejiang University, Hangzhou, China, in 2002. He obtained a Master's degree in computer science from the University of Electrical Science and Technology of China in 2005, and later his Ph.D. from the University of Connecticut, Storrs. His main research interests cover the design, modeling, optimization, development, and experimental evaluation of wireless sensor networks, embedded/distributed systems, and their applications in the unique underwater environment.

Zhijie Shi (zshi@engr.uconn.edu) is currently an associate professor in the Department of Computer Science and Engineering at the University of Connecticut. He received his Ph.D. degree from Princeton University in 2004, and his M.S. and B.S. degrees from Tsinghua University, China, in 1996 and 1992, respectively. He received a U.S. National Science Foundation CAREER award in 2006. His current research interests include underwater sensor networks, network security, hardware mechanisms for secure and reliable computing, and primitives for cipher designs.

Call for Papers
IEEE Communications Magazine
Communications Standards Supplement

Background

Communications standards enable the global marketplace to offer interoperable products and services at affordable cost. Standards development organizations (SDOs) bring together stakeholders to develop consensus standards for use by a global industry. The importance of standards to the work and careers of communications practitioners has motivated the creation of a new publication on standards that meets the needs of a broad range of individuals, including industrial researchers, industry practitioners, business entrepreneurs, marketing managers, compliance/interoperability specialists, social scientists, regulators, intellectual property managers, and end users. This new publication will be incubated as a Communications Standards Supplement in *IEEE Communications Magazine*, which, if successful, will transition into a full-fledged new magazine. It is a platform for presenting and discussing standards-related topics in the areas of communications, networking, and related disciplines. Contributions are also encouraged from relevant disciplines of computer science, information systems, management, business studies, social sciences, economics, engineering, political science, public policy, sociology, and human factors/usability.

Scope of Contributions

Submissions are solicited on topics related to the areas of communications and networking standards and standardization research, in at least the following topical areas:

Analysis of new topic areas for standardization, either enhancements to existing standards or of a new area. The standards activity may be just starting or nearing completion. For example, current topics of interest include:

- 5G radio access
- Wireless LAN
- SDN
- Ethernet
- Media codecs
- Cloud computing

Tutorials on, analysis of, and comparisons of IEEE and non-IEEE standards. For example, possible topics of interest include:

- Optical transport
- Radio access
- Power line carrier

The relationship between innovation and standardization, including, but not limited to:

- Patent policies, intellectual property rights, and antitrust law
- Examples and case studies of different kinds of innovation processes, analytical models of innovation, and new innovation methods

Technology governance aspects of standards focusing on both the socio-economic impact as well as the policies that guide it. These would include, but are not limited to:

- The national, regional, and global impacts of standards on industry, society, and economies
- The processes and organizations for creation and diffusion of standards, including the roles of organizations such as IEEE and IEEE-SA
- National and international policies and regulation for standards
- Standards and developing countries

The history of standardization, including, but not limited to:

- The cultures of different SDOs
- Standards education and its impact
- Corporate standards strategies
- The impact of open source on standards
- The impact of technology development and convergence on standards

Research-to-standards, including standards-oriented research, standards-related research, research on standards

Compatibility and interoperability, including testing methodologies and certification to standards

Tools and services related to any or all aspects of the standardization life cycle

Proposals are also solicited for Feature Topic issues of the Communications Standards Supplement.

Articles should be submitted to the *IEEE Communications Magazine* submissions site at

http://mc.manuscriptcentral.com/commag-ieee

Select “Standards Supplement” from the drop-down menu of submission options.

Design Guidelines for Opportunistic Routing in Underwater Networks

Rodolfo W. L. Coutinho, Azzedine Boukerche, Luiz F. M. Vieira, and Antonio A. F. Loureiro

The authors discuss the two main building blocks for the design of opportunistic routing protocols for underwater sensor networks: candidate set selection and candidate coordination procedures. They propose classifying candidate set selection procedures into sender-side, receiver-side, and hybrid approaches, and candidate coordination procedures into timer-based and control-packet-based approaches.

ABSTRACT

The unique characteristics of the underwater acoustic channel impose many challenges that limit the utilization of underwater sensor networks. In this context, opportunistic routing, which has been extensively investigated in terrestrial wireless ad hoc network scenarios, has greater potential for mitigating drawbacks from underwater acoustic communication and improving network performance. In this work, we discuss the two main building blocks for the design of opportunistic routing protocols for underwater sensor networks: candidate set selection and candidate coordination procedures. We propose classifying candidate set selection procedures into sender-side, receiver-side, and hybrid approaches, and candidate coordination procedures into timer-based and control-packet-based approaches. Based on this classification, we discuss particular characteristics of each approach and how they relate to underwater acoustic communication. Furthermore, we argue that those characteristics should be considered during the design of opportunistic routing protocols for different scenarios in underwater sensor networks.

INTRODUCTION

Research about oceans has become increasingly necessary in recent years. Oceans represent around 2/3 of the Earth's surface and play an important role in sustaining human life. They are a substantial source of primary global production, absorb most of the carbon dioxide (CO_2) emitted into the atmosphere, and regulate the Earth's climate. Despite the oceans' importance, it is estimated that 95 percent of their volume remains unknown.

Underwater sensor networks (UWSNs) have great potential to help change the aforementioned reality. UWSN has been proposed as an alternative solution for observing and exploring aquatic environments against the traditional wired and communicationless technologies. By providing nodes with underwater wireless communication capabilities, UWSNs enable real-time monitoring and actuation, online system reconfiguration, and failure detection [1]. This novel technology has enabled a new era in scientific and industrial underwater monitoring applications, such as ocean exploration, oceanographic data collection, ocean and offshore sampling, navigation assistance, and tactical surveillance applications.

Recently, opportunistic routing (OR) has been proposed for tackling channel fading, which diminishes the routing performance of traditional routing paradigms. Instead of a unique next-hop forwarder selected in traditional multihop routing, OR selects a set of next-hop forwarder candidates that can overhear the packet transmission and continue forwarding it in a prioritized way toward the destination. Packet retransmission occurs only if it is not received by any candidate. The importance of OR for underwater networks arises from the challenges imposed by underwater acoustic communication, which is characterized by strong attenuation and ambient noise, time-varying multipath propagation, and low-speed sound propagation (≈1500 m/s). These aspects result in a high delay and error rate, temporary loss of connectivity, limited bandwidth capacity, and high energy communication cost. Thus, OR can help mitigate underwater channel effects and enhance the poor underwater acoustic physical links by taking advantage of the broadcast nature of the wireless transmission medium.

In this article, we examine the opportunistic routing protocol design problem for underwater sensor networks. More specifically, our contributions include:

- A thorough review of underwater acoustic channel characteristics that present different challenges to data delivery and motivate the research on OR in underwater networks
- Investigation of candidate set selection and transmission coordination procedures for the design of OR protocols for UWSNs
- Discussion of design considerations of OR protocols according to the candidate set selection and transmission coordination approaches

BACKGROUND AND PROBLEM SETTING

In this article, we consider the network architecture depicted in Fig. 1. Each sensor node is responsible for monitoring its surrounding region and sending the gathered data to a surface sonobuoy through multihop underwater acoustic communication links. A surface sonobuoy (sink) is responsible for receiving, through acoustic

Rodolfo W. L. Coutinho is with the Federal University of Minas Gerais and the University of Ottawa; Azzedine Boukerche is with the University of Ottawa; Luiz F. M. Vieira and Antonio A. F. Loureiro are with the Federal University of Minas Gerais.

0163-6804/16/$25.00 © 2016 IEEE

communication, collected data from the sensor nodes and sending it, through radio frequency communication, to a monitoring center.

Currently, the underwater acoustic channel is the only feasible technology for medium- to long-range underwater wireless communication. Therefore, UWSN links are affected by high path loss, noise, multipath signal propagation, limited bandwidth capacity, Doppler spreading, and high power consumption. Path loss is mainly caused by geometrical spreading and signal attenuation associated with frequency-dependent absorption. Noise originates from ambient and site-specific sources. Ambient noise is always present. Turbulence, shipping activities, breaking waves, and thermals are the primary sources of ambient noise. Another important characteristic of the underwater acoustic channel is that the bandwidth depends on transmission power and distance, and radio frequency. For a more detailed overview of the characteristics of underwater acoustic channels, the interested reader might refer to [2].

The aforementioned characteristics severely degrade underwater wireless link reliability. Due to underwater acoustic channel characteristics, a link between neighbors can perform poorly or even be down at any given moment. Temporary connectivity loss can occur due to shadow zones. A major result of this is a potential increase in retransmissions as an attempt to deliver data packets. Such an increase would create more packet collisions, delay, and energy consumption.

In this context, OR has proven a promising paradigm to design routing protocols for UWSNs [3]. In traditional multihop routing, a packet is transmitted to a specific next-hop node using unicast communication. If the next-hop node does not receive it, the packet should be retransmitted. After a finite number of unsuccessful retransmissions, the packet is discarded. In OR, a set of candidate nodes is involved for advancing the packet toward the destination. Accordingly, a packet is sent leveraging the broadcast nature of the wireless transmission. The candidates that receive the packet will continue, in coordination, to forward it in a prioritized way, such that a low-priority node transmits the packet if none of the high-priority nodes have done so. Therefore, a packet is retransmitted only if none of the candidates have received it.

The OR paradigm has benefits and drawbacks that must be considered when it is used to design routing protocols for underwater networks. Opportunistic routing increases packet delivery and decreases the number of packet collisions, since the probability that at least one candidate would correctly receive the packet is high compared to traditional unicast routing. However, the packet delivery end-to-end delay is high due to the nodes' transmission coordination. Also, the harsh underwater communication environment can result in poor transmission nodes' coordination, culminating in redundant packet transmissions, increasing packet collisions, and delay and energy consumption. Moreover, the assignment of the same high priority for some candidates may deplete their energy sooner, leading to partitions in the network.

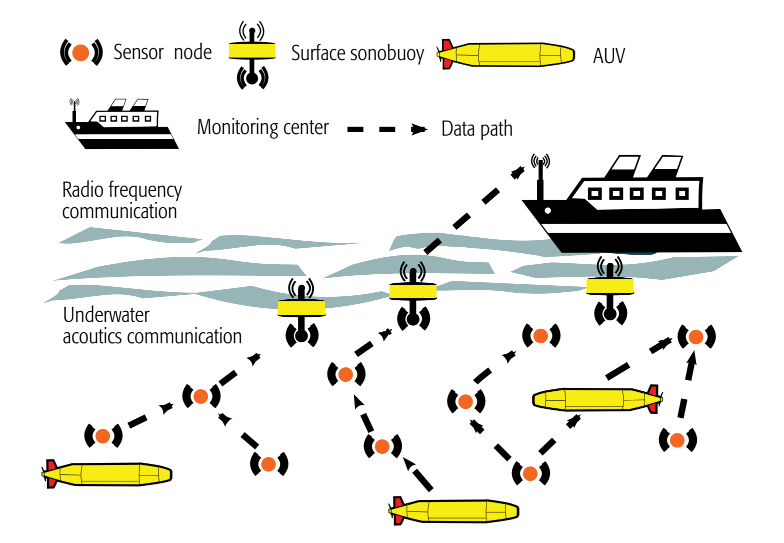

Sensor node
Surface sonobuoy
AUV
Monitoring center
Data path
Radio frequency communication
Underwater acoutics communication

Figure 1. Underwater sensor network architecture.

We can state the problem of OR for underwater sensor networks as follows. Let an underwater sensor network be represented by a Euclidean graph G=($\mathcal{V}$, $\mathcal{E}$) with the following properties:

- $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$ is a finite set of nodes.
- $\mathcal{E}$ is a set of links, where $(i, j) \in \mathcal{E}$ if v_i reaches v_j, that is, the distance between them is less than the communication range r.
- $\mathcal{N}_i$ is the neighboring set of node i, composed of all vertices j, such that $(i, j) \in \mathcal{E}$.
- Each link $(i, j) \in \mathcal{E}$ has an associated cost representing packet delivery error between nodes i and $j \in \mathcal{V}$.

For a source node v_i, an OR protocol must determine a next-hop forwarder candidate set $\mathcal{C}_i \subseteq \mathcal{N}_i$, according to a metric such as distance progress, transmission count, or one-hop throughput. In the candidate set $\mathcal{C}_i$, nodes are ordered in a prioritized way, $v_1 > v_2 > \ldots > v_n \in \mathcal{C}_i$, such that the predecessor node will continue forwarding the packet toward the destination only if its successor node fails to do so.

The Components of Opportunistic Routing

Opportunistic routing protocols are composed of two main building blocks.

Candidate Set Selection: This procedure is responsible for selecting a subset of neighboring nodes to continue forwarding the packet toward the destination. In this article, we categorize candidate set selection procedures in sender-side-based, receiver-side-based, and hybrid approaches. These terms are elucidated in Definitions 1, 2, and 3, respectively.

Candidate Set Coordination: This procedure is responsible for coordinating the packet forwarding operation between the next-hop candidate nodes. Moreover, this procedure is also responsible for determining suppression of the redundant packet transmissions of low-priority nodes. In this article, we categorize the candidate set coordination in timer-based and control-packet-based approaches. These terms are elucidated in Definitions 4 and 5, respectively.

Definition 1 (sender-side-based candidate set selection): Procedures in which the candidate

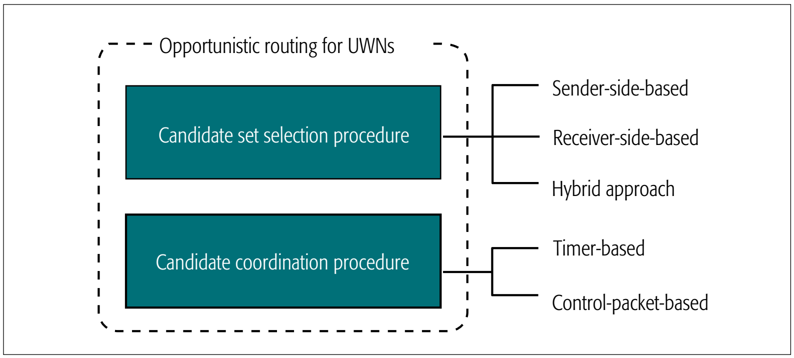

Opportunistic routing for UWNs
Candidate set selection procedure
Candidate coordination procedure
Sender-side-based
Receiver-side-based
Hybrid approach
Timer-based
Control-packet-based

Figure 2. Classification of opportunistic routing building blocks for underwater sensor networks.

set is determined by the current forwarder node when it has a data packet to transmit.

Definition 2 (receiver-side-based candidate set selection): Procedures in which the candidate set is determined by the neighbors when they receive the data packet, that is, each neighbor is responsible for verifying whether it is a next-hop forwarder candidate.

Definition 3 (hybrid candidate set selection): Procedures in which the candidate set is determined cooperatively by the current forwarder node and its neighbors.

Definition 4 (timer-based candidate set coordination): Procedures that employ holding time to coordinate candidates' data packet transmissions.

Definition 5 (control packet-based candidate set coordination): Procedures that employ control packet exchange to coordinate candidates' data packet transmissions.

Figure 2 depicts the OR building blocks and the categorization we have proposed to better describe some design principles of each approach. If we use an OR protocol in underwater sensor networks, we must execute the following steps when a node has a data packet to deliver to a surface sonobuoy. First, the candidate set selection procedure determines a subset of the neighboring nodes (candidate set) to forward the packet. Second, either the candidate nodes' ID or other indicative information is included in the packet header to be used by the candidate nodes. After that, the current forwarder node broadcasts the packet. Candidate nodes that have successfully received the packet initiate the forwarding procedure according to their priority levels. Finally, low-priority candidate nodes suppress the packet transmission when it is sent by a high-priority node.

Candidate Set Selection Procedures

The candidate set selection procedure is responsible for choosing a subset of the neighboring nodes to continue forwarding the packet. A fitness function or a single metric, such as the expected distance progress or expected transmission count, is used to determine the suitability of each neighbor. The neighbors are sorted according to their suitability, and are then included in the next-hop candidate set, which eventually becomes subject to restriction, such as a limited number of candidates or a one-hop candidate set delay.

In terrestrial wireless networks, it is common to have candidate set selection procedures that consider the whole network topology, or a recurrence function to determine the fitness of each neighbor. For instance, Li *et al.* [4] analyze the candidate set selection problem, developing a dynamic programming algorithm to determine an optimal solution that will minimize the expected number of transmissions. A similar strategy seems inappropriate for an underwater sensor network due to the high cost of disseminating the topology information to all the nodes, given the unique characteristics of the underwater acoustic channel and node mobility.

We concentrate only on local procedures for the candidate set selection. In those approaches, the next-hop candidate set is determined by each neighboring node or by the current forwarder node with k-hop neighborhood information for a small value of k (e.g., $k = 1$ or $k = 2$). In a broader sense, candidate set selection procedures can be classified as *sender-side*, *receiver-side*, and *hybrid* procedures. In the following, we describe each category, and present some proposed OR protocols in underwater sensor networks.

Sender-Side Procedures

In sender-side candidate set selection procedures, the current forwarding node is responsible for selecting the next-hop forwarder candidate set. It selects the forwarder nodes based on neighborhood information. Usually, nodes use periodic beaconing to acquire neighborhood information. Based on the neighborhood information and the next-hop selection metric or function, the current forwarder node determines which neighbors are enabled to continue forwarding the packet toward the destination. After that, the enabled neighbors are sorted and included in the next-hop candidate set, according to their priorities. Finally, the unique ID of the chosen nodes is included in the packet header. Frequently, a Bloom filter or membership checking data structures are used to avoid long packet headers, which would increase the packet error rate.

With approaches that fall into this category, as the neighborhood is known in advance, more complex and multi-objective fitness functions, which consider application and underwater acoustic channel characteristics, can be used to select the next-hop candidate set. For instance, buffer and distance information of the neighbors can be used by the next-hop candidate set selection procedure in a real-time application (e.g., oil spill monitoring) to ensure an acceptable packet reception ratio with limited delay. Moreover, only nodes close enough to hear each other's transmission can be selected to belong to the next-hop candidate set in order to mitigate the problem of hidden terminals. The main drawback of this approach stems from the need to keep updated neighborhood information at the nodes, considering the node mobility caused by ocean currents. The use of beacon messages can overload the channel and increase packet collisions because of the slow propagation through acoustic channels.

HydroCast [5] routing protocol is an example of underwater OR protocol that implements a sender-side-based candidate set selection pro-

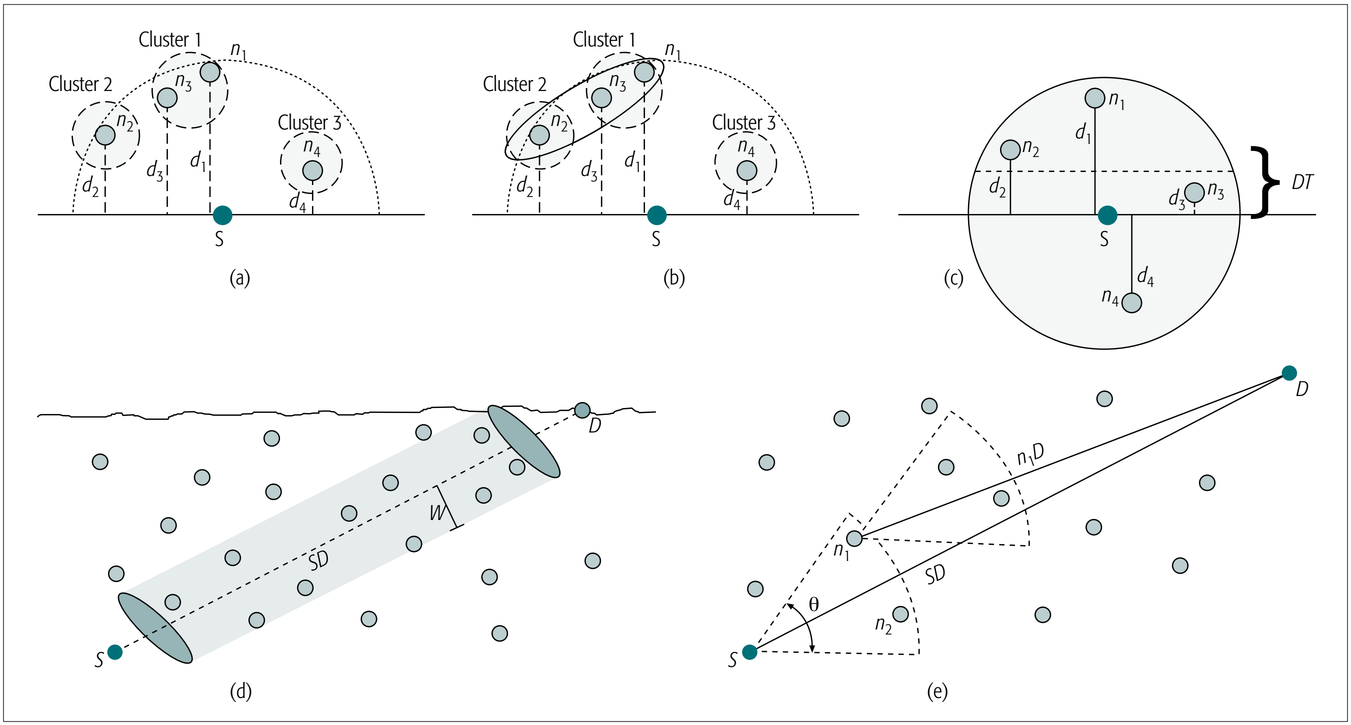


Figure 3. Candidate set selection procedures: a) HydroCast candidate set selection procedure — part 1; b) HydroCast candidate set selection procedure — part 2; c) DBR candidate set selection procedure; d) VBF candidate set selection procedure; e) FBR candidate set selection procedure.

cedure. Its candidate set selection procedure employs neighbors with positive packet advancement (i.e., closer to the sea surface). The protocol first calculates the fitness of each node using normalized advancement (NADV). Second, the neighbor with the highest NADV and all nodes distant at most βR of it are chosen to form a cluster. Those steps are repeated until no neighbor remains. For instance, in Fig. 3a, *clusters 1*, *2*, and *3* are determined after the aforementioned step. Each cluster is then expanded by including all nodes with a shorter distance to nodes already in the cluster than the communication range. In the example of Fig. 3b, node n_2 is included in *cluster 1*, and nodes n_1 and n_3 in *cluster 2*. Finally, the cluster with the greatest expected progress (a normalized sum of advancements made by the nodes) is selected, and the IDs of its nodes are included in the packet header. With minor variations, the aforementioned next-hop candidate set selection procedure is used by many other OR protocols in underwater sensor networks, such as GEDAR [6]. A main disadvantage of HydroCast is that it employs a simplified underwater channel model to calculate the NADV, which may not reflect the reality in some scenarios.

Receiver-Side Procedures

In receiver-side candidate set selection procedures, the neighboring nodes are responsible for determining whether they are included in the next-hop forwarder candidate set of a received packet. The current forwarder includes some control information (e.g., its depth or distance to the destination) in the packet header, and then broadcasts it. Each receiver node, from the information contained in the packet, determines locally whether it is a next-hop candidate node according to the rule adopted by the protocol. If the neighboring node is a candidate, it determines its forwarding priority and forwards the packet if and when it should do so.

The receiver-side-based next-hop candidate set selection procedures are simple and scalable, requiring no neighborhood information. Because of this, energy conservation and increased underwater acoustic channel utilization can be achieved. Moreover, these procedures are indicated for high traffic load applications, such as pollution monitoring, since there are no control packets competing for access to the acoustic channel and colliding with the data packets. However, the candidate coordination and redundant packet suppression can be inefficient, leading to a high number of duplicated packet transmissions, which unnecessarily consume energy and do not provide any innovative information to the destination, being discarded when it is reached.

VBF [7] and DBR [8] routing protocols are examples of underwater OR protocols that implement a receiver-side-based candidate set selection procedure. In the DBR routing protocol, the current forwarder node includes its depth information in the packet and broadcasts it. Upon receiving a packet, each neighboring node compares its depth with the sender's depth. The neighbor is a next-hop candidate if it is closer to the sea surface than the current forwarder node, and this depth difference is higher than a *DT* depth threshold. For instance, considering the current forwarder node *S* in Fig. 3c, nodes n_3 and n_4 will discard the received packet because n_3 and n_4 do not have a depth difference higher than *DT*, and do not advance the packet toward the surface, respectively. Nodes n_1 and n_2 are

While a candidate set with only two or three nodes may be sufficient for obtaining a high packet delivery probability in terrestrial radio frequency-based wireless networks, in underwater networks this number should be greater due to high signal attenuation and shadow zones.

candidates for forwarding the packet. In this example, node n_1 is the high-priority node. Node n_2 only forwards the packet if n_1 fails to do so. In the VBF routing protocol, packets are routed along a virtual pipe (please refer to Fig. 3d). The location information of the sender and the destination is included into the packet header to be used for the purpose of next-hop candidate selection. When a node receives a packet, it determines whether it is inside the routing pipe by calculating its distance to the vector formed by the sender and destination location information. If its distance to the vector is smaller than W, the node is a next-hop candidate. Otherwise, the node discards the packet. VBF and DBR protocols suffer from similar drawbacks. Networks with higher densities are susceptible to a large number of duplicated transmissions. This is because both protocols do not employ any type of mechanism for restricting the number of candidates in the set, and do not address the hidden terminal problem. In low-density scenarios, data delivery will be compromised, as VBF and DBR do not employ a communication void region mechanism.

Hybrid Procedures

In hybrid candidate set selection procedures, the next-hop forwarder candidate set is determined by the current forwarder node and its neighbors in two distributed steps. When a node has a packet to transmit, it reports its situation and requests information (e.g., battery level or link reliability) from its neighborhood through a broadcast of a control packet. Neighbor nodes meeting the criteria (e.g., positive progress to the destination) will respond to the current forwarder node with the requested information. In the end, the current forwarder node selects the next-hop candidate set based on received packets.

In the solutions belonging to this category, the neighborhood condition is known on demand, in contrast to the procedures based on the sender-side-based candidate set selection, where beaconing happens periodically. Moreover, short packets to the next-hop candidate selection and coordination are less likely to be lost, resulting in an improvement in candidate coordination performance. This approach is highly suitable for low traffic load monitoring applications, such as periodic ocean temperature or salinity monitoring applications. However, this two-way procedure can increase end-to-end delay, due to the slow propagation through underwater acoustic channels.

FBR [9] and CARP [10] are examples of underwater routing protocols that implement a hybrid candidate set selection procedure. It is worth noting that these protocols are not opportunistic in the sense that only one neighbor node is selected as next-hop forwarder (i.e., only one candidate as the next hop). However, we present them as an example of a hybrid solution, since they can easily be extended to incorporate the OR nature by determining a sorted list of next-hop candidate nodes. In both routing protocols, when a node has a packet to send, it broadcasts a control packet to inform its neighbors and obtain additional information. Each neighbor meeting the desired requirement to be a forwarder candidate replies to the current forwarder node. The desired requirement is to lie within a cone of angle ±$\theta/2$ emanating from the transmitter toward the destination in FBR (e.g., nodes n_1 and n_2 in Fig. 3e) and having a hop count to the destination less than the source node in CARP. The next-hop candidate is selected among the neighbors that replied to the sender request, based on their load or proximity to the destination, and a link quality goodness function in FBR and CARP, respectively.

Candidate Coordination Procedures

The transmission coordination of the next-hop forwarder candidate nodes is a crucial component of OR design protocol. With the inclusion of many nodes in the candidate set, link reliability improves, and the average number of transmissions required to deliver a packet is reduced. However, the candidates must forward the packet in a coordinated manner, such that a lower-priority node will transmit the packet only if the higher-priority nodes fail to do so. This will avoid transmission of unnecessary and redundant packets, which will consume energy and fail to provide any additional information, being discarded at the destination.

The coordination of candidate transmission is a more difficult problem in underwater networks than in terrestrial networks. While a candidate set with only two or three nodes may be sufficient for obtaining a high packet delivery probability in terrestrial radio-frequency-based wireless networks, in underwater networks this number should be greater due to high signal attenuation and shadow zones.

In this article, we categorize candidate coordination procedures based on *timer-based coordination* and *control-packet-based coordination*. We anticipate that current state-of-the-art OR protocols have mostly used timer-based procedures for candidate coordination. This is due to its simplicity and the absence of extra packet transmissions, which might overload the underwater acoustic channel and increase energy consumption. In the following we discuss the advantages and disadvantages of each category.

Timer-Based Candidate Coordination

Upon receiving a packet in the timer-based coordination procedure, each candidate node holds it for a period of time or a number of transmission slots, according to its priority. The candidate will suppress its transmission if it receives an indication during its waiting period that the packet was forwarded by a high priority node. Usually, this indication is the reception of the same packet, now coming from a high-priority node or an acknowledgment (ACK) packet. Otherwise, the node forwards the packet when its holding time expires.

The main advantage of this approach is the absence of extra control packets, which can further degrade network performance. However, as no control packet is used to coordinate and inform low-priority nodes, duplicated data packet transmissions can occur when the nodes are far from one another and cannot hear transmissions. In this case, duplicated data packet transmissions have a more negative impact on network performance and energy consumption than the use of control packets due to their differences in size.

HydroCast [5], VBF [7], and DBR [8] routing protocols are examples of OR protocols for underwater sensor networks that employ timer-based candidate coordination procedures. Their packet holding time function is presented in Table 1, where R is the node's communication range and d is the packet advancement made by the node. The VBF holding time function has two terms. The first term determines holding time based on the node's desirability factor (in other words, its priority) and a predefined maximum delay, such that it waits less time when the node's priority is higher. The second term computes the remaining time needed for all the nodes in the current forwarder's communication range to receive the packet. DBR and HydroCast implement a linear function of a receiver node, such that the closer the node approaches the sea surface, the less time it waits. The difference between their holding time function is the manner in which the constant factor is determined. Since the node in DBR does not know about the other candidates, the constant factor considers the packet propagation time in the worst scenario possible, in which the highest- and lowest-priority nodes are distant δ m from each other. In HydroCast, this packet propagation time can be computed at each packet transmission, since the next-hop candidate nodes and their priorities are known by the current forwarder node.

OR protocol	Holding time function
VBF [7]	$T_{h_{VBF}} = \sqrt{\alpha} \times T_{delay} + \dfrac{R-d}{v}$
DBR [8]	$T_{h_{DBR}} = \dfrac{2\tau}{\delta} \cdot (R-d)$
HydroCast [5]	$T_{h_{HydroCast}} = (R-d)$

Table 1. Holding time function of some timer-based candidate coordination procedures.

Control-Packet-Based Candidate Coordination

Using a control-packet-based candidate coordination procedure, a candidate node, upon receiving a packet, will respond with a short control packet if high-priority candidates have failed to do so. The control packet transmission notifies the current forwarder node of the successful receipt of the packet, and informs the other low candidate nodes that they should suppress their transmissions. Usually, control-packet-based candidate coordination approaches designed for terrestrial wireless networks fall into one of the categories below [11]:

- ACK-based: The current forwarder node sends the data packet, and the candidates respond with an ACK packet according to their priorities.
- RTS/CTS-based: The current forwarder node sends a request to send (RTS) packet, the candidates respond with a clear to send (CTS) packet according to priority, and finally, the data packet is sent to the candidate that replied to the RTS transmission.

The above-mentioned design principles have severe drawbacks when applied to underwater sensor network scenarios. ACK-based approaches require all candidates close enough to the transmission to be heard by others [11]. Otherwise, duplicated packet transmissions will take place and degrade network performance. RTS/CTS-based candidate coordination might perform poorly in underwater sensor network scenarios, since a high-priority node may successfully receive a short RTS packet, become the next-hop forwarder responding with the CTS packet, and eventually become unable to receive the data packet, as it is longer than the RTS, and its delivery probability is significantly lower.

Currently, to the best of our knowledge, there is no OR protocol designed for underwater sensor networks, which exclusively use control-packet-based candidate coordination. HydroCast [5], which employs timer-based candidate coordination, uses an ACK packet transmission before the candidate forwards the data packet; this functions as an alert reinforcement to inform low-priority nodes, since it is more probable that the neighbor will receive the short ACK packet than the high-priority node data packet transmission.

Design Consideration of OR Protocols for Underwater Sensor Networks

In the previous sections, we have discussed the advantages and disadvantages of candidate set selection and coordination categories. Herein, we highlight the resulting benefits and drawbacks of OR protocol design for UWSNs when both procedures are combined. In describing the combination of these OR protocol design building blocks, we have organized this section according to the candidate set selection procedure.

To further motivate the aforementioned discussion, we plot some performance evaluation results obtained from simulations. In our simulations, we have a two-hop network configuration, in which one underwater node generates data packets to be delivered to surface sonobuoys (sinks), and uses its one-hop neighbors to do so using the OR paradigm. We vary the neighbor densities, which results in a varying number of candidates in the forwarder set. The nodes' transmission power is set to 150 dB re μPa. We use the model described by Urick [12] to simulate the underwater acoustic channel. We simulate a sender-side and receiver-side candidate set selection OR protocol based on the HydroCast [5] and DBR [8] routing protocols, respectively. The results correspond to the average value and a confidence interval of 95 percent.

Figure 4 portrays the results we have obtained. The packet delivery ratio of sender-side and receiver-side candidate set selection approaches, shown in Fig. 4a, increases when the number of neighbors increase. The receiver-side approach performs better than the sender-side approach. The reason for both trends is that more nodes are helping deliver data packets, as corroborated by the results depicted in Fig. 4d. The increasing number of candidates necessary for improving data delivery has undesired consequences. First,

The HydroCast routing protocol uses an ACK packet transmission before the candidate forwards the data packet; this functions as an alert reinforcement to inform low priority nodes, since it is more probable that the neighbor will receive the short ACK packet than the high priority node data packet transmission.

Using sender-side-based candidate set selection and timer-based candidate coordination procedures, OR routing protocols can be robust, efficiently addressing specific aspects of acoustic channel impairment. However, high end-to-end delay can be experienced.

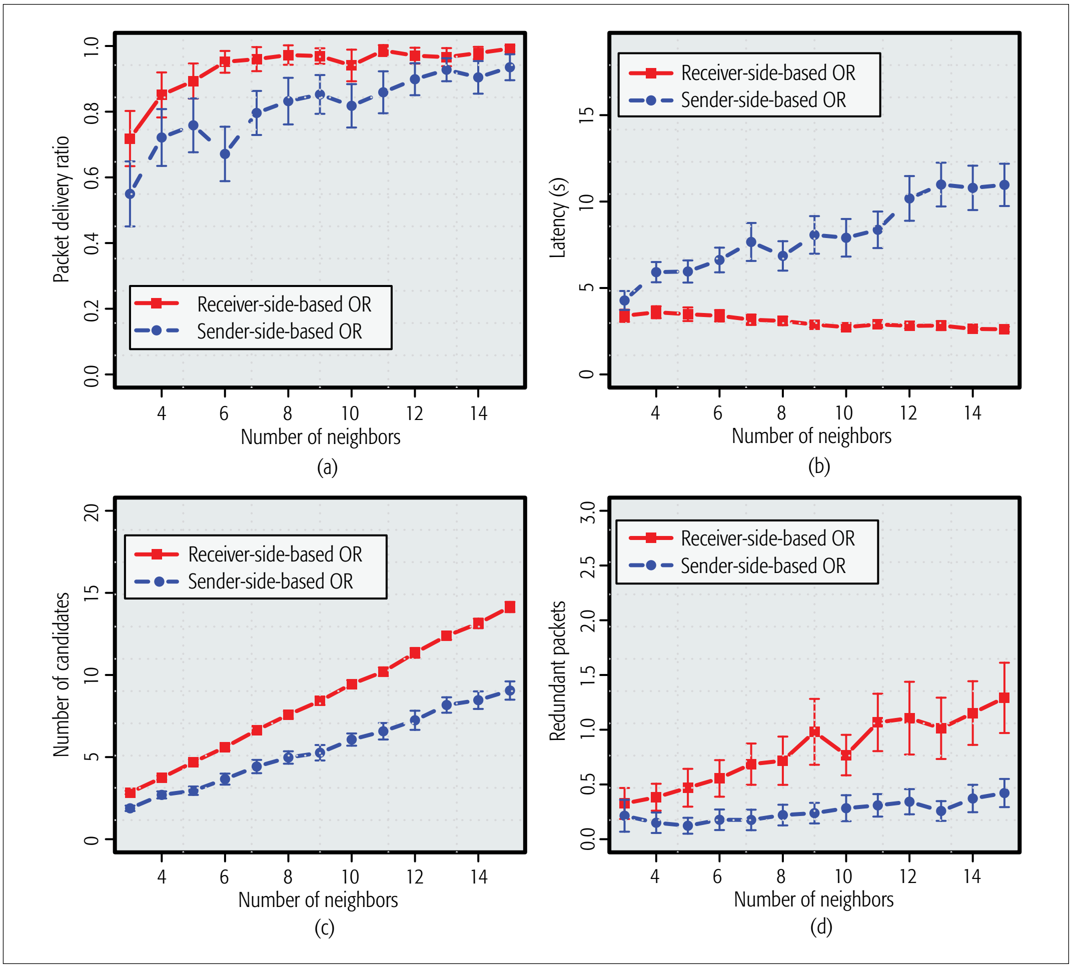


Figure 4. Results: a) packet delivery ratio; b) average end-to-end delay; c) average number of redundant packets; d) average number of candidates.

the latency increases because of the waiting time inherent in timer-based candidate coordination procedures, as shown in Fig.4b. Second, more redundant transmissions will take place because of unreliability of the underwater acoustic channel, which will prevent some low-priority candidates from hearing high-priority candidates' transmissions. These observed effects are discussed in more detail in the following sections.

OR Protocol Using Sender-Side Candidate Set Selection

Using sender-side-based candidate set selection and timer-based candidate coordination procedures, OR routing protocols can be robust, efficiently addressing specific aspects of acoustic channel impairment. However, high end-to-end delay can be experienced. This is due to the fact that in timer-based candidate coordination, low-priority nodes should hold the packet for a time according to its priority. Moreover, given the overhead, data packets should compete with beacon packets to access the underwater acoustic channel.

If control-packet-based candidate coordination is chosen, end-to-end delay can be reduced, since a low-priority node should wait as little time as possible to receive an ACK packet from high-priority nodes. However, the network performance might be severely degraded when OR protocols based on sender-side candidate set selection and control-packet-based candidate coordination are used in high mobility and traffic load application scenarios. This is due to the high energy cost relative to the communication operation and the low probability packet delivery characteristics of the underwater acoustic channel, combined with the high overhead as a result of beacons and control packets.

OR Protocol Using Receiver-Side Candidate Set Selection

Using receiver-side-based candidate set selection and timer-based candidate coordination procedures, OR protocols tend to be simple and scalable, proving very attractive for high mobility and dense underwater sensor network scenarios. Moreover, despite the intrinsic delay of timer-based coordination procedures, their end-to-end delays have a tendency to be low when compared to joint sender-side and timer-based approaches. This is due to the simple receive and forward mechanism, since there is no need to perform complex computations for candidate set selection. However, a high fraction of duplicated packets can occur because of the fully local candidate coordination combined with the hidden terminal problem and the impairments of acoustic communication, which increase packet collisions and spend unnecessary energy.

The aforementioned drawback can be reduced by using control packet-based candidate coordination, given that the use of a short control packet to direct a low-priority node to suspend its transmissions is more likely to be received than data packets. It is worth mentioning that the duplicated packet transmission effect, which is the main drawback of OR protocols, might be used as an advantage for overcoming the difficulties inherent in underwater acoustic communication. For instance, the Energy-Efficient Depth-Based Routing (EEDBR) protocol [13] reduces the suppression of the packet transmission when the packet delivery ratio is less than a particular threshold.

OR Protocol Using Hybrid Candidate Set Selection

We envision that OR protocols for UWSNs, based on hybrid-based candidate set selection with timer-based or control-packet-based candidate coordination, are very suitable for data routing in low packet generation rate applications and moderate- to high-mobility network scenarios. The main advantage of these combinations is that robust OR protocols can be designed with low cost in packet overhead. The choice between timer-based or control-packet-based coordination in this case can be guided by the candidate fitness function. For instance, when too many nodes are selected to the set, timer-based coordination is preferable. The main disadvantage of these OR protocols is the high end-to-end delay.

Future Research Issues

While significant research has been conducted regarding opportunistic routing design for underwater networks, there are several directions that require further exploration. The following are some open research problems that warrant additional investigation.

- There is a lack of modeling works that investigate OR for UWSNs. Theoretical limits of OR performance in UWSNs, which consider optimal frequency selection, data rate transmission adaptation, transmission scheduling, and interference cancellation techniques, are still open research issues. Further research in this direction is of great interest due to the challenging nature of the underwater acoustic communication environment.
- In general, there is a particular number of candidates in the forwarding set, such that the addition of more nodes will not improve routing performance. On the contrary, an excessive number of candidates will increase end-to-end delay and redundant packet transmissions. This characterization is a challenging open research issue in underwater networks, because the number of candidates may not be fixed or the same for all nodes due to the multipath and time-varying underwater acoustic signal propagation and shadow zones.
- Opportunistic routing may shorten the underwater network lifetime due to redundant transmissions. In this context, duty cycling holds promise for reducing energy consumption. Coutinho *et al.* [14] proposed a modeling framework for evaluating the collision of duty cycling and OR in underwater networks. However, an ideal sleep and awake interval selection and their adjustment on the fly are still open research issues.
- Usually, candidate priorities will be unchangeable during the network's operation. This can result in network partitions, as high-priority nodes may deplete their batteries sooner. Therefore, rotation of candidates' priorities should be investigated in order to prolong the network lifetime where the link reliability is increased.

Usually, candidate priorities will be unchangeable during the network's operation. This can result in network partitions, as high-priority nodes may deplete their batteries sooner. Therefore, rotation of candidates' priorities should be investigated in order to prolong the network lifetime where the link reliability is increased.

Final Remarks

This article has presented detailed design guidelines for opportunistic routing protocols in underwater networks. We have described a general framework of OR for UWSNs in particular, and have investigated the design of their candidate set selection and coordination procedures. We have divided candidate set selection procedures into sender-side-based, receiver-side-based, and hybrid. We have categorized the candidate coordination procedures into timer-based and control-packet-based approaches. For each category of candidate set selection and coordination procedures, we have discussed its principles, advantages, and disadvantages, relating them to the underwater acoustic communication characteristics, since these characteristics (together with the network scenario and application) guide the design of opportunistic routing protocols. Finally, we have put into perspective the advantages and disadvantages of the candidate set selection and coordination categories combined during the design of OR protocols for underwater sensor networks.

References

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater Acoustic Sensor Networks: Research Challenges," *Ad Hoc Networks*, vol. 3, May 2005, pp. 257–79.

[2] M. Stojanovic and J. Preisig, "Underwater Acoustic Communication Channels: Propagation Models and Statistical Characterization," *IEEE Commun. Mag.*, vol. 47, Jan. 2009, pp. 84–89.

[3] L. F. M. Vieira, "Performance and Trade-Offs of Opportunistic Routing in Underwater Networks," *Proc. IEEE Wireless Commun. and Networking Conf.*, Apr. 2012, pp. 2911–15.

[4] Y. Li, W. Chen, and Z.-L. Zhang, "Optimal Forwarder List Selection in Opportunistic Routing," *Proc. IEEE 6th Int'l. Conf. on Mobile Adhoc and Sensor Systems*, Oct. 2009, pp. 670–75.

[5] U. Lee *et al.*, "Pressure Routing for Underwater Sensor Networks," *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.

[6] R. W. L. Coutinho *et al.*, "GEDAR: Geographic and Opportunistic Routing Protocol with Depth Adjustment for Mobile Underwater Sensor Networks," *Proc. IEEE ICC*, June 2014, pp. 251–56.

[7] P. Xie, J.-H. Cui, and L. Lao, "VBF: Vector-Based Forwarding Protocol for Underwater Sensor Networks," *Proc. 5th Int'l. IFIP-TC6 Networking Conf.*, May 2006, pp. 1216–21.

[8] H. Yan, Z. Shi, and J.-H. Cui, "DBR: Depth-Based Routing for Underwater Sensor Networks," *Proc. 7th Int'l. IFIP-TC6 Networking Conf.*, May 2008, pp. 72–86.

[9] J. M. Jornet, M. Stojanovic, and M. Zorzi, "Focused Beam Routing Protocol for Underwater Acoustic Networks," *Proc. 3th ACM Int'l. Wksp. Underwater Net.*, Sept. 2008, pp. 75–82.

[10] S. Basagni *et al.*, "Channel-Aware Routing for Underwater Wireless Networks," *Proc. OCEANS - Yeosu*, May 2012, pp. 1–9.

[11] A. Boukerche and A. Darehshoorzadeh, "Opportunistic Routing in Wireless Networks: Models, Algorithms, and Classifications," *ACM Comp. Surveys*, vol. 47, no. 2, Nov. 2014, pp. 22:1–22:36.

[12] J. R. Urick, *Principles of Underwater Sound*, McGraw-Hill, 1983.

[13] A. Wahid *et al.*, "EEDBR: Energy-Efficient Depth-Based Routing Protocol for Underwater Wireless Sensor Networks," *Advanced Comp. Sci. and Info. Tech.*, vol. 195, Sept. 2011, pp. 223–34.

[14] R. W. L. Coutinho *et al.*, "Modeling and Analysis of Opportunistic Routing in Low Duty-Cycle Underwater Sensor Networks," *Proc. 18th ACM Int'l. Conf. Modeling, Analysis and Simulation of Wireless and Mobile Sys.*, Nov. 2015, pp. 125–32.

Biographies

Rodolfo Wanderson Lima Coutinho (rwlc@dcc.ufmg.br, rlimaco2@site.uottawa.ca) is currently a Ph.D. candidate at the Federal University of Minas Gerais (UFMG), Brazil, and the University of Ottawa, Canada. He received his Bachelor's degree in 2009 and Master's degree in 2010, both from the Federal University of Para (UFPA), Brazil. Currently, he is a Ph.D. visiting scholar at PARADISE Research Lab, University of Ottawa, Canada, where is conducting research in the area of underwater sensor networks, wireless networking, and mobile computing.

Azzedine Boukerche [F] (boukerch@site.uottawa.ca) is a full professor and holds a Canada Research Chair Tier-1 position at the University of Ottawa. He is founding director of the PARADISE Research Laboratory and the DIVA Strategic Research Centre at the University of Ottawa. He has received the C. Gotlieb Computer Medal Award, Ontario Distinguished Researcher Award, Premier of Ontario Research Excellence Award, G. S. Glinski Award for Excellence in Research, IEEE Computer Society Golden Core Award, IEEE CS-Meritorious Award, IEEE TCPP Leaderships Award, IEEE ComSoc ASHN Leaderships and Contribution Award, and University of Ottawa Award for Excellence in Research. He serves as an Associate Editor for several IEEE transactions and ACM journals, and is also a Steering Committee Chair for several IEEE and ACM international conferences. His current research interests include wireless ad hoc and sensor networks, wireless networking and mobile computing, wireless multimedia, QoS service provisioning, performance evaluation and modeling of large-scale distributed and mobile systems, and large-scale distributed and parallel discrete event simulation. He has published extensively in these areas and received several best research paper awards for his work. He is a Fellow of the Engineering Institute of Canada, a Fellow of the Canadian Academy of Engineering, and a Fellow of the American Association for the Advancement of Science.

Luiz Filipe Menezes Vieira (lfvieira@dcc.ufmg.br) is an assistant professor at UFMG. He received his degree in computer science from UFMG and his Ph.D. degree from the University of California at Los Angeles. His current research interests include network coding, wireless networks, and underwater sensor networks.

Antonio Alfredo Ferreira Loureiro (loureiro@dcc.ufmg.br) is a full professor at UFMG, where he leads the research group on mobile ad hoc networks. He received his B.Sc. and M.Sc. degrees in computer science from UFMG, and his Ph.D. degree in computer science from the University of British Columbia, Canada. He was the recipient of the 2015 IEEE Ad Hoc and Sensor (AHSN) Technical Achievement Award. He is a regular visiting professor and researcher at the PARADISE Research Laboratory at the University of Ottawa and is an international research partner of DIVA Strategic Research Networks. His main research areas include wireless sensor networks, mobile computing, and distributed algorithms. In the last 10 years, he has published regularly in international conferences and journals related to those areas, and has also presented tutorials at international conferences.

A Journey toward Modeling and Resolving Doppler in Underwater Acoustic Communications

Fengzhong Qu, Zhenduo Wang, Liuqing Yang, and Zhihui Wu

ABSTRACT

Underwater acoustic (UWA) communications is the only reliable method for long-distance communications underwater, and is widely used in commercial, scientific, and military scenarios. However, the UWA channel is most challenging due to its double dispersion property in both long time delay and large Doppler spread, resulting in severe multipath spread and time variation. Among these, Doppler spread is one of the most critical challenges, and researchers have proposed various models in order to resolve Doppler spread in UWA channels. In this article, an overview of Doppler modeling and resolving is provided, divided into four stages: the quasi-static model of the mid-1980s, the uniform Doppler shift model in the 1990s, the basis expansion model and uniform path speed models of the late 1990s, and the recently developed non-uniform path speed model. Furthermore, the UWA channel sparsity property utilized by each of those models will also be discussed.

INTRODUCTION

Underwater acoustic (UWA) communications are widely believed to be the only approach feasible for long-distance communications underwater, and are widely used in various scenarios. The need for high-quality underwater wireless communications arises in many military, scientific, and civilian applications, including communications among submarines, underwater security surveillance, scientific data collection at ocean bottom stations, off-shore oil explorations by autonomous underwater vehicles (AUVs), and data exchanges in underwater sensor networks for environmental monitoring. The 20th century witnessed the evolution of UWA communications from analog noncoherent techniques to digital coherent ones. However, as shown in Fig. 1, UWA communications is uniquely challenging due to the underwater environment and UWA propagation properties. A specific presentation of Fig. 1 is provided in the following three paragraphs.

On one hand, due to the low propagation speed at 1500 m/s of acoustic waves underwater, a relatively large number of arrivals could be distinguished at the receiver side in a shallow water scenario, which usually lasts for tens of milliseconds. This large multipath spread results in long time delay characteristic of the UWA channel and severe inter-symbol interference (ISI) when information from different arrivals collapse into each other, leaving channel equalization a demanding task for reliable UWA communications. Thus, multipath spread should be taken into account when modeling the UWA channel. In addition, the UWA channel also features significant Doppler effects due to platform and sea surface motion, causing large Doppler spread. The large Doppler spread may contaminate the communications, so it needs to be carefully mitigated. The reason the Doppler effect is difficult to solve lies in the fact that it could have different values for different frequencies throughout the bandwidth. Due to its nature, the Doppler effect is widely considered to be more difficult to accommodate than the multipath effect. Unfortunately, UWA channels have severe time variance, where the channel property may differ as time passes by. This explains why the UWA channel time-varying property is another important consideration in modeling. In general, due to long time delay and large Doppler spread, UWA channels are often characterized as doubly spread, especially in shallow water, while terrestrial RF channels are not unless the transceiver mobility is very high.

On the other hand, acoustic waves usually propagate at low frequencies, typically tens of kilohertz, compared to electromagnetic waves that propagate at gigahertz. The fractional bandwidth for a UWA communication system, defined as the ratio between bandwidth against carrier frequency, is usually 1000 times greater than that of terrestrial wireless systems. Thus, a UWA communication system is a typical wideband system. Due to the low propagation speed of acoustic wave at 1500 m/s in water compared to electromagnetic waves propagating at 3×10^8 m/s in air, Doppler spread for UWA communications will have much greater relative values than terrestrial RF communications. Furthermore, UWA communication systems have different Doppler spread values for different frequencies due to their wideband property, unlike narrowband systems where Doppler spread could be considered constant over the entire bandwidth.

The above discussions indicate that more severe challenges are present in UWA channels

The authors provide an overview of Doppler modeling, divided into four stages: the quasi-static model of the mid-1980s; the uniform Doppler shift model in the 1990s; the basis expansion model and uniform path speed models of the late 1990s; and the recently developed non-uniform path speed model.

Fengzhong Qu, Zhenduo Wang, and Zhihui Wu are with Zhejiang University; Liuqing Yang is with Colorado State University.

0163-6804/16/$25.00 © 2016 IEEE

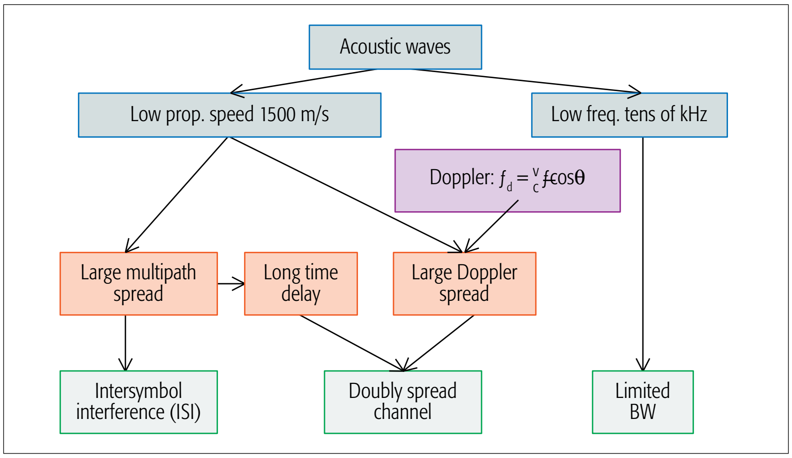


Figure 1. a) Uniform Doppler shift model; b) frequency compression.

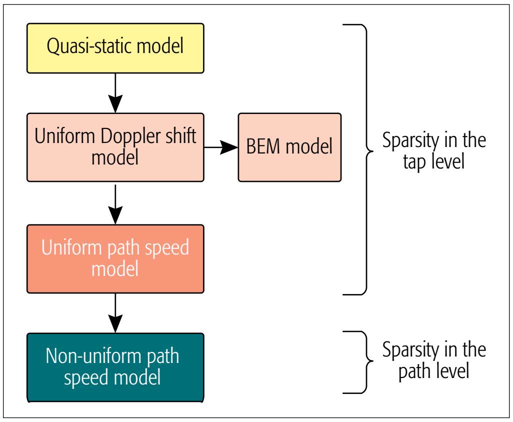

Quasi-static model
Uniform Doppler shift model
BEM model
Uniform path speed model
Non-uniform path speed model
Sparsity in the tap level
Sparsity in the path level

Figure 2. Doppler spread modeling in five stages.

than those in terrestrial RF channels. Several topics of interest have been focused on in UWA communications research in recent years, such as Doppler modeling and resolution, channel equalization for ISI cancellation, UWA channel doubly-spread mitigation in both time delay and Doppler scale, and bandwidth efficiency improvement. Among all these research interests, we believe that Doppler modeling and resolution is one of the most crucial aspects in present UWA communications, and also partly supports the physical layer research in UWA communications at the same time.

In this article, an overview of Doppler modeling and resolution is provided for ever better understanding of the nature of UWA communication channels, and through its development the progress of UWA communications research is also demonstrated. Unlike the distinguished review of UWA communications [1], where the major discussions about UWA communications are channel modeling, equalization, time reversal and passive phase conjugation, multicarrier system, and so on, this article mainly focuses on Doppler modeling and resolution, and tries to reflect the development of UWA communications via this approach. Throughout this article, the Doppler spread concept means that the exact Doppler value varies for different frequencies throughout the bandwidth, while for the term Doppler shift, which is a simplified version of Doppler spread, only a carrier frequency offset (CFO) is considered.

The modeling of Doppler spread can be divided into five stages in a chronological manner, which is shown in Fig. 2: quasi-static model, uniform Doppler shift model, basis expansion model (BEM), uniform path speed model, and non-uniform path speed model. In the mid-1980s, the quasi-static model considered UWA channels as time-invariant in a certain time interval, incapable of tracking channel time variation. The problem solving approach during that time was vague. Later, in the 1990s, a constant CFO (i.e., a uniform Doppler shift value throughout the bandwidth) was introduced into the uniform Doppler shift model, which can be compensated by frequency shifting at the receiver side. In the late 1990s, researchers started to use the BEM, which comprises a series of basis functions to fit UWA channel time variation, purely from the mathematical point of view and skipping its physical propagation properties. Around the same time as the BEM arose, the uniform path speed model started to be used in UWA communications, where a constant path speed is assumed. This model leads to the same Doppler expansion (or compression) ratio for various frequencies, and this expansion (or compression) could be compensated by resampling the received signal in the time domain. The most recent model, the non-uniform path speed model, which arose in the last few years of the 2000s, involves different path speeds for various paths. One possible approach to achieving the non-uniform path speed model is to parameterize both the amplitude and the time delay in a path-wise manner. Therefore, precise Doppler spread modeling, estimation, and compensation could be performed with high precision. UWA channel sparsity is also an important issue to be addressed, where the quasi-static model, uniform Doppler shift model, BEM, and uniform path speed model exploit the sparsity in the tap level, while the non-uniform path speed model utilizes the sparsity in the path level. The usage of sparsity is discussed later.

This article is organized as follows. The quasi-static and uniform Doppler shift models are discussed, respectively. BEM is given. We focus on the uniform and non-uniform path speed models, respectively. We give insight on UWA channel sparsity. Finally, we provide a summary and possible future prospectives.

Quasi-Static Model

In the early stage, around the mid-1980s, UWA communication systems were achieved with limited data rates, which were usually under 1000 b/s. Knowledge on the UWA channel itself was quite limited. The UWA channel was modeled as quasi-static, that is, time-invariant within a certain time interval. This approach was not able to track the variation of the UWA channel; a long channel coherence time is required to achieve reliable communications. Doppler spread is treated at the expense of lowering the data rate and bandwidth efficiency. In addition, the time reversal (TR) signal processing technique for UWA communications was also developed based on the quasi-static channel assumption, utilizing the principle of the TR mirror in the 1990s. TR signal processing is in fact a channel equalization technique utilizing the auto-correlation property of channel impulse response (CIR), where a Dirac-shaped impulse response will be generated at the receiver side, further enabling the matched filter processing [2]. From another point of view, TR signal processing for UWA communications could be considered as a time-space matched filter for CIR by taking advantage of the ocean itself.

Uniform Doppler Shift Model

Later on, in the 1990s, the quasi-static model developed into the uniform Doppler shift model, where a fixed Doppler shift Δf_d, that is, a carrier frequency offset (CFO) as indicated in Fig 3a, was assumed throughout the whole bandwidth. In the later processing stage of received signals, this offset would easily be removed by shifting the frequency band opposite to that CFO. For example, as in [3], a frequency-shift estimator combined with a time-scale interpolator is utilized to estimate the Doppler shift, remove the offset, and then interpolate to shift the timescale of the data. The uniform Doppler shift is estimated from the training data at the start of each received packet and is computed across the range of Doppler shifts to maximize the ambiguity function, which is a two-dimensional function of the time delay and the Doppler frequency. As the previous step removes coarse Doppler shift, further operation is dedicated to deal with the residual Doppler shift; still, the performance might not be perfect.

In multiple-carrier UWA communication systems, such as an orthogonal frequency-division multiplexing (OFDM) system, which is very sensitive to the Doppler spread since it destroys the orthogonality among subcarriers, a general CFO was often modeled and estimated for all subcarriers by means of the null-subcarriers-based approach. Due to the Doppler spread, there would be energy leakages to null subcarriers, and by minimizing the energy on those null subcarriers, a CFO estimate could be obtained [4]. However, some necessary and sufficient conditions on the number of null subcarriers and their placement should be considered cautiously, since they may relate to issues like CFO identifiability. In summary, as the uniform Doppler shift model considers static Doppler shift in a certain time interval throughout the whole frequency band, it could also be regarded as a generalized quasi-static model.

Basis Expansion Model

Apart from treating the UWA channel as a quasi-static or uniform Doppler shift model, the focus of BEM, developed in late 1990s, is to model the UWA channel's time variation purely from a mathematical point of view. Taking the assumption that Doppler spread for a certain time-variant UWA channel is limited and under some maximum value, a series of basis functions could be used to fit its time variance, and those basis functions actually span an orthogonal signal space. The basic idea for BEM is that it truncates CIR in the time domain while the remaining channel taps are negligible, and then selects a series of basis functions and estimates the correspond coefficients to model the UWA channel.

The advantage of BEM is that it reduces the degree of freedom of the UWA channel since it uses limited basis functions with their corresponding coefficients to fit the UWA channel. For example, in order to get the exact CIR, an N-points sample is required. But with the BEM method, when zero values are set for those channel taps with negligible magnitudes, typically K basis functions are assigned to model the CIR, where K is smaller than N. It could be concluded from this observation that the BEM approach exploits the sparsity in the time domain for channel taps. However, the drawback of BEM is that the construction of orthogonal space is essentially performing a truncation in the time domain. This will lead to model error and frequency leakage in the high frequency band, deteriorating the estimation accuracy [5].

A brief introduction to various BEM models is provided in [6]. The discrete Fourier transform (DFT) BEM model is a common BEM model that utilizes low-frequency components in an inverse DFT matrix as basis functions, and channel fitting may be performed based on least square (LS) or minimum mean square error (MMSE) criterion. However, DFT BEM actually uses a window function to truncate CIR in the time domain, and this induces non-existent high-frequency components, resulting in modeling bias. Other BEMs include discrete prolate spherical sequences (DPSS) BEM and Karhunen–Loève (KL) BEM. DPSS BEM uses the set of eigenvectors corresponding to the largest eigenvalues of the band-limited rectangular power spectrum signal (i.e., discrete prolate spherical sequence) as the basis functions for UWA channel fitting. Since DPSS BEM has a double orthogonality property over an infinite and a finite time interval, it takes fewer approximations compared to DFT BEM. In addition,

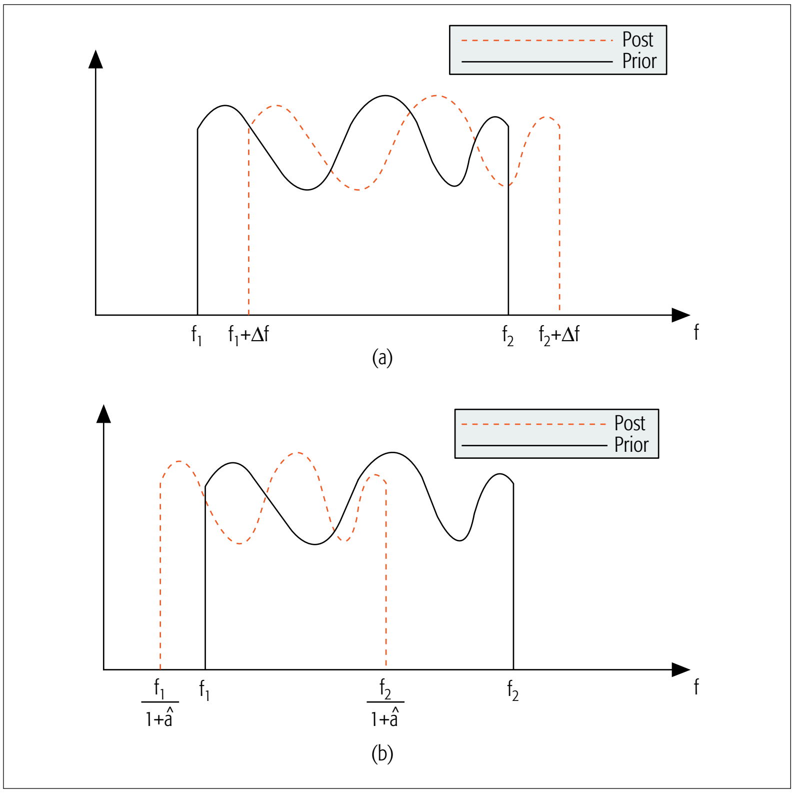


Figure 3. a) Uniform Doppler shift model; b) frequency compression.

channel estimation could be performed with higher resolution by DPSS BEM since the matrix formulated by its basis function matrix is a uniform matrix the inverse of which could be calculated easily. The disadvantage of the DPSS BEM approach is that it influences the channel fitting bias distribution since it colors the noise, being worse than the DFT BEM approach. The tradeoff of DPSS BEM should relate to when to use it. KL BEM uses the eigenvectors corresponding to the largest eigenvalues of the covariance matrix of the channel as basis functions. However, a channel covariance matrix is usually not available when applying KL BEM. This yields limited usage for the KL BEM approach.

By catching the channel variation, BEM transfers the time-variant channel into its time-invariant equivalent, usually within a block time, further enabling two possible applications. One is to perform channel estimation based on BEM, since BEM reduces the number of coefficients to be estimated. DFT BEM is preferred, but the frequency leakage problem should also be considered. A possible solution is to design the window function with fewer sidelobes [5]. After the block-wise time-invariant equivalent channel model is obtained, another application is to apply differential coding at both the transmitter and receiver. By differential coding, the previously received symbols could be used as the channel reference for the current symbols, and channel estimation could be skipped.

Uniform Path Speed Model

The aforementioned model merely uses an average Doppler shift in the frequency domain to model the Doppler spread that UWA communications systems actually suffer from, and the BEM approach does not take advantage of the physical propagation property. Since the UWA communications system is a typical wideband system, where Doppler shifts each frequency component by a different amount, a more general Doppler spread should be considered. Therefore, researchers transformed the uniform Doppler shift model to the uniform path speed model for better performance.

In the uniform path speed model, for the dominant arriving paths, a constant relative speed between transmitter and receiver platform and a certain constant acoustic speed are assumed. And a complete time scaling, that is, the same ratio, throughout the frequency band is utilized to describe the Doppler expansion (or compression) in the frequency domain, which is shown in Fig 3b. In other words, the absolute Doppler spread values may differ from different frequencies, but these values share the same scaling ratio.

Therefore, a more realistic Doppler compensation strategy is to compress or extend the received signal in the frequency domain according to that ratio, which corresponds to upsampling or downsampling in the time domain. The advantage of resampling is that it could compensate for varying Doppler spread for different frequencies, which was widely used later on. As the uniform path speed model introduces a constant relative speed, the acceleration of transmitter or receiver platform is neglected. In fact, the constant relative speed is sufficient, because during a certain time block, the speed variation is quite limited, and the acceleration for transmitter or receiver platform, if any, is very little.

As the uniform path speed model introduces a constant relative speed, the acceleration of transmitter or receiver platform is neglected. In fact, the constant relative speed is sufficient, because during a certain time block, the speed variation is quite limited, and the acceleration for transmitter or receiver platform, if any, is very little.

The example of resampling is to use a Doppler-insensitive signal, such as a linear frequency modulation (LFM) signal, placed before and after the transmitted data block, to calculate the ratio between the duration time at transmitter side T_{tx} and that at receiver side T_{rx}. By dividing T_{tx} by T_{rx}, the resample ratio $\hat{\alpha}$ can be obtained. Later on, the resampling method was extensively used in both single-carrier and multicarrier UWA communications systems, and hence became a dominant approach for average Doppler spread mitigation.

In a single-carrier UWA communications system, resampling is usually performed prior to symbol demodulation. In order to achieve high accuracy, sample rate conversion should be performed cautiously. This demanding task is usually accomplished by converting the sampling rate by a rational number [7] by which the received signal is interpolated in the first place, continued with a finite impulse receiver (FIR) filter, and finally decimated. Computational load should be improved by using linear interpolation to calculate each new sample.

Resampling has a similar application for average Doppler removal in multi-carrier UWA communications system. To achieve better perform- ance, some fine Doppler spread compensation strategy is required for more precise compensation. As shown in [8], the received signal for an OFDM communication system is resampled to remove average Doppler spread prior to being transferred to the baseband for fine compensation, converting the wideband problem into a narrowband problem. And the high-resolution uniform residual Doppler compensation is performed corresponding to the narrowband model for the best inter-carrier interference (ICI) reduction.

The application scenarios for uniform path speed model and resampling Doppler spread compensation are somehow limited, since it would lose effectiveness in shallow water scenarios when paths with excessive reflections exist, as shown in Fig. 4. For some paths (e.g., direct, surface-reflected, and bottom-reflected paths), the uniform path speed model may work well, since the difference between various arriving angles is little, yielding similar Doppler spreads. But for those with excessive reflections, a constant path speed is not adequate, and those paths need to be involved for reliable system performance. Therefore, a precise path-based Doppler modeling approach is required in addition to the uniform path speed model.

Non-Uniform Path Speed Model

The aforementioned methods (e.g., the uniform Doppler shift or uniform path speed model) use either a constant uniform Doppler shift or a constant speed to model the Doppler shift, and the BEM approach skips the channel propagation property. All three models are tap-based and unable to fully utilize the UWA channel propagation properties (the multipath effect, etc.). Thus, a new path-based approach for UWA channel modeling is required, emphasizing the channel physical propagation properties.

As shown in Fig. 4, the multipath effect of

UWA channels could be demonstrated as the effect of the direct path, the surface-reflected path, the bound-reflected path, the multi-reflected paths, and so on. The differences between these paths lay in the path variant amplitudes, time delays, and speeds, that is, Doppler spreads. The Doppler shift for the direct acoustic ray is $D_1 := fv/c$, and the Doppler shift for the acoustic ray with largest arriving angle θ is $D_2 : = fv/c \cdot \cos\theta$, so the Doppler spread is defined as $D_s := D_1 - D_2$, where v stands for the relative speed between the transmitter and receiver platforms, c is the sound speed, and f is the carrier frequency. Some insights could be obtained from the above expression when modeling Doppler spread for a specific path with path speed v, which provides the possibility of realizing the non-uniform path speed model. The UWA channels scattering function in the RACE08 sea trial is shown in Fig 5. RACE 08 was performed at Narragansett Bay near the University of Rhode Island with the help of WHOI. The water depth was 9 to 14 m, and the distance was 1000 m. Twelve hydrophones with 0.12 m spacing and three transducers were used in the experiment. Figure 5 indicates different Doppler spreads on different paths, where in good channel conditions the Doppler varies from approximately –0.2 Hz to 0.2 Hz, and it varies from approximately –1.5 Hz to 0.5 Hz in bad channel conditions. Furthermore, the sparsity property of UWA channels could also be exploited, together with compressive sensing. This is discussed in the next section.

In UWA related applications, some modeling approaches referring to physical propagation effects have been investigated (e.g., the ray-based acoustic scattering approach). In addition, the path-based UWA channel model is also investigated, in which the Doppler spread is modeled at the path level. The advantage of the path-based UWA channel model is that the Doppler spread could be treated as a mean Doppler spread that is the same for all the paths plus an individual residual Doppler spread for each path. The mean Doppler spread could be compensated by easily resampling to reduce computational complexity. And for the residual Doppler spread for each path, some advanced modeling and problem-solving approaches could be applied for better performance.

A prevalent method is to use polynomials to fit amplitude variation and time delay for different arrivals in UWA channels, first applied in terrestrial wireless communication systems and then transferred into UWA communications systems. Three coefficients are required to model each dominant ray: amplitude A, time-delay τ_p, and Doppler spread β. Specifically, as shown in [9], a polynomial up to N order is used to model the channel amplitude variation, and another one- or two-order polynomial is used to model the time delay for UWA channels. Usually, one-order polynomial is able to represent the constant speed of the transmitter and receiver platforms (i.e., the Doppler spread), while two-order polynomial is capable of modeling the linear acceleration for transmitter and receiver platforms, but additional application complexity is introduced as indicated by the theoretical analysis.

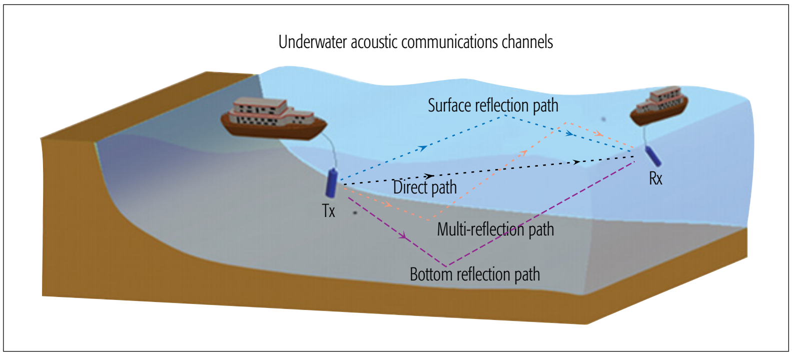


Figure 4. UWA channel multipath property.

In the next step, UWA channel estimation is investigated. One possible approach is to simultaneously search in both the timescale and Doppler spread scale for estimating amplitude variation, time delay, and Doppler spread. This approach would introduce a relatively high computational load and application complexity. The computational complexity and communication overhead are traded off for achieving reliable UWA communications. When more coefficients are involved, the overhead is most likely to increase linearly depending on the transmission scheme, while the computational complexity may increase exponentially when a matrix inversion operation is required.

For further improvement, a two-stage UWA channel estimation method was proposed in 2013 [10], as shown in Fig. 6. In this approach, time delays (squares) are estimated in the timescale for each ray in the first place, and then the delay-Doppler two-dimensional grid is constructed based on the previous estimated time delay (circles). As in the two-stage approach, the estimations of two dimensions are done sequentially, and the total number of candidates for this approach is the sum of the number of candidates in each dimension. However, in the one-stage approach, as the estimation is done simultaneously in both the timescale and Doppler spread scale, the total number of candidates becomes the multiplication of the number of candidates in each dimension. Furthermore, the two-stage approach could achieve higher accuracy with the same amount of pilots or reduce the amount of required pilots to have the same accuracy as the one-stage approach.

In summary, the non-uniform speed path model exploits the propagation effect while modeling UWA channels and enables specific Doppler spread estimation on the ray-level, which are the two major contributions to UWA communications research.

Besides the difference in the approach for modeling the UWA channel for the aforementioned five models, they also vary in application scenarios. For quasi-static and uniform Doppler shift models, they are most applicable in calm sea environments with fewer water dynamics and little relative movement between the transmitter and the receiver, and the symbol duration is considerably less compared to channel coherence time. The uniform path speed model is suitable when distance between the transmitter and the receiver is large so that the reflection angle for acoustic rays is quite little. For BEM, since it is

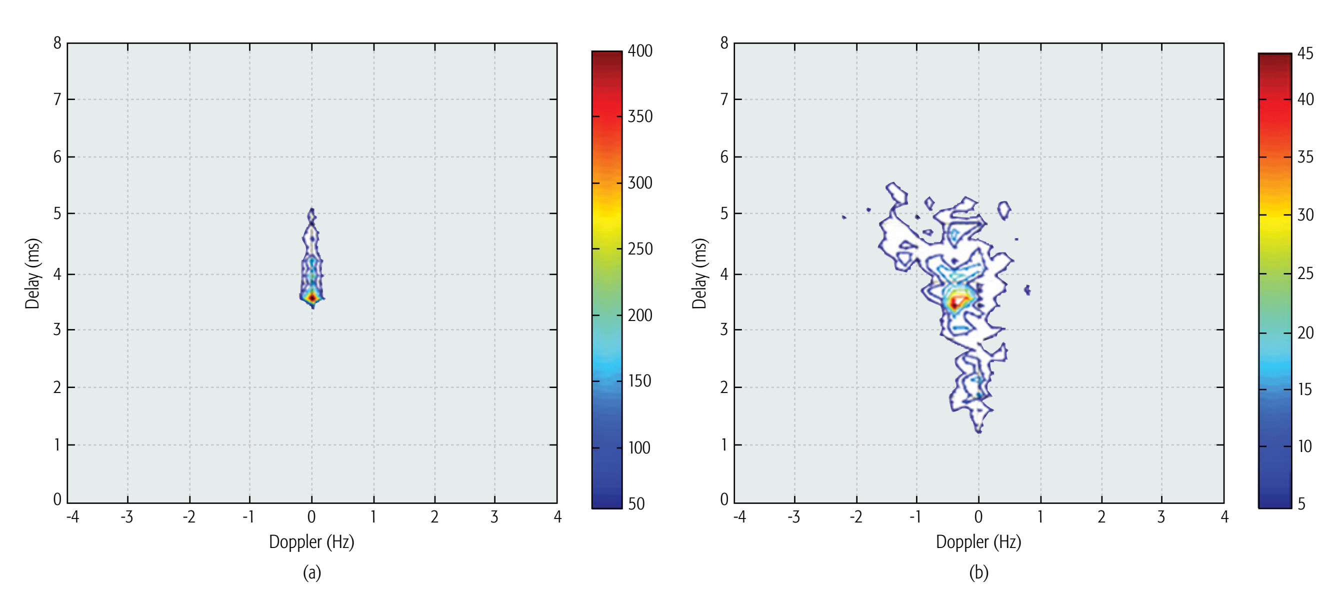


Figure 5. Scattering function for UWA channels in RACE08: a) for good channel conditions; b) for bad channel conditions.

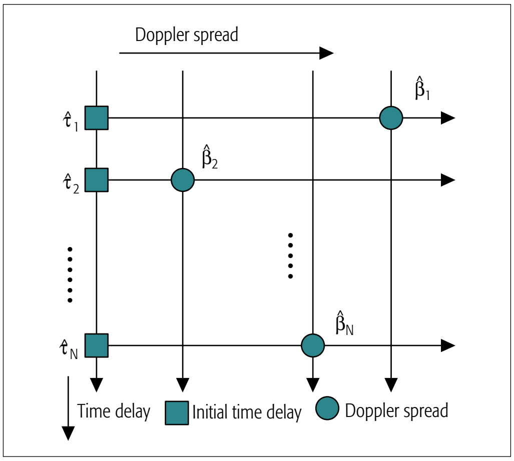


Figure 6. Search pattern for two-stage UWA channel estimation.

an approximation of a UWA channel's variation, it makes sense when the largest Doppler value is accessible. The non-uniform path speed model has the optimal performance in dynamic water circumstances with multiple paths.

UWA Channel Sparsity

From the previous sections, we see that throughout the development of UWA communications, researchers are always trying to obtain a more precise model to describe the Doppler spread, that is, the quasi-static, Doppler shift, uniform path speed, and non-uniform path speed models, and BEM, which uses basis functions to fit UWA channel time variation.

However, as more coefficients are involved in channel modeling, problems arise such as computational complexity and communication overhead. The model accuracy and the above expense seems a trade-off for achieving reliable UWA communications. One possible solution to this problem is to explore UWA channel sparsity and combine it into the demanding model task. The aforementioned models utilize the UWA channel sparsity to different extents, as discussed in the following parts.

The quasi-static and uniform Doppler shift model only utilize a constant uniform Doppler shift, which could be treated as the sparsest but least accurate modeling approach. Also, the uniform path speed model performs the Doppler modeling by means of a constant speed, so specific Doppler spread values for different frequencies are actually required for compensation, which increases its complexity but with better system performance. BEM uses a series of basis functions to fit the time variation of UWA channels, and it truncates the CIR with a certain window function, neglecting the rest of the channel taps. In one word, the above models utilize UWA channel sparsity at the tap level.

When it comes to the non-uniform path speed model, taking advantage of UWA channel sparsity becomes an important problem, since more coefficients are about to be involved, which are in two aspects: channel modeling and estimation.

First, given that there are multiple arriving acoustic rays due to reflections, a question arises naturally: Is it really compulsory to take every path into consideration? Generally, when more paths are involved, a more precise channel model is obtained, and better performance is thereby achieved with less modeling error and lower error rate. For convenience, the multipath effect of a UWA channel could be limited to some degrees about which only the dominant paths need to be concerned, while the rest could be neglected due to excessive reflections or low energy.

In the second stage, sparse channel estimation based on greedy algorithms for UWA communications could also be applied, especially when a UWA channel is modeled based in the path level. These methods include the basis pursuit (BP) and orthogonal matching pursuit (OMP) algorithms. When the UWA channel is modeled based on the non-uniform path speed model where each path is determined by amplitude variation, time delay, and Doppler spread, we could search for those optimal values within a predefined dictionary with the assistance of those greedy algorithms. The intention of those greedy algorithms is to iteratively search for the optimal estimation, while BP and OMP are two applications.

A certain sparse channel estimation is shown in [11] in which the sparsity property of UWA channels is investigated by analyzing channel scattering function, and several greedy-algorithm-based channel estimation methods are also performed. This combined non-uniform path speed model and sparse UWA channel estimation approach is also investigated in [10], where the estimations of time delay and Doppler spread are divided into two stages.

Summary and Prospects

For the past three decades, researchers have been concentrating on achieving a more precise model for Doppler spread, and performing problem-solving for Doppler estimation and compensation with higher accuracy. In this article, Doppler spread modeling and solution is categorized in five stages: the quasi-static model developed around mid-1980s, the uniform Doppler shift model of the 1990s, BEM and the uniform path speed model from the late 1990s, and the non-uniform path speed model developed recently. Throughout the development, we observe that efforts have been made in order to get a more precise model with fewer approximations to reflect the reality of UWA channels. In a chronological manner, the five models discussed in this article provide higher and higher accuracy to reflect the reality of UWA channels step by step, by carefully considering the hardware voltage responses, the computational complexity, and the necessary coefficients that must be involved. Fortunately, some remarkable milestones have been accomplished in this process, and these great contributions improve the reliability of UWA communications dramatically.

Based on the continuous investigation of UWA channel modeling, it is worthwhile for the next stage to look into the physical process that each path experiences (e.g., sediment property for bound reflect, surface wave analysis for surface reflect). This investigation could in turn enable a more precise model and problem solving for Doppler spread in UWA communications.

Acknowledgments

This article is in part supported by the National Natural Science Foundation of China under Grants 61172105, and 61371093.

References

[1] A. B. Baggeroer, "An Overview of Acoustic Communications from 2000–2012," *J. Underwater Commun.: Channel Modelling & Validation*, vol. 5, no. 1, Sept. 2012, pp. 201–07.

[2] T. C. Yang, "Correlation-based Decision-Feedback Equalizer for Underwater Acoustic Communications," *IEEE J. Oceanic Eng.*, vol. 30, no. 4, Oct. 2005, pp. 865–80.

[3] J. Mark, L. Freitag, and M. Stojanovic, "Improved Doppler Tracking and Correction for Underwater Acoustic Communications," *Proc. IEEE Int'.l Conf. Acoustics, Speech, and Signal Processing*, vol. 1, Munich, Germany, Apr. 21–24, 1997, pp. 575–78.

[4] M. Ghogho, A. Swami, and G. B. Giannakis, "Optimized Null-Subcarrier Selection for CFO Estimation in OFDM over Frequency-Selective Fading Channels," *Proc. IEEE GLOBECOM*, vol. 1, San Antonio, TX, Nov. 25–29, 2001, pp. 202–06.

[5] F. Qu and L. Yang, "On the Estimation of Doubly-Selective Fading Channels," *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, Apr. 2010, pp. 1261–65.

[6] Z. Tang *et al.*, "Pilot-Assisted Time-Varying Channel Estimation for OFDM Systems," *IEEE Trans. Signal Processing*, vol. 55, no. 5, May. 2007, pp, 2226–38.

[7] B. S. Sharif *et al.*, "A Computationally Efficient Doppler Compensation System for Underwater Acoustic Communications," *IEEE J. Oceanic Eng.*, vol. 25, no. 1, Jan. 2000, pp. 52–61.

[8] B. Li, S. Zhou *et al.*, "Multicarrier Communication over Underwater Acoustic Channels with Nonuniform Doppler Shifts," *IEEE J. Oceanic Eng.*, vol. 33, no. 2, Apr. 2008, pp. 198–209.

[9] X. Xu *et al.*, "Parameterizing Both Path Amplitude and Delay Variations of Underwater Acoustic Channels for Block Decoding of Orthogonal Frequency Division Multiplexing," *J. Acoustical Soc, America*, vol. 131, no. 6, 2012, pp. 4672–79.

[10] F. Qu, X. Nie, and W. Xu, "A Two-Stage Approach for the Estimation of Doubly Spread Acoustic Channels," *IEEE J. Oceanic Eng.*, vol. pp, Mar. 2014, pp. 1–13.

[11] W. Li and J. C. Preisig, "Estimation of Rapidly Time-Varying Sparse Channels," *IEEE J. Oceanic Eng.*, vol. 32, no. 4, Oct. 2007, pp. 927–39.

Biographies

Fengzhong Qu [S'07, M'10, SM'15] (jimqufz@zju.edu.cn) received his B.S. and M.S. degrees from Zhejiang University, Hangzhou, China, in 2002 and 2005, both in electrical engineering. He received his Ph.D. degree from the Department of Electrical and Computer Engineering at the University of Florida, Gainesville, in 2009. Since 2011, he has been with the Ocean College at Zhejiang University, Hangzhou, China, where he is presently an associate professor. His current research interests include underwater acoustic communications and networking, and sea floor observatories.

Zhenduo Wang (zhenduowang@zju.edu.cn) received his M.Sc. degree in mechatronics, mechanical engineering, from Royal Institute of Technology, Stockholm, Sweden, in 2012, and now is a Ph.D. student in the area of underwater acoustic communications and underwater observatories at the Ocean College, Zhejiang University.

Liuqing Yang (lqyang@engr.colostate.edu) received her Ph.D. degree from the University of Minnesota, Minneapolis, in 2004. She is currently a professor with Colrado State University, Fort Collins. Her main research interests include communications and signal processing. She is the recipient of the ONR YIP Award in 2007, the NSF CAREER Award in 2009, and Best Paper Awards at IEEE ICUWB '06, ICCC '13, ITSC '14, and GLOBECOM '14.

Zhihui Wu (zhihuiwu@zju.edu.cn) received her B.S. degree in automation from Nanjing University of Post and Telecommunication, China, in 2014, and is currently pursuing her M.S. degree in naval architecture and ocean engineering from Zhejiang University. Her research interests include wireless communication and underwater acoustic communication.

Based on the continuous investigation of UWA channel modeling, it is worthwhile for the next stage to look into the physical process that each path experiences. This investigation could in return enable a more precise model and problem-solving for Doppler spread in UWA communications.

Impulse Response Modeling for General Underwater Wireless Optical MIMO Links

Huihui Zhang and Yuhan Dong

In underwater wireless optical communications, channel modeling plays a key role in investigating the propagation properties of light beams through UWOC links as well as evaluating overall system performance. We consider UWOC MIMO systems with multiple light sources and detectors, and focus on the impulse response that is capable of characterizing the temporal behavior of UWOC links.

ABSTRACT

In underwater wireless optical communications (UWOC), channel modeling plays a key role in investigating the propagation properties of light beams through UWOC links as well as evaluating overall system performance. We consider UWOC multiple-input and multiple-output (MIMO) systems with multiple light sources and detectors, and focus on the impulse response that is capable of characterizing the temporal behavior of UWOC links. We propose a weighted Gamma function polynomial (WGFP) to model the impulse response of general UWOC MIMO links with arbitrary numbers of light sources and detectors. Numerical results of Monte Carlo simulations have validated the proposed WGFP model in a turbid water environment.

INTRODUCTION

In the past decade, underwater wireless optical communications (UWOC) has received considerable attention due to the advantages of a much higher data rate, bandwidth, and security over traditional underwater acoustic communications. Although light beams suffer from absorption and scattering and are applicable for relatively short ranges compared with acoustic waves, UWOC is still a promising technology and has many more potential applications such as underwater observation and monitoring, especially for the transmission of large volume data under water [1].

Prior studies have shown that the absorption and scattering processes may introduce energy loss and direction changing of light beams. Among visible spectrum, blue/green light with wavelengths from 450 nm to 580 nm has the lowest absorption in seawater, and is typically adopted for UWOC applications. Light scattering may affect the impulse response, which is able to characterize the temporal behavior of UWOC links. The negligible effect of scattering has been validated on channel impulse response in clear water [2]. However, in a turbid seawater environment, scattering will broaden the impulse response and therefore introduce inter-symbol interference (ISI) and degrade system performance such as channel bandwidth and bit-error-rate (BER). Some research has been done on impulse response modeling for UWOC links by analytical analysis [2, 3] or experimental measurements [4]. Gabriel *et al.* [1] simulated the trajectories of emitted photons propagating in water to study the impulse response and quantified the temporal dispersion for different water types, link distances, and transmitter/receiver characteristics. Tang *et al.* [3] used the Monte Carlo method to reach a closed-form expression of the impulse response for UWOC links under the condition of collimated source and precisely aligned link in turbid seawater. Cochenour *et al.* [4] measured the frequency response of a UWOC channel and investigated the impact of scattering function, receiver field of view (FOV), and pointing angle between the transmitter and receiver on spatial and temporal dispersions.

Multi-input and multi-output (MIMO) technology has been widely used and accepted as an effective approach to improve system performance as measured by throughput and robustness. However, all these prior works (see [2–4] and references therein) only focus on the UWOC single-input and single-output (SISO) links. Compared with traditional UWOC SISO systems, UWOC MIMO systems are promising to provide even higher data rates and/or larger communication ranges. To the best of our knowledge, our prior study [5] is the first to present a simple closed-form expression of weighted double Gamma functions (WDGF) to model the impulse response of 2 × 2 UWOC MIMO systems with two light sources and two detectors. In this article we extend our prior study to more general UWOC MIMO systems with arbitrary numbers of light sources and detectors. We have proposed a closed-form expression of weighted Gamma function polynomial (WGFP) to model the impulse response of the general UWOC MIMO links, which facilitates performance evaluation and system design for general UWOC MIMO systems. Numerical results of Monte Carlo simulations have validated the proposed WGFP model in turbid water environments.

SYSTEM MODEL

BASIC MIMO PRINCIPLE

We consider a general UWOC MIMO link geometry over which N detectors (receivers (Rx)) are illuminated by M light sources (transmitters (Tx)) where M and N are arbitrary integers. As shown in Fig. 1, M transmitters and

The authors are with Tsinghua University. Y. Dong is the corresponding author.

N receivers are wrapped into linear arrays at each side, located in the xoy-plane and the paralleled receiving plane, and centered on the x axis and the line parallel to the x axis, respectively. Denote the intensity of photons emitted by the M-th transmitter as s_m, $m = 1, \cdots, M$. In the absence of noise, the intensity of photons captured by the N-th receiver can be computed as $r_n = \Sigma_{m=1}^{M} h_{nm} * s_m$, where h_{nm} is the impulse response from the M-th transmitter to the N-th receiver, $n = 1, \cdots, N$, and * denotes the convolution operator. Note that each impulse response h_{nm} characterizes the temporal behavior of the UWOC SISO link between the corresponding pair of transmitter and receiver.

Monte Carlo Approach

The Monte Carlo approach can be used for addressing the absorption and scattering processes in underwater environments by generating a large number of photons and simulating their trajectories. Therefore, the channel characteristics can be evaluated by the Monte Carlo approach numerically instead of solving the radiative transfer equation (RTE) [6], which is hard to be achieved analytically.

We apply a similar Monte Carlo approach for UWOC SISO links [2, 3, 5] to this study of UWOC MIMO links. At the side of the transmitter array, each emitted photon will be assigned the following four basic attributes: the position in Cartesian coordinates, transmission direction, propagation time, and weight, also known as intensity. Transmission direction refers to the pair of zenith angle and azimuth angle of a photon in spherical coordinates. At the very beginning, each photon will be uniformly emitted by a Gaussian light source with a small beam radius into a solid angle confined to the divergence angle and elevation angle of the source. Here the elevation angle represents the angle between the principal optic axis of the corresponding source and the normal of the receiving plane, as shown in Fig. 1.

During propagation, a photon will travel a step distance between two successive scatterers, which correspondingly results in an increment of propagation time and change of Cartesian coordinates. After interacting with some particle in water, the photon will suffer weight loss and direction deviation due to absorption and scattering, respectively. Single scattering albedo $b(\lambda)/c(\lambda)$ is the attenuation factor of weight loss where λ is the light wavelength, and $a(\lambda)$, $b(\lambda)$, and $c(\lambda) = a(\lambda) + b(\lambda)$ are coefficients of the absorption, scattering, and extinction, respectively.

The direction deviation is the rotated angle relative to the previous direction and characterized by the scattering phase function (SPF). Typically, SPF also depends on λ and can be approximated by the Henyey-Greenstein function [6], with parameter g defined as the average cosine of the scattering angle in all scattering directions. In practice, the parameter set (a, b, c, g) distinguishes different types of water.

At the receiver side, a photon can be detected by a receiver only when its position and arrival angle are within the receiver aperture and FOV, respectively, and its weight is higher than the threshold. Then the impulse response observed at each receiver can be evaluated by histogramming the weight versus the propagation time of detected photons.

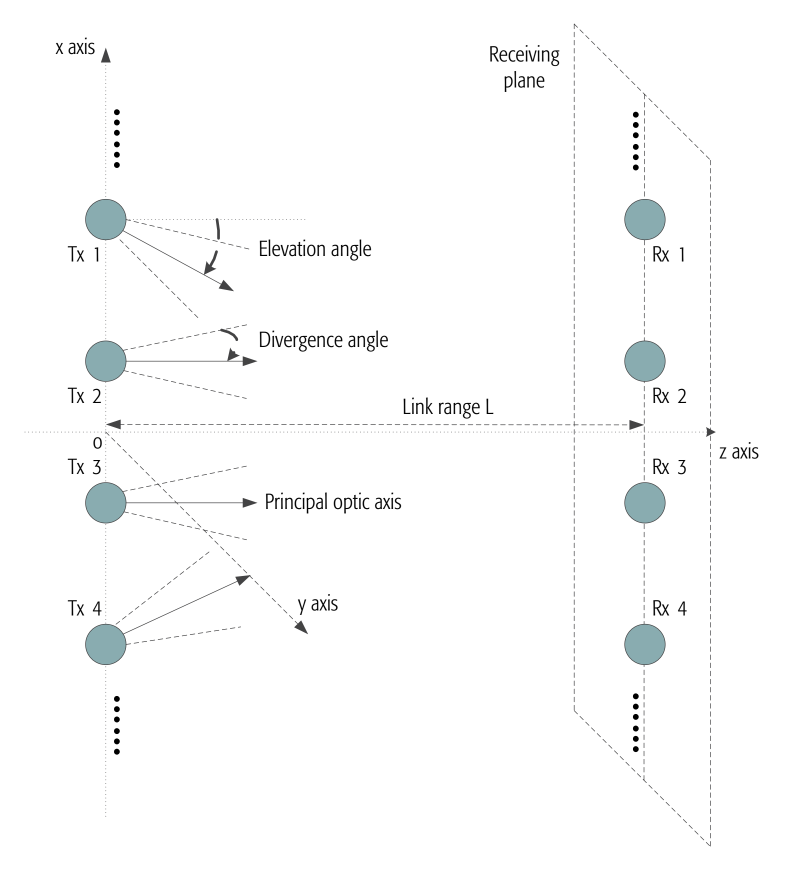


Figure 1. A general UWOC MIMO link geometry.

Weighted Gamma Function Polynomial

In this section we propose a weighted Gamma function polynomial to model the impulse response for general $M \times N$ UWOC MIMO links. Note that the impulse response of an UWOC system varies in different water types and communication distances. We therefore adopt the attenuation length $\tau = c(\lambda)L$ with L as the link range to indicate the joint configuration of water type and communication distance.

For small values of τ, the path loss versus τ follows Beer's law [6], where the non-scattering (absorption) effect dominates. In this case, the impulse response of UWOC SISO links has negligible temporal dispersion and can be modeled by an ideal delta function [2]. However, scattering effect becomes dominant for large attenuation lengths in underwater environments as scattered photons are captured by the detector [8]. In turbid water environments where τ is relatively large, Tang *et al.* [3] modeled the channel impulse response of precisely aligned UWOC SISO links by double Gamma functions (DGF) with two terms representing the relatively low order and high order scattering components, respectively. Moreover, our prior study [5] was the first to present a simple closed-form expression of WDGF to model the impulse response of 2 × 2 UWOC MIMO systems.

For general $M \times N$ UWOC MIMO systems

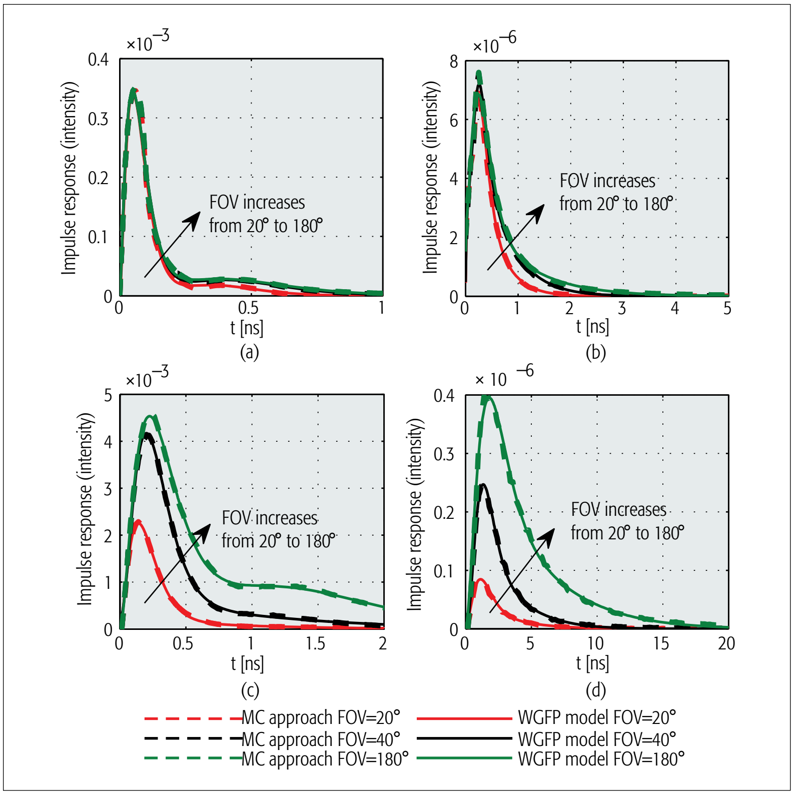


Figure 2. Impulse response of the 2 × 2 MIMO link: a) L = 12 m link in coastal water τ = 3.66; b) L = 30 m link in coastal water τ = 9.15; c) L = 5 m link in harbor water τ = 10.85; and d) L = 12 m link in harbor water τ = 26.04.

as shown in Fig. 1, the photons emitted from different sources will suffer entirely different scattering and absorption processes during propagation and then reach different detectors in the sense of probability. Consequently, M transmitters may contribute unequally to each of N receivers. Motivated by our prior work [5], we model the impulse response of general UWOC MIMO links with relatively large values of τ by M-order WGFP, which is the superposition of contributions from all transmitters to the impulse response. The closed-form expression of WGFP is given by

$$h(t)=\sum_{m=1}^{M} C_m \Delta_t^{\alpha_m} e^{-D_m \Delta_t}, t \geq t_0, \tag{1}$$

where α_m is the m-th weight coefficient used to adjust or balance the contribution from the m-th single Gamma function, $(C_m, D_m, \alpha_m)_{m=1}^{M}$ is the parameter set to be solved, and $\Delta_t = t - t_0$ with t_0 is the theoretical minimum arrival time. Based on the nonlinear least square (NLS) criterion, the parameter set $(C_m, D_m, \alpha_m)_{m=1}^{M}$ in Eq. 1 can be determined by minimizing the square error of the WGFP model using Monte Carlo simulation results.

Note that the WDGF proposed to model the impulse response of 2 × 2 UWOC MIMO links [5] has exactly the same form of 2-order WGFP and is therefore the simplest case of the WGFP model. Moreover, the DGF model in [3] is the summation of two single Gamma functions with fixed weights, and then is a special case of WDFG as well as WGFP.

Validation

In this section we present numerical examples to validate the proposed WGFP model for the impulse response of general UWOC MIMO links. We choose M light sources with 532 nm wavelength and 10° divergence angle as the transmitter array, and N photon detectors of 50 cm aperture as the receiver array to set up a UWOC MIMO link geometry as shown in Fig. 1. Transmitter and receiver arrays are located symmetrically in the xoy-plane and perpendicular to the z axis, and have their respective centers precisely aligned and far apart with a distance of L. As mentioned earlier, the impulse response of UWOC links has negligible temporal dispersion and can be modeled by an ideal delta function when the attenuation length of τ is relatively small. We therefore adopt turbid water environments with relatively large τ such as coastal and harbor water, which are widely used and have the values of parameter set (a, b, c, g) as (0.088,0.216,0.305,0.9470) and (0.295,1.875,2.17,0.9199), respectively [6].

Without loss of generality, we consider three typical UWOC MIMO systems of 2 × 2, 3 × 3, and 4 × 4 configurations, with link geometry as shown in Fig. 1, and evaluate the WGFP model of the corresponding impulse response. In these UWOC MIMO systems, each source of the transmitter array will generate and emit 6 × 10^8 photons to propagate through the turbid water environment. Then the desired detector of the receiver array can observe the impulse response by capturing all the arrival photons that may come from different sources and have their positions and arrival angles within the receiver aperture and FOV, respectively, and weights higher than the threshold. Based on these results of Monte Carlo simulations, we apply an NLS criterion to solve the parameter set $(C_m, D_m, \alpha_m)_{m=1}^{M}$ and therefore determine the impulse response.

For the simplest case of 2 × 2 UWOC MIMO link geometry, two transmitters are placed symmetrically in the xoy-plane and centered on the x axis with coordinates of –1 m and 1 m, and two receivers are located in the receiving plane parallel to the xoy-plane and centered on the line parallel to the x axis with x-coordinate of –0.5 m and 0.5 m, respectively. The WDGF model has been proposed in our prior study [5] to represent the impulse response of these 2 × 2 UWOC MIMO links with various link ranges and receiver FOVs in coastal and harbor water, and fits well with Monte Carlo simulations as shown in Fig. 2. In this figure, the impulse response (intensity) versus propagation time is plotted for 20°, 40°, and 180° receiver FOVs and various link ranges in coastal water and harbor water, respectively. Due to the symmetry of link geometry, we only plot the impulse response observed at one receiver. As mentioned earlier, the WDGF is actually 2-order WGFP.

Another insight can be obtained by examining the temporal behavior of the impulse response. In coastal water with relatively short range, FOV has less impact on the temporal behavior of the impulse response, as shown in Fig. 2(a). The small angle approximation (SAA) [7] implies that the scattering occurs mainly in the forward directions for short link range in less turbid environments such as coastal water. Therefore, the receivers with different FOVs may detect similar intensity of photons, which results in similar impulse responses. However, the impulse response disperses heavily as FOV increases for relatively large values of τ, as shown in Figs. 2(c) and 2(d). This is due to the fact that photons suffer more scattering for longer propagation distance and/or in rich scattering environments such as harbor water, which is more turbid than coastal water. Then the receiver with larger FOV can capture more scattered photons and observe an impulse response with a heavily temporal dispersion [8]. Moreover, we also consider the impact of MIMO links on time dispersion, which is defined similarly to [2] as the time interval of an impulse response falling 20 dB below the peak. With the same configurations, e.g. 12 m link range in both coastal and harbor water, the time dispersion of 2 × 2 UWOC MIMO links is slightly larger than that of UWOC SISO links, which is described in [3]. This is due to the fact that two sources emit more photons, and the desired receiver then has a higher probability to detect more scattered photons and observe a relatively large time dispersion.

Further examinations are carried out for 3 × 3 and 4 × 4 UWOC MIMO systems with link geometry depicted in Fig. 1. Figure 3 and Fig. 4 plot the impulse response of 3 × 3 UWOC MIMO links observed at the outer and inner receivers with 20°, 40°, and 180° FOVs for various link ranges in coastal and harbor water, respectively. From these figures we can observe that the proposed WGFP model fits well with Monte Carlo simulations for both water types regardless of receiver FOV and position. Similar phenomena can be observed from the impulse response for 4 × 4 UWOC MIMO systems, as shown in Fig. 5 and Fig. 6.

The coefficient of determination (R-square) [9] is a widely used similarity metric for curve fitting, and is adopted in this work to indicate how well the proposed WGFP model fits with Monte Carlo results. R-square takes a value from 0 to 1 and has a larger value to represent a higher similarity. All test cases above have the values of R-square higher than 99.51 percent, which implies that WGFP can well model the impulse response of $M \times N$ UWOC systems in turbid water environments.

Discussion

In this section we briefly discuss applicable underwater environments and system configurations for the proposed WGFP model such as the water type, link range, transmitter divergence angle, receiver aperture and FOV, as well as elevation angles and inter-spacings of Tx/Rx arrays. Since temporal dispersion of the impulse response is mainly caused by the scattering effect in the propagation medium, the proposed WGFP model then works well in the scattering dominant region and may break down in other regions. As mentioned earlier, the impulse response suffers negligible temporal dispersion and can be modeled by an ideal delta function in clear water where the WGFP model is no longer suitable for modeling the impulse response.

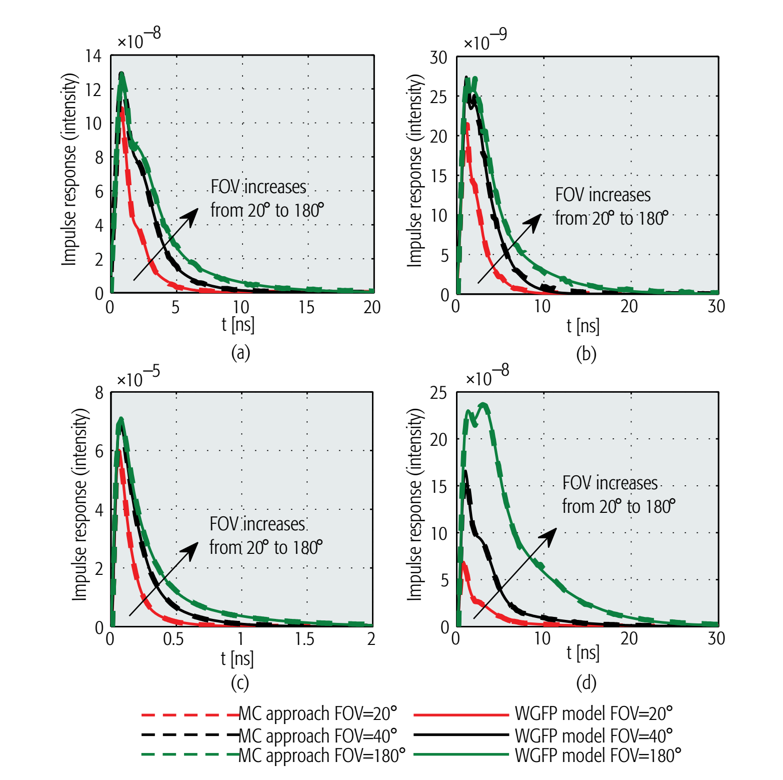


Figure 3. Impulse response observed at two outer receivers of the 3 × 3 MIMO link: a) L = 50 m link in coastal water τ = 15.25; b) L = 60 m link in coastal water τ = 18.3; c) L = 5 m link in harbor water τ = 10.85; and d) L = 12 m link in harbor water τ = 26.04.

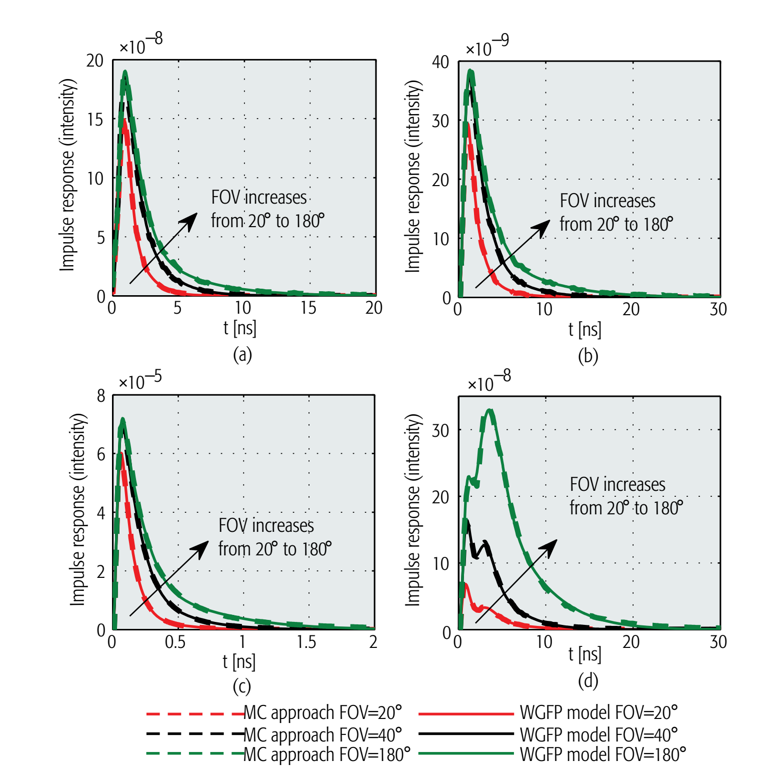


Figure 4. Impulse response observed at one inner receiver of the 3 × 3 MIMO link: a) L = 50 m link in coastal water τ = 15.25; b) L = 60 m link in coastal water τ = 18.3; c) L = 5 m link in harbor water τ = 10.85; and d) L = 12 m link in harbor water τ = 26.04.

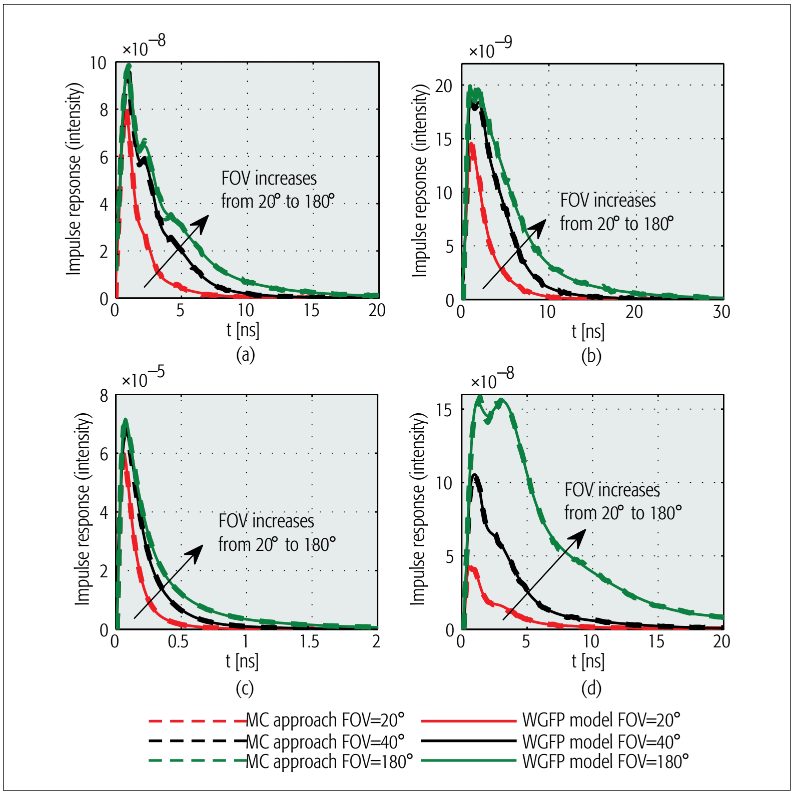


Figure 5. Impulse response observed at two outer receivers of the 4 × 4 MIMO link: a) $L = 50$ m link in coastal water $\tau = 15.25$; b) $L = 60$ m link in coastal water $\tau = 18.3$; c) $L = 5$ m link in harbor water $\tau = 10.85$; and d) $L = 12$ m link in harbor water $\tau = 26.04$.

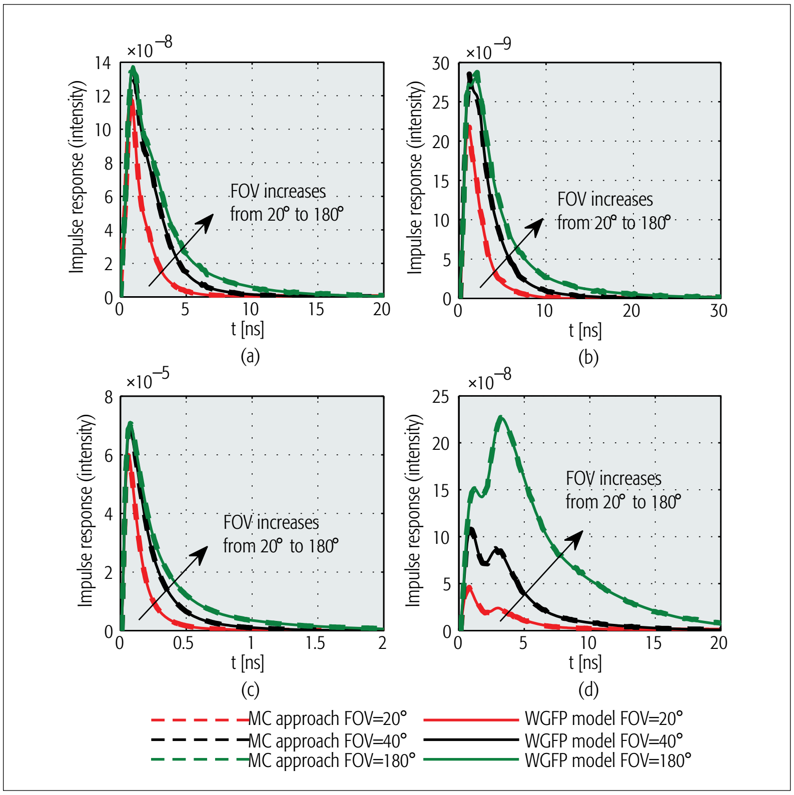


Figure 6. Impulse response observed at two inner receivers of the 4 × 4 MIMO link: a) $L = 50$ m link in coastal water $\tau = 15.25$; b) $L = 60$ m link in coastal water $\tau = 18.3$; c) $L = 5$ m link in harbor water $\tau = 10.85$; and d) $L = 12$ m link in harbor water $\tau = 26.04$.

Similar to [3], we consider narrow configuration systems with relatively small divergence of sources, compact receivers, and narrow FOV, and wide configuration systems with relatively large divergence of sources, large receiver aperture size, and wide FOV, respectively. For narrow configuration systems, we have verified that WGFP is valid only for large enough attenuation lengths that imply turbid water type and/or long link range. This is intuitive since turbid water contains more scatterers, and a long link range increases the chances for photons to be scattered. However, wide configuration systems alleviate the requirement for large enough attenuation lengths since the receiver can still capture the scattering light of large displacement or arrival angle and therefore make the WGFP model valid even for relatively small attenuation lengths.

For UWOC MIMO systems, the inter-spacing between two neighboring sources at the transmitter array or detectors at the receiver array also has an important effect on the validity of the WGPF model. At the side of the transmitter array, small inter-spacing may introduce channel correlation, which breaks down the WGFP model. A possible solution is to increase the order of the proposed WGFP model since *M* Gamma functions are not enough to characterize the temporal behavior of correlated UWOC MIMO links. For consideration of uncorrelated UWOC MIMO links, we set the inter-spacing of the transmitter array as two meters for 2 × 2 systems and three meters for 3 × 3 and 4 × 4 systems. For the simplest case of 2 × 2 UWOC MIMO links with 3.66 attenuation length as shown in Fig. 2(a), the value of R-square for impulse response modeling tends to be lower than 99 percent when the inter-spacing of the transmitter array is less than 2 m. Furthermore, the inter-spacings at both sides may affect the actual communication distance as well as the observation of impulse response at a given point on the receiver side. In this study, we set the Tx and Rx inter-spacings in 3 × 3 and 4 × 4 UWOC MIMO systems as three meters and larger than those in the 2 × 2 case.

The elevation angles of the Tx/Rx arrays also affect the modeling of the impulse response. For simplicity, a mirror symmetry of Tx elevation angles is adopted as shown in Fig. 1. For 2 × 2 UWOC MIMO links with 180° receiver FOV and 12 m link range in coastal water, we have observed that 2-order WGFP is valid when Tx elevation angle is zero (default value) and 30° but fails to fit Monte Carlo results for 10° or 20° Tx elevation angle. As the Tx elevation angle varies from 0° to 30°, light beams from different transmitters become close to increase the channel correlation and then far away to decrease the correlation. As mentioned earlier, the proposed WGFP model may break down for high channel correlation when light beams from different transmitters get close enough. We have also investigated the joint effect of attenuation length and Tx elevation angle on channel correlation. Compared with the failed case of 12 m link range with 10° Tx elevation angle in coastal water, the proposed WGFP model fits well with Monte Carlo simulations for 30 m link range in coastal water and 12 m link range in harbor water with the same 10° Tx elevation angle. This observation shows that a relatively large attenuation length, i.e. more turbid water and/or longer link range, enriches the scattering environment and therefore may alleviate the correlation caused by improper Tx elevation angles. A similar conclusion can be drawn that a relatively large attenuation length can reduce the channel correlation introduced by small inter-spacings at both sides.

Generally speaking, the proposed weighted Gamma function polynomial can well model the impulse response of general UWOC MIMO links in the scattering dominant region with common system configurations, taking into consideration the water type, link range, transmitter divergence angle, receiver aperture and FOV, elevation angles, and inter-spacings of Tx/Rx arrays. Compared with UWOC SISO systems, UWOC MIMO systems can enjoy the benefits of MIMO techniques such as high power efficiency while also introducing channel correlation as well due to small inter-spacings at the Tx/Rx arrays. Based on the analysis above, channel correlation will affect the validity of the proposed WGFP model, which can be improved by increasing the attenuation length.

Conclusion

In this article we investigated the impulse response of UWOC MIMO links with arbitrary numbers of light sources and detectors. We proposed a closed-form expression of weighted Gamma function polynomial to model the impulse response of these general UWOC MIMO links. Numerical examples of 2 × 2, 3 × 3, and 4 × 4 UWOC MIMO links suggest that the proposed WGFP model fits well with Monte Carlo simulations in turbid water environments such as coastal and harbor water. We also investigated the applicable underwater environment and system configurations of the proposed WGFP model, including the water type, link range, transmitter divergence angle, receiver aperture and FOV, as well as elevation angles and inter-spacings of the Tx/Rx arrays. Small inter-spacings and improper Tx/Rx elevation angles at both sides may introduce channel correlation and break down the WGFP model, while more turbid water and/or longer link range may alleviate the effect of correlation and therefore revalidate the proposed model. It is plausible that the simple closed-form expression of WGFP for impulse response modeling can facilitate the performance evaluation of UWOC MIMO systems as well as system design and enhancement by making full use of MIMO techniques.

It is plausible that the simple closed-form expression of WGFP for impulse response modeling can facilitate the performance evaluation of UWOC MIMO systems as well as system design and enhancement by making full use of MIMO techniques.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 61201184.

References

[1] J. R. Potter, M. B. Porter, and J. C. Preisig, "UComms: A Conference and Workshop on Underwater Communications, Channel Modeling, and Validation," *IEEE J. Oceanic Eng.*, vol. 38, no. 4, Oct. 2013, pp. 603–13.

[2] C. Gabriel *et al.*, "Monte-Carlo-based Channel Characterization for Underwater Optical Communication Systems," *J. Opt. Commun. Net.*, vol. 8, no. 1, Jan. 2013, pp. 1–12.

[3] S. Tang, Y. Dong, and X. Zhang, "Impulse Response Modeling for Underwater Wireless Optical Communication Links," *IEEE Trans. Commun.*, vol. 62, no. 1, Jan. 2014, pp. 226–34.

[4] B. Cochenour, L. Mullen, and J. Muth, "Temporal Response of the Underwater Optical Channel for High-Bandwidth Wireless Laser Communications," *IEEE J. Oceanic Eng.*, vol. 38, no. 4, Oct. 2013, pp. 730–42.

[5] Y. Dong, H. Zhang, and X. Zhang, "On Impulse Response Modeling for Underwater Wireless Optical MIMO Links," *Proc. IEEE/CIC Int'l. Conf. Commun. China (ICCC '14)*, Shanghai, China, 2014, pp. 151–55.

[6] C. D. Mobley, *Light and Water: Radiative Transfer in Natural Waters*, New York, NY, USA: Academic/Elsevier, 1994, chs. 3, 5.

[7] W. H. Wells, "Theory of Small Angle Scattering," *Optics of the Sea*, vol. 61, ser. AGARD Lect. Brussels, Belgium: IEEE, 1973, ch. 3.3.

[8] W. Cox, "Simulation, Modeling, and Design of Underwater Optical Communication Systems," Ph.D. dissertation, Dept. Elect. Comput. Eng., North Carolina State Univ., Raleigh, NC, USA, 2012.

[9] N. J. Nagelkerke, "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, vol. 78, no. 3, Sept. 1991, pp. 691–92.

Biographies

Huihui Zhang [S] is a master candidate in electronic engineering, and is currently with the Modern Communication Laboratory at the Graduate School at Shenzhen, Tsinghua University. He received the B.S. degree in information engineering from Southeast University in 2012. His research interests include channel modeling and system design of underwater wireless optical communications for both SISO and MIMO systems.

Yuhan Dong [M] (dongyuhan@sz.tsinghua.edu.cn) is with the Graduate School at Shenzhen, Tsinghua University, where he is currently an associate professor and leads the Underwater Wireless Optical Communication Group. He received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, and the Ph.D. degree in electrical engineering from North Carolina State University, Raleigh, NC, USA, in 2002, 2005, and 2009, respectively. He is a member of the OSA.

ADVANCES IN OPTICAL COMMUNICATIONS NETWORKS

Osman Gebizlioglu

Vijay Jain

The global telecommunications industry has been focusing on network functions virtualization (NFV) and software-defined networking (SDN) to provide service providers with the tools for more effective operation and management of communications networks. These important trends are expected to gain additional momentum in 2016. Recent developments in the Internet of Things and cloud computing promise to make programmable network management and control an imperative. In 2015, we witnessed the evolution of more efficient optical transport capabilities delivered by the optical components and systems suppliers as high-speed optical and data center networks expanded globally. Underlying this evolution has been the ever accelerating development of high-speed interconnects and optical transceivers for data center networks. We expect to see continued migration to all-optical networks in 2016 and beyond.

In this issue, we have selected four contributions that address the coexistence of Wi-Fi and visible light communications (VLC), 1 Gb/s service provisioning with next generation passive optical networking (NG-PON) technologies, migration strategies for active optical networks, and SDN for data center optical interconnection.

In the first contribution, "Coexistence of Wi-Fi and Li-Fi toward 5G: Concepts, Opportunities, and Challenges," M. Ayyash, H. Elgala, A. Khreishah, V. Jungnickel, T. Little, S. Shao, M. Rahaim, D. Schulz, J. Hilt, and R. Freund present a status review of the wireless communications network capacity to meet the needs of current and future multimedia applications. Wireless heterogeneous networks (HetNets) are expected to play an important role toward the goal of using a diverse spectrum to provide high quality of service (QoS) in indoor data consumption environments. An additional capability in the wireless HetNets concept is from indoor gigabit small cells (SCs). The use of light as a new mobile access medium is considered promising. In this article, the authors describe the general characteristics of Wi-Fi and VLC (or Li-Fi) and demonstrate a practical framework for both technologies to coexist.

In the second contribution, "Provisioning 1 Gb/s Symmetrical Services with Next-Generation Passive Optical Network Technologies," R. Sanchez, J. A. Hernandez, J. Montalvo Garcıa, and D. Larrabeiti present a technical and economic comparison of four NG-PON standard optical access technologies: GPON, XGPON, WDM-PON, and the emerging TWDM-PON. Service providers have been making large investments to upgrade their broadband access networks, and optical fiber has been considered as the technology of choice in the long term due to its transmission rate and reach. Optical access technologies have a distinct advantage over other broadband access technologies for symmetrical downstream and upstream transmission. In this contribution, the authors analyze the delivery of symmetrical 1 Gb/s access to residential users with a target temporal guarantee at the least cost using NG-PON technologies. Their analysis shows that only TWDM-PON can provide 1 Gb/s service guaranteed at a moderate cost when compared to a fully dedicated 1 Gb/s point-to-point connection.

In the third contribution, "Migration Strategies for FTTx Solutions based on Active Optical Networks," K. Wang, A. Gavler, C. M. Machuca, L. Wosinska, K. Brunnström, and J. Chen present migration strategies for the active optical network (AON) from the data plane, topology, and control plane perspectives. They discuss results of their investigations on the impact of these strategies on the total cost of ownership (TCO). The AON has been one of the most widely deployed fiber access solutions in Europe, and service providers have been facing the need to upgrade their AONs to keep up with the ever growing bandwidth demand driven by new applications and services. As service providers migrate their AONs, they aim at achieving their primary goal of savings in capital and operational expenditures.

In the fourth contribution, "SUDOI: Software Defined Networking for Ubiquitous Data Center Optical Interconnection," H. Yang, J. Zhang, Y. Zhao, J. Han, Yi Lin, and Y. Lee present a new software-defined data center optical interconnection (SUDOI) architecture to enable extensive user access from the perspective of multi-layer networking modes. The feasibility and efficiency of the proposed architecture are experimentally demonstrated on an optical-as-a-service testbed with OpenFlow-enabled optical nodes, and compared in terms of blocking probability and resource occupation rate. The functional modules of SUDOI architecture, including the core elements of various controllers, are described in detail. The cooperation in user-access-oriented interconnection, and multi-layer resource integration in inter- and intra-data center service modes is investigated. Future capabilities enabled are also explored by the authors.

In this first Optical Communications Series (OCS) issue of 2016, we thank all authors and reviewers for their contributions to the OCS in 2015. This issue marks the end of our three-year term as Guest Editors. It is our great pleasure to hand over this privilege to a new OCS GE team of Professor Admela Jukan, Technical University of Braunschweig, Germany, and Dr. Xiang Liu of Futurewei Technologies, Inc./Huawei Technologies US R&D Center, Bridgewater, New Jersey. We wish them well and request your support.

CALL FOR PAPERS
IEEE COMMUNICATIONS MAGAZINE
GREEN COMMUNICATIONS AND COMPUTING NETWORKS SERIES

BACKGROUND

Green Communications and Computing Networks is published semi-annually as a recurring Series in *IEEE Communications Magazine*. The objective of this Series is to provide a premier forum across academia and industry to address all important issues relevant to green communications, computing, and systems. The Series will explore specific green themes in depth, highlighting recent research achievements in the field. Contributions provide insight into relevant theoretical and practical issues from different perspectives, address the environmental impact of the development of information and communication technologies (ICT) industries, discuss the importance and benefits of achieving green ICT, and introduce the efforts and challenges in green ICT. This Series welcomes submissions on various cross-disciplinary topics relevant to green ICT. Both original research and review papers are encouraged. Possible topics in this series include, but are not limited to:

- Green concepts, principles, mechanisms, design, algorithms, analyses, and research challenges
- Green characterization, metrics, performance, measurement, profiling, testbeds, and results
- Context-based green awareness
- Energy efficiency
- Resource efficiency
- Green wireless and/or wireline communications
- Use of cognitive principles to achieve green objectives
- Sustainability, environmental protections by and for ICT
- ICT for green objectives
- Non-energy relevant green issues, and/or approaches
- Power-efficient cooling and air-conditioning
- Green software, hardware, device, and equipment
- Environmental monitoring
- Electromagnetic pollution mitigation
- Green data storage, data centers, contention distribution networks, and cloud computing
- Energy harvesting, storage, transfer, and recycling
- Relevant standardizations, policies, and regulations
- Green smart grids
- Green security strategies and designs
- Green engineering, agenda, supply chains, logistics, audit, and industrial processes
- Green building, factory, office, and campus designs
- Application layer issues
- Green scheduling and/or resource allocation
- Green services and operations
- Approaches and issues of social networks used to achieve green behaviors and objectives
- Economic and business impact and issues of green computing, communications, and systems
- Cost, OPEX and CAPEX for green computing, communications, and systems
- Roadmap for sustainable ICT
- Interdisciplinary green technologies and issues
- Recycling and reuse
- Prospect and impact on carbon emissions and climate policy
- Social awareness of the importance of sustainable and green communications and computing

SUBMISSION GUIDELINES

Prospective authors are strongly encouraged to contact the Series Editor with a brief abstract of the article to be submitted, before writing and submitting an article in order to ensure that the article will be appropriate for the Series. All manuscripts should conform to the standard format as indicated in the submission guidelines at

http://www.comsoc.org/commag/paper-submission-guidelines

Manuscripts must be submitted through the magazine's submissions web site at

http://mc.manuscriptcentral.com/commag-ieee

You will need to register and then proceed to the Author Center. On the manuscript details page, please select "Green Communications and Computing Networks Series" from the drop-down menu.

SCHEDULE FOR SUBMISSIONS

Scheduled Publication Dates: Twice per year, May and November

SERIES EDITORS

Jinsong Wu, Alcatel-Lucent, China, wujs@ieee.org
John Thompson, University of Edinburgh, UK, john.thompson@ed.ac.uk
Honggang Zhang, UEB/Supelec, France; Zhejiang Univ., China, honggangzhang@zju.edu.cn
Daniel C. Kilper, University of Arizona, USA, dkilper@optics.arizona.edu

Optical Communications

Coexistence of WiFi and LiFi Toward 5G: Concepts, Opportunities, and Challenges

Moussa Ayyash, Hany Elgala, Abdallah Khreishah, Volker Jungnickel, Thomas Little, Sihua Shao, Michael Rahaim, Dominic Schulz, Jonas Hilt, and Ronald Freund

The authors describe the general characteristics of WiFi and VLC (or LiFi) and demonstrate a practical framework for both technologies to coexist. They explore the existing research activity in this area and articulate current and future research challenges based on their experience in building a proof-of-concept prototype VLC HetNet.

Abstract

Smart phones, tablets, and the rise of the Internet of Things are driving an insatiable demand for wireless capacity. This demand requires networking and Internet infrastructures to evolve to meet the needs of current and future multimedia applications. Wireless HetNets will play an important role toward the goal of using a diverse spectrum to provide high quality-of-service, especially in indoor environments where most data are consumed. An additional tier in the wireless HetNets concept is envisioned using indoor gigabit small-cells to offer additional wireless capacity where it is needed the most. The use of light as a new mobile access medium is considered promising. In this article, we describe the general characteristics of WiFi and VLC (or LiFi) and demonstrate a practical framework for both technologies to coexist. We explore the existing research activity in this area and articulate current and future research challenges based on our experience in building a proof-of-concept prototype VLC HetNet.

Introduction

The number of multimedia-capable and Internet-connected mobile devices is rapidly increasing. Watching HD streaming videos and accessing cloud-based services are the main user activities consuming data capacity, now and in the near future. Most of this data consumption occurs indoors, and increasingly in spaces such as aircraft and other vehicles. This high demand for video and cloud-based data is expected to grow and is a strong motivator for the adoption of new spectrum, including the use of optical wireless media. In terms of network topology, heterogeneous networks (HetNets) will play an important role in integrating a diverse spectrum to provide high quality-of-service (QoS), especially in indoor environments where there is localized infrastructure supporting short-range directional wireless access. We envision multi-tier HetNets that utilize a combination of macrocells providing broad lower-rate services, RF small-cells (RF-SCs) providing improved coverage at locations occupied by users, and LiFi small cells that provide additional capacity through the use of the optical spectrum. Indoor RF-SCs, including licensed femtocells and/or unlicensed WiFi access points (APs), deployed under coverage of macrocells, can take over the connection when moving indoors. In this manner, WiFi enables traffic offloading from these capacity-stressed licensed macrocells or RF-SCs [1]. According to Cisco Visual Networking Index (*Global Mobile Data Traffic Forecast Update (2014–2019)*), approximately 50 percent of this traffic is expected to be offloaded to WiFi in 2016.

The State of Wireless and Mobile Communications

Except in dense WiFi networks, where contention is possible, high signal strength in indoor access WiFi networks is an indicator of a fast and reliable WiFi connection. In a building with different types of walls and other obstructions, and as distance increases, the WiFi signal strength is attenuated. Accordingly, if in one room the signal strength is much attenuated, WiFi users experience poor connectivity and slow speed. Slow connectivity is also caused by high interference signal from neighboring WiFi APs and/or multiple active users sharing the limited bandwidth of a WiFi AP.

The WiFi evolution considers higher frequencies with new spectrum to reach multi-Gb/s peak data rates (WiGig (www.wigig.com) at 60 GHz) indoors and to serve multiple users in parallel. While the IEEE 802.11ad (WiGig) wireless local area network (WLAN) implementations are beginning to reach the consumer market in tri-band products (2.4 GHz, 5 GHz, and 60 GHz), optical wireless communications (OWC) systems, specifically based on visible light communications (VLC) technology, also called LiFi, offer dual-functionality to transmit data on the intensity of optical sources (lighting concurrent with data communication) [2]. The authors in [3] describe an integrated architecture for 5G mobile networks that includes SCs and enhanced WiFi as the main scaling factor for wireless capacity. However, and especially in dense deployments, the sustainable performance of WiFi can be reduced, as the carrier sense multiple access with collision avoidance (CSMA/CA) allows only one link to be active at

Moussa Ayyash is with Chicago State University; Hany Elgala is with State University of New York (SUNY) – Albany; Abdallah Khreishah and Sihua Shao are with New Jersey Institute of Technology; Volker Jungnickel, Dominc Schulz, Jonas Hilt, and Ronald Freund are with Fraunhofer Heinrich Hertz Institute; Thomas DC Little and Michael Rahaim are with Boston University.

0163-6804/16/$25.00 © 2016 IEEE

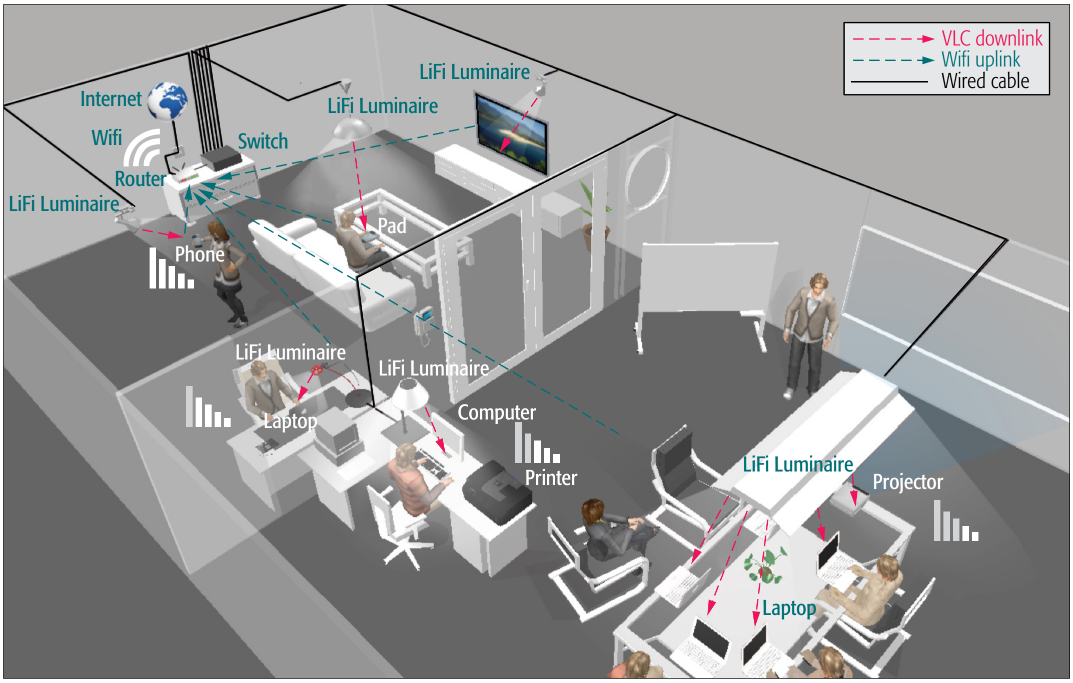


Figure 1. The proposed Li+WiFi HetNet.

Security enhancement is an obvious result because visible light doesn't penetrate through walls and improved indoor positioning is a result of a better resolution in a centimeter range compared to other RF based technologies including WiFi.

once as it is somewhat random, demand-driven, and not always fair. For example, the first user detecting an unused channel is allowed to start transmission, independent of its channel quality. However, if there is a demand from another user having a better channel at some later time, such demand cannot be served because the first link is not interrupted due to the CSMA/CA rule that the next transmission starts only if the channel is free. This situation is exacerbated with the increased adoption of IP video streaming, which increases both data utilization and the need for continuous gap-free data delivery.

Therefore, concurrent multiuser transmission is used in WiFi as a next step, similar to the enabled multiuser multiple-input and multiple-output (MU-MIMO) in Long-Term Evolution (LTE). In dense environments, cooperative beamforming between adjacent APs is also considered [2].

However, a big standardization effort is needed to define such a new mode of simultaneous transmissions to multiple users that must remain backward-compatible. Moreover, there are complexity limits with larger numbers of antennas. It is well known that the complexity of linear MIMO equalizers scales with N^3, where N is the number of antennas, while optimal scheduling problems, in particular between the beams of multiple adjacent APs, are NP-hard. Recently, a practical solution has been developed (see [3] and references therein). Due to these standardization, scalability, and complexity issues, and due to the increasing demand for WiFi, scalability is limited and there is a rationale to consider other wireless media.

Getting to High Capacity and Density

Given the aforementioned challenges, we envision an additional tier in wireless HetNets comprised of indoor gigabit SCs to offer additional wireless capacity where it is needed the most. LiFi-enabled indoor luminaires (lights) can be modeled as optical SCs (O-SCs) in a HetNet, where a three-layer network formed by RF macrocells, RF-SCs, and O-SCs are deployed. Offloading traffic to the most localized and directional LiFi is expected to enhance the performance of a single WiFi AP or across multiple WiFi APs. Besides high-speed traffic offloading with seamless connectivity, the proposed Li+WiFi system also offers new interesting features, such as enhanced security in O-SC and improved indoor positioning [4]. Security enhancement is an obvious result because visible light does not penetrate through walls, and improved indoor positioning is a result of a better resolution in a centimeter range compared to other RF based technologies, including WiFi.

Operators say that 80 percent of mobile traffic occurs indoors; therefore, the combination of LiFi and WiFi has great potential to be a breakthrough technology in future HetNets, including next generation (5G) mobile telecommunications systems [5, 6]. To our knowledge, the state-of-the-art research is currently focused on enhancing the performance of each of the technologies alone, while there is a clear need for reliable WiFi and LiFi coexistence solutions [7].

As shown in Fig. 1, stationary and quasi-stationary mobile users are provided data access via LiFi-enabled light fixtures, or luminaires, in lighting parlance. This approach can alleviate congestion and free RF resources to serve users who are more mobile or outside the LiFi coverage area. More highly mobile users will be able to fall back on the broader coverage of the WiFi network.

In the Li+WiFi network, user devices (UDs) must be LiFi-enabled. To evaluate the development of LiFi-enabled devices, the evolution of cellular networks can be used for reference. Evolving from 1G to 4G, mobile technologies blaze the trail for marketing more advanced and

The issue of overlapping and non-overlapping coverage of the distributed luminaires needs careful examination. It has a major impact on the handover not only between WiFi and LiFi-enabled luminaires but also among the distributed luminaires themselves.

more expensive user devices. By delivering richer mobile broadband experiences, LiFi-enabled smartphones offer manufactures considerable profitability. Actually, most modern smartphones already support multiple radios and protocols. Even though the Li+WiFi network is likely to be asymmetric with LiFi as the downlink, this should free up WiFi system capacity to accommodate any future growth in traffic-uploading. This is due to challenges to overcoming upward link alignment, glare, and energy consumption factors in the handset. But despite the asymmetry, the benefits of the added VLC channel are significant. Our work, and this article, are motivated by promising preliminary results using high-throughput LiFi transceivers utilized in a proof-of-concept hybrid Li+WiFi demonstration [8, 9].

A HetNet Vision Incorporating VLC and Current Research Activities

Central issues in designing and managing a Li+WiFi network include dealing with how a UD attaches to the network, how mobility is supported as a device moves from cell to cell and between networks, and how multiple users are accommodated. Ultimately, the combined performance of the LiFi and WiFi networks aggregate to match available capacity to where devices need it. In this section, we describe the proposed Li+WiFi network with a goal to provide seamless connectivity and to optimally distribute resources among users. Also, we consider some of the most relevant recent works addressing present challenges.

Multiple Links and Aggregation

Because luminaires are distributed throughout our living spaces, it is often possible to "see" more than one at a time. This fact can be exploited using a multichannel receiver. Imagine that the lighting infrastructure is potentially enabling MIMO transmission using a multi-detector UD. However, reconciling the optimal link or links involving one or more luminaires in the presence of multiple UDs is challenging. This is more difficult with mobility and changing UD orientation. Therefore, reliable sensing of the optical link quality between individual luminaires within the UD receiver's field-of-view is critical and requires careful investigation. Previous work assumes that the transmitter exactly knows the channel state information (CSI) from each UD in the room. However, accurate CSI may be relatively easier to obtain in a static condition, and from a practical perspective in the case of user mobility, obtaining the CSI is an estimation problem that cannot be error free. Therefore, it is important to understand the effect of the channel estimation error on the system throughput in a multiuser environment for time-varying single-input single-output (SISO) and MIMO wireless channels.

On the other hand, connecting a user on multiple optical channels might be an advantage, whenever the application needs high throughput. Since multiple LiFi-enabled luminaires are in each room, modulation frequency sub-bands and wavelengths can both be reused at some distance to achieve a higher throughput. Carrier and channel aggregation, similar to LTE-Advanced, is one key approach to increase the overall transmission bandwidth. Performing aggregation in the Li+WiFi network needs efficient methods to split the overall traffic between the RF and optical links, to handle packet drops on the individual links, and to reorder the packets, accordingly. These issues clearly affect higher layer protocols such as the transmission control protocol (TCP). In scenarios in which a user can be attached to a single luminaire (SISO configuration) or simultaneously to multiple luminaires (MIMO configuration), three possible access scenarios can be considered. Initially, the user is served by a single luminaire providing the highest link quality. Multiple luminaires serving a single user are allowed to satisfy the user's requirements. However, and to insure fairness and minimum QoS among multiple users, especially in a dense user scenario, the number of luminaires serving a single user can be managed depending on resource availability.

MIMO research activities on LiFi typically consider the single-user MIMO (SU-MIMO) scenario, where a single multi-detector UD is communicating with a single multi-chip LED based luminaire or multiple distributed luminaires. The limited spatial separation between the different detectors on a single UD suggests pointing them to different directions to maximize receiver diversity. As shown in Fig. 2, for a SU-MIMO, the singular value decomposition (SVD) based MIMO transmission can ideally support parallel links and maximize the capacity while satisfying illumination constraints [10]. However, SU-MIMO LiFi channels can be highly correlated [10], which needs a joint rank-adaptation and rate-adaptation to the channel similar to RF wireless links.

As already mentioned, optical beamforming, e.g. through a spatial light modulator (SLM), can provide enhanced spatial separation and channel quality [11]. In a MU-MIMO, the rank of the MIMO channel can be improved depending on the selected user locations. Multiple luminaires can send signals to multi-detector UDs to serve these multiple users in parallel. Note that such parallel transmissions are common in RF communications, while multiple-source, multiple-access schemes, also including multi-color luminaires, are only just emerging from early lab prototypes. In a practical indoor VLC deployment, target illumination and color quality must be maintained while maximizing the system throughput and supporting each user's mobility.

Mobility and Medium Access

The issue of overlapping and non-overlapping coverage of the distributed luminaires needs careful examination. It has a major impact on the handover not only between WiFi and LiFi-enabled luminaires but also among the distributed luminaires themselves [9]. The handover mechanism may also involve information about UD location, which can be realized using both technologies, while LiFi is probably more precise.

Resource allocation and scheduling are important aspects of QoS support in wireless networks. In order to support mobility, they need adaptation to changing channels on both slow

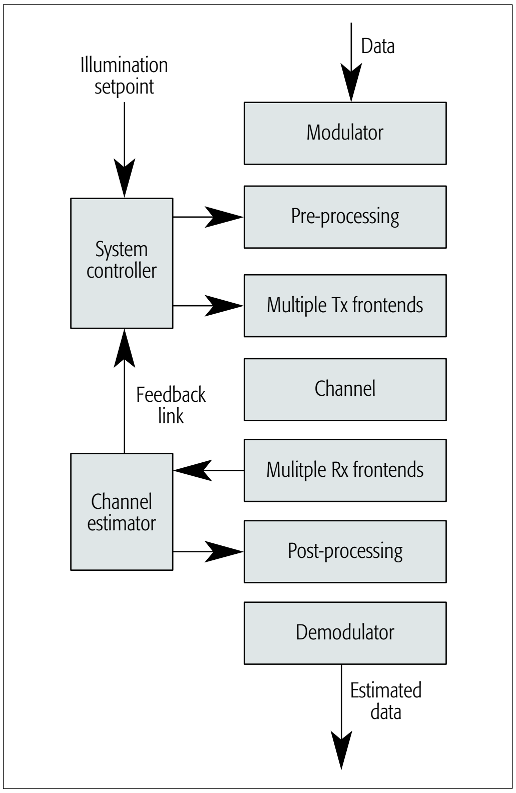


Figure 2. The SU-SVD-MIMO concept can be used to avoid interference and maintain target illumination. The SVD is used to decompose the MIMO channel into parallel SISO sub-channels, enabling interference-free spatial multiplexing. At the receiver, and after estimating the channel, the information needed to pre-process and post-process the signals at the transmitter and receiver, respectively, and the illumination set point (room brightness) is available on the feedback channel, to extract the parallel SISO channels.

and fast time scales. While the LiFi link changes more slowly, as the instantaneous signal power is proportional to the integral of the optical power over the detector surface, the WiFi link is subject to fast fading where the radio channel can fade randomly over a few centimeters passed during a few milliseconds.

Moreover and as discussed earlier, the drawback of CSMA/CA in WiFi is particularly notable in scenarios where low latency is required for multiple users in parallel [12]. Moreover, WiFi standards are backward compatible, and typical environments with a mix of clients and protocols do not achieve the peak performance specified in standards. These WiFi issues are solved using MU-MIMO and coordinated beamforming (see [2]). By offloading the data of users with high-quality channels on optical links, WiFi CSMA/CA fairness of resource allocation issues can be improved. Also, offloading removes congestion and interference within the same WLAN and other networks in the area.

Maintaining continuous connectivity for mobile users is the first challenge. Handover on the same wireless access technology is needed due to the small coverage area created by each luminaire as well as the limited number of luminaires per room. Hence, user mobility triggers frequent switching among the O-SCs, resulting in connectivity losses and/or undesired latency. This handover may thus be complemented by a second handover mechanism, where the traffic from a UD is rerouted from O-SCs to RF-SCs and vice versa [6]. Handover in RF cellular networks is an important research area, where the signal-to-interference and noise ratio (SINR) is commonly the optimal metric for decisions regarding channel selection between cells within a tier. In multi-tier and/or HetNets, a preference to connect is often given to SCs. This is due to the aggregate performance improvement that dense networks provide. The sensitivity of LiFi to occlusions and vulnerability due to sudden losses in the LOS path also requires additional metrics. Specifically, a history of previous losses should be considered in the decision process because large overhead due to frequent handover may make the LiFi connection less desirable than the RF macrocell or SC.

The sensitivity of LiFi to occlusions and vulnerability due to sudden losses in the LOS path also requires additional metrics. Specifically, a history of previous losses should be considered in the decision process because large overhead due to frequent handover may make the LiFi connection less desirable than the RF macrocell or SC.

A new protocol considering mobility combined with access is presented in [13]. The handover between the SCs of the same technology and between SCs of a different technology (O-SCs to RF-SCs and vice versa) are combined using orthogonal frequency-division multiple access (OFDMA). In OFDMA, data is transmitted on orthogonal narrow-band subcarriers, where users are allocated subcarrier-groups to enable concurrent transmissions. In this OFDMA scheme, system complexity is relatively increased compared to CSMA/CA, because transmission needs a tight coordination of resource assignment in the entire network. Alternatively, and while targeting fairness among users, a parallel transmission MAC (PT-MAC) protocol containing both the CSMA/CA algorithm and parallel transmission is proposed in [4]. This PT-MAC protocol improves the throughput and efficiency of the hybrid (IEEE 802.11n and VLC) network.

Motion information can also be considered as an important and distinctive metric in the utility function for traffic routing and handover in Li+WiFi systems. For example, a predictive handoff scheme is proposed in [14] using real-time user tracking information (e.g. user location, moving direction, and velocity). This approach minimizes the number of luminaires involved in the handoff mechanism while maintaining a seamless transition. The mobility models of users and several performance metrics, such as file size, average connectivity, and system throughput, are considered in [14]. The results in [14] show that the hybrid WLAN-VLC is always better than VLC or WLAN when individually implemented for both single and multi-user cases.

A VLC network coordinator is introduced in [7] to provide a bi-directional interface between WiFi uplink and optical downlink. While the first steps have already been made, these problems need to be further investigated.

A Prototype System Proof of Concept and Results

Through a partnership among researchers from the Fraunhofer Heinrich Hertz Institute, the New Jersey Institute of Technology, Chicago

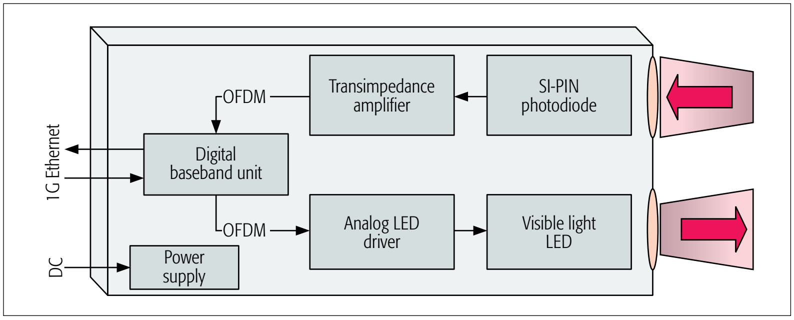


Figure 3. The LiFi transceivers.

State University, and Boston University, we have implemented a proof-of-concept Li+WiFi HetNet prototype system. In this section, we describe the various components of the system and show performance results from experimental data gained from the prototype.

Capabilities of the LiFi Transceivers:

The proposed Li+WiFi HetNet is tested using bidirectional high-speed LiFi transceiver devices that satisfy real-time data delivery and achieve layers 1 and 2 of the OSI protocol stack. The device, the principle of which is shown in Fig. 3, uses a conventional lighting-grade high-power phosphorus-converted LED (PC-LED), and it realizes both illumination and data transmission in parallel. A proprietary LED driver is used to enable an analog modulation bandwidth of up to 180 MHz. At the receiver, a large-area high-speed silicon PIN photodiode is used together with a trans-impedance amplifier (TIA). A plano-convex 1” lens is used at both the LED and the photodiode to concentrate the beam and to enlarge the receiving area, respectively.

Behind the analog transmitter and receiver circuits, a digital baseband unit (BBU) is used to convert Ethernet packets into DC-biased orthogonal frequency division multiplexing (OFDM) signals, and vice versa. The OFDM signals have a bandwidth of 70 MHz. The BBU performs pilot-assisted channel estimation and frequency-domain equalization to reconstruct the received symbol constellations. From the received pilot sequence, the error vector magnitude (EVM) is measured, and this information is fed back to the transmitter. Depending on the channel quality as a function of frequency, the bit loading is adapted. The data rate is increased as much as possible so that no errors occur after forward error correction. Thanks to the techniques used in link adaptation, implemented in real-time as a closed-loop, the achievable data rate is realized while avoiding outages due to changing channel conditions such as varying illumination levels. The relation between the data rate and the illumination level is explicitly given in [15]. Each transceiver is equipped with an external power supply and a standard RJ45 1 Gb/s Ethernet connector. Altogether, a gross and net data rate of 500 Mb/s and 270 Mb/s are possible, respectively, with one-way latency of approximately 10 ms, independent of the data rate [15].

Performance of Indoor and Outdoor LiFi Links

Indoor and outdoor experiments are conducted to measure the achievable throughput of the LiFi frontends. The distance between the transmitter and receiver is varied in the range of 2–15 meters and 2–10 meters for the indoor and outdoor experiments, respectively. In an indoor deployment, distance represents the vertical range of the O-SC. The throughput is also measured at different points away from the center of the light beam representing the horizontal distance within the coverage area of the O-SC.

Figure 4 (left) shows that the achieved throughput is 74 Mb/s and 25 Mb/s at a vertical distance of 2 m and 5 m, respectively. Note that the vertical distance will be in this range for most of the indoor applications. The data rate offered by our LiFi devices is already reduced at such distance due to the wide transmitter beam formed by the 1 inch aperture lens. Results are further reduced by using a white LED and measuring the throughput at the application layer. In [15], monochromatic LEDs were used with a 2 inch lens so that a higher throughput was measured at the physical layer. Despite those practical limitations, the single-user throughput achieved with LiFi is higher than what can be achieved using current WiFi devices based on “up to 54 Mb/s” mode (Fig. 6). Due to the small coverage area for the O-SC, the total through-

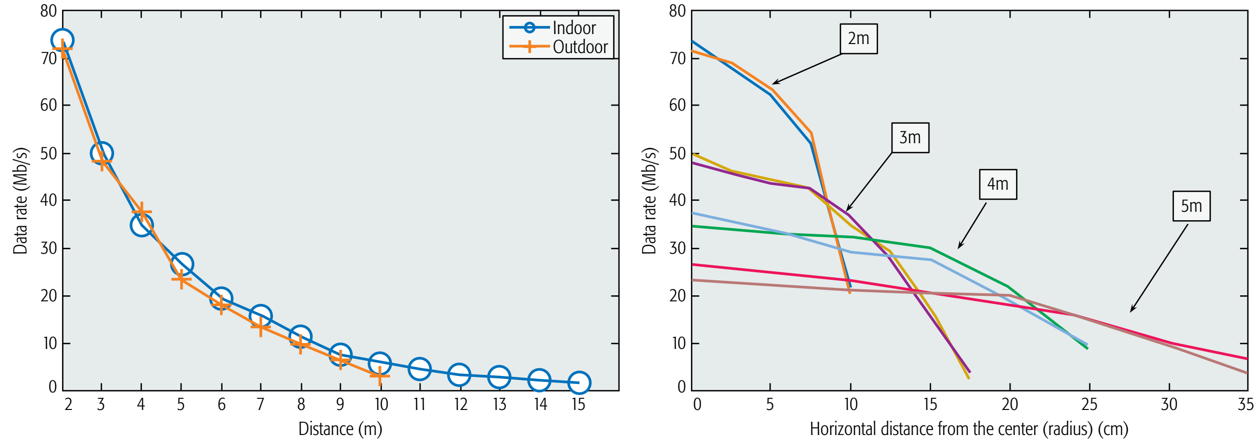


Figure 4. Vertical and horizontal distance between LiFi transceivers.

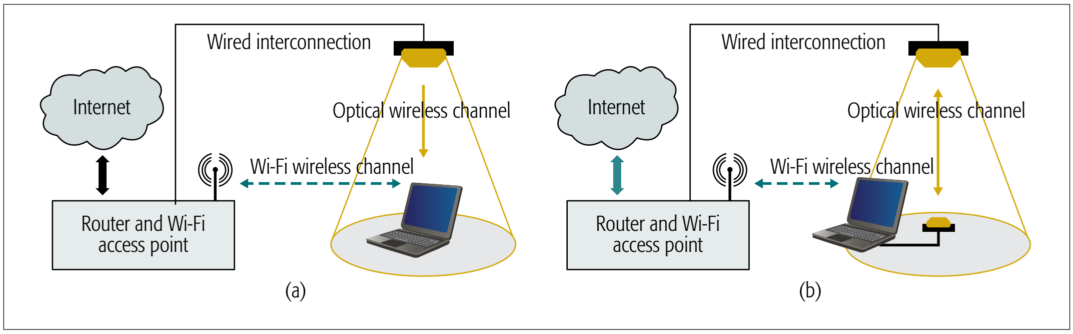


Figure 5. Configurations of the a) hybrid system, and b) the aggregated system.

put can be significantly increased by spatial reuse of the optical spectrum if multiple O-SCs are deployed serving multiple users in parallel. The results for the outdoor setting obtained during a sunny day are very close to those for the indoor setting. The results indicate that the optical frontends are robust even in outdoor conditions. While direct sunlight was avoided as it would probably disconnect the link, scattered sunlight, e.g. from back-illuminated clouds, only degrades the signal-to-interference-and-noise ratio (SINR) due to increased shot noise. In this case, the VLC transceivers adapt the data rate according to the reduced SINR.

Proof-of-Concept Experiment

A proof-of-concept hybrid Li+WiFi setup in which there is a single WiFi AP and a single LiFi AP is implemented [8, 9]. Here, three systems are compared. In the first system, the WiFi is only used to connect to the Internet. The second system, referred to as a hybrid system, is the same as the first one, but the downlink of one of the users is connected through a LiFi link. In the third system, referred to as an aggregated system, one user is connected to both WiFi and LiFi in parallel. Figure 5 depicts the configurations of the hybrid system (a) and the aggregated system (b). In the hybrid system, the unidirectional LiFi link is exploited to supplement the conventional WiFi downlink, while in the aggregated system, both bi-directional WiFi and LiFi links are fully utilized to improve the achievable throughput and provide robust network connectivity.

Figure 6a shows the average throughput of the three systems measured at different distances between the WiFi and LiFi frontends. In this setup, the LiFi frontends are strictly aligned (i.e. zero off-axis displacement). The mode of the WiFi router is selected as “up to 54 Mb/s” to provide robust connectivity in a crowded environment. Although the signaling scheme of WiFi depends on the received SNR in principle, the WiFi-only throughput shown in Fig. 6a is almost constant in the coverage area of the LiFi AP because the throughput degrading of WiFi will occur when the distance increases up to 25 meters, where the connectivity of VLC already becomes unavailable.

The hybrid system more than doubles the throughput near the LiFi AP, while degrading quickly as the distance increases. The throughput of WiFi-only surpasses that of the hybrid system when the distance is increased to around 4.1 m, because as the distance increases, the downlink capacity of LiFi decreases with distance, eventually becoming insignificant. Note that the throughput results of the hybrid VLC system depend only on the capacity of the LiFi downlink.

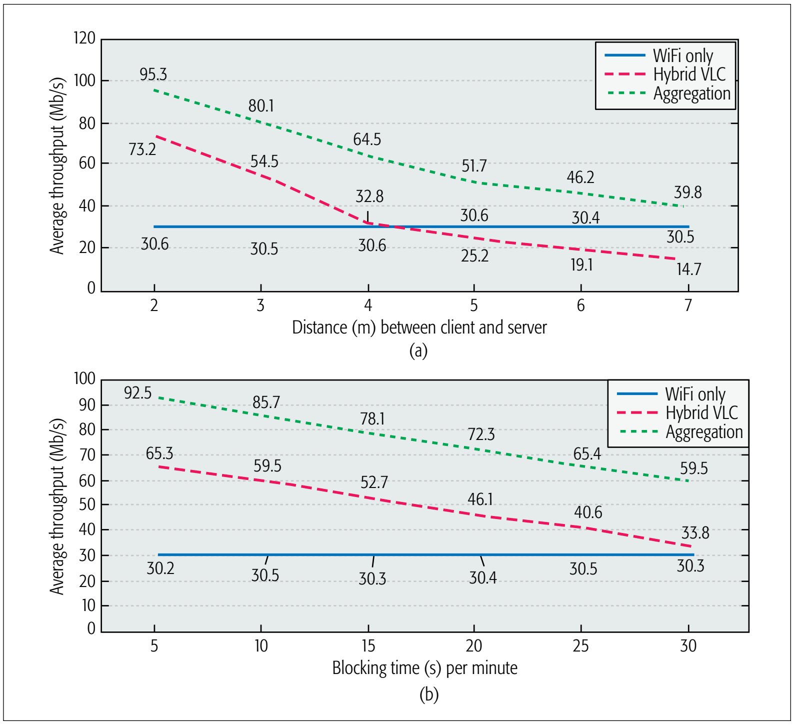


Figure 6. a) throughput vs. distance; b) throughput vs. blockage duration.

The aggregated system triples the achievable average throughput, and its lowest bound is higher than the average throughput of WiFi-only. Therefore, the aggregation technique not only enhances the available integrated bandwidth, but also provides reliable network communication. Due the inherent short-range property of LiFi, much better performance can be reached close to the LiFi AP for individual users. Note also that LiFi and WiFi users can be served in parallel inside and outside this limited coverage area.

Considering that mobile devices can have irregular movements, LiFi channel blockage can be a significant aspect that is mitigated by the hybrid solution. Figure 6b shows the average throughput achieved by the three systems with the variation of periods in which the LiFi link was blocked from 5 s to 30 s per minute. The distances between the WiFi and LiFi frontends are both set to 2 meters. It is observed that even if the LiFi link is blocked 50 percent of the time, Due the inherent short-range property of LiFi, much better performance can be reached close to the LiFi AP for individual users. Note also that LiFi and WiFi users can be served in parallel in and outside this limited coverage area.

while the user is moving, the hybrid system outperforms the WiFi-only system.

Future Research Opportunities

Based on our experience with the proof-of-concept system, there are considerable opportunities in future work in this area. In this section we outline an agenda for the combined Li+WiFi approach proposed in this article.

First, both technologies will experience further evolution to higher data rates. LiFi allows Gb/s throughput using higher bandwidth, monochromatic LEDs or lasers together with wavelength-division multiplexing as well as MIMO. WiFi is currently also upgraded by using more antennas and more bandwidth.

Besides unlicensed WiFi APs, research is needed to explore potential effects of LiFi data offloading when licensed indoor femtocells and outdoor macrocells are included in the system. The obtained results will yield a complete picture and offer first insights into a practical multi-tiered HetNet under practical illumination constraints (e.g. meeting lighting standards for office lighting) [6]. A proper system design must carefully consider the unique illumination qualities and services of individual spaces and applications to achieve the best compromise between VLC performance and illumination needs.

Another opportunity is to study the coexistence and further evolution of CSMA/CA and OFDMA in the proposed HetNet, including closed-loop link adaptation envisioned for both LiFi and enhanced WiFi networks. It is important to manage proportional fairness among the users, meaning that each of N users would get a constant fraction of the bandwidth when being alone in a combination of both LiFi and WiFi channels [13].

Channel aggregation of Li+WiFi is another interesting challenge. Two models are of interest:

- Aggregating channels from one access technology
- Aggregating channels from different access technologies

These can include multiple channels within either RF or optical spectrum [8]. Both approaches can be implemented on different layers of the OSI reference model ranging from the data link to the application layer. Relying on higher layers requires modifying both the client and server sides. Aggregation at lower layers must remain compatible with higher-layer protocols such as TCP, otherwise cross-layer aggregation must be achieved.

User mobility is also an important consideration for the provision of seamless connectivity and is required in order to properly evaluate the performance of the proposed Li+WiFi network. Physical layer (PHY) techniques can be used to enhance the performance of Li+WiFi in multi-user scenarios. For example, user separation can be performed by assigning separate color clusters to the users analogous to frequency reuse or subcarrier isolation in RF-cellular systems. One strategy is to leverage difference color shift keying (CSK) triplets in neighboring cells under the IEEE 802.15.7 model. A multi-color enabled VLC receiver allows separation of the individual channels in the color domain using a filtering technology. Optimized multi-color multi-user MIMO solutions based on the hybrid nature of the Li+WiFi network are not well investigated. UD battery drain and the impact of the user population and density on performance, while maintaining target illumination, are important research problems.

Finally, there is further need for experimental measurements to provide insights into the practical deployment of Li+WiFi networks and to attract industry interest in the most promising solutions. Therefore, a testbed is needed to investigate and realize Li+WiFi networks using different configurations and to evaluate the most promising solutions and algorithms for the integration. The fact that high-speed VLC frontends using existing baseband processing solutions are already available allows for early experiments also at the higher protocol layers that combine WiFi and LiFi with increasing sophistication [8, 9]. Of course, the available optical frontends need further development. Investigating the use of multiple colors and of fully software-defined digital signal processing will allow intervention at all protocol layers. There is a great deal of research opportunity for heterogeneous Li+WiFi networks.

Conclusion

The coexistence between WiFi and LiFi is a new promising research area. We have discussed the primary characteristics of both technologies and the possibility for them to coexist. We have demonstrated that a close integration of both technologies enables off-loading opportunities for the WiFi network to free resources for more mobile users because stationary users will preferably be served by LiFi. In this way, LiFi and WiFi can efficiently collaborate. We have implemented several ways of channel aggregation for the suggested coexistence, and demonstrated by proof-of-concept results, using state-of-the-art LiFi and WiFi frontends, that both technologies together can more than triple the throughput for individual users and offer significant synergies, yielding a combined solution that can adequately address the need for enhanced indoor coverage with the highest data rates needed in the 5th generation of mobile networks (5G). Finally, we have outlined a roadmap for future research opportunities toward the integration of both technologies.

Acknowledgments

The authors are grateful for partial support by NSF grant ECCS-1331018, the Engineering Research Centers Program of the National Science Foundation under NSF Cooperative Agreement No. EEC-0812056, and the German Ministry for Education and Research (BMBF) for the support of the collaborative project OWI-CELLS grant 16KIS0199K.

References

[1] J. G. Andrews *et al.*, "Femtocells: Past, Present, and Future," *IEEE JSAC*, vol. 30, no. 3, 2012, pp. 497–508.

[2] J. Kim and I. Lee, "802.11 WLAN: History, and Enabling MIMO Techniques for Next Generation Standards," *IEEE Commun. Mag.*, vol. 53, no. 3, 2015, pp. 134–40.

[3] V. Jungnickel *et al.*, "The Role of Small Cells, Coordinated Multipoint, and Massive MIMO in 5G," *IEEE Commun. Mag.*, vol. 52, no. 5, 2014, pp. 44–51.

[4] W. Guo *et al*., "A Parallel Transmission MAC Protocol in Hybrid VLC-RF Network," *J. Commun.*, vol. 10, no. 1, 2015.

[5] S. Wu, H. Wang, and C.-H. Youn, "Visible Light Communications for 5G Wireless Networking Systems: From Fixed to Mobile Communications," *IEEE Network*, vol. 28, no. 6, 2014, pp. 41–45.

[6] M. Rahaim, A. Vegni, and T. Little, "A Hybrid Radio Frequency and Broadcast Visible Light," *Proc. GLOBECOM Wksps.*, 2011.

[7] Z. Huang and Y. Ji, "Design and Demonstration of Room Division Multiplexing-Based Hybrid VLC Network," *Chinese Optics Lett.*, vol. 11, no. 6, 2013, pp. 1671–7694.

[8] S. Shao *et al*., "An Indoor Hybrid WiFi-VLC Internet Access System," *Proc. Wksp. CellulAR Traffic Offloading to Opportunistic Networks (CARTOON)*, Philadelphia, 2014.

[9] S. Shao *et al.*, "Design and Analysis of a Visible-Light-Communication Enhanced WiFi System," *OSA/IEEE J. Optical Commun. Netw. (JOCN)*, vol. 7, no. 10, 2015, pp. 960–73.

[10] P. M. Butala, H. Elgala, and T. Little, "SVD-VLC: A Novel Capacity Maximizing VLC MIMO System Architecture under Illumination Constraints," *Proc. 4th IEEE Wksp. Optical Wireless Commun.*, Atlanta, 2013.

[11] K. Kim and S. Kim, "Wireless Visible Light Communication Technology using Optical Beamforming," *Lasers, Fiber Optics, and Commun.*, vol. 52, no. 10, 2013, pp. 1–6.

[12] R. Nishioka *et al*., "A Camera and LED-Based Medium Access Control Scheme for Wireless LANs," *IEICE Trans. Commun*., vols. E98-B, no. 5, 2015, pp. 917-92.

[13] X. Bao *et al*., "Protocol Design and Capacity Analysis in Hybrid Network of Visible Light Communication and OFDMA Systems," *IEEE Trans. Vehic. Tech.*, vol. 63, no. 4, 2014, pp. 1770–78.

[14] H. Chowdhury and M. Katz, "Cooperative Data Download on the Move in Indoor Hybrid (Radio-Optical) WLAN-VLC Hotspot Coverage," *Trans. Emerging Telecommun. Technologies*, vol. 25, no. 6, 2014, pp. 666–77.

[15] L. Grobe *et al*., "High-Speed Visible Light Communication Systems," *IEEE Commun. Mag.*, vol. 51, no. 12, 2013.

Biographies

Moussa Ayyash [SM] (mayyash@csu.edu) is an associate professor in the Department of Information Studies at Chicago State University. He is the Director of the Center of Information and National Security Education and Research. He received his B.Sc. in electrical and computer engineering (ECE) from Mu'tah University, his M.Sc. in ECE from the University of Jordan, and a Ph.D. in ECE from IIT/Chicago. He is a member of the ACM.

Hany Elgala (helgala@albany.edu) is an assistant professor in the Computer Engineering Department, at the University of Albany–State University of New York (SUNY). Before moving to SUNY he was a research professor at Boston University and the Communications Testbed leader at the National Science Foundation Smart Lighting Engineering Research Center. His research focuses on visible light communications (VLC) or LiFi, wireless networking, and embedded systems. He is a member of the IEEE and IEEE Communications Society.

Abdallah Khreishah (abdallah@njit.edu) is an assistant professor in the Department of ECE at NJIT. His research interests are in the areas of visible light communications, green networking, network coding, wireless networks, and network security. He received his B.S. degree in computer engineering from Jordan University of Science and Technology in 2004, and his M.S. and Ph.D. degrees in ECE from Purdue University in 2006 and 2010, respectively. He is the chair of North Jersey IEEE EMBS chapter.

Volker Jungnickel (volker.jungnickel@hhi.fraunhofer.de) received doctorate and habilitation degrees from Humboldt University in 1995 and Technical University in 2015, respectively, both in Berlin. In 1997 he joined the Fraunhofer Heinrich Hertz Institute, where he is leading the metro, access, and in-house systems group. Volker contributed to high-speed optical wireless links, a first 1 Gb/s mobile radio link, the first real-time trials of LTE and the first coordinated multipoint trials. He has contributed to 180 papers, 10 books, and 25 patents.

Thomas DC Little (tdcl@bu.edu) is a professor of electrical and computer engineering in the College of Engineering at Boston University. He is also an associate director and principal investigator for the National Science Foundation Smart Lighting Engineering Research Center. Little received his B.S. degree in biomedical engineering from RPI in 1983, and his M.S. degree in electrical engineering and Ph.D. degree in computer engineering from Syracuse University in 1989 and 1991, respectively.

Sihua Shao [S] (ss2536@njit.edu) is a Ph.D. student in the Department of Electrical and Computer Engineering at New Jersey Institute of Technology. His current research interests include wireless communication, visible light communication, and heterogeneous networks. He received his B.S. degree in electrical and information engineering from South China University of Technology in 2011, and his M.S. degree in electrical and information engineering from Hong Kong Polytechnic University in 2012.

Michael Rahaim (mrahaim@bu.edu) is a postdoctoral researcher in the Department of Electrical and Computer Engineering at Boston University, working with the NSF funded Smart Lighting Engineering Research Center. His research focuses on software defined radio, visible light communication, HetNets, and smart lighting. He received his B.S. in electrical and computer systems engineering from Rensselaer Polytechnic Institute in 2007, and his M.S. and Ph.D. in computer engineering from Boston University in 2011 and 2015, respectively.

Dominic Schulz (dominic.schulz@hhi.fraunhofer.de) received his MS in communications engineering from Berlin University of Applied Sciences in 2012. In 2013 he joined the Department of Photonic Networks and Systems at Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute. He is currently working toward his Ph.D. in the field of optical wireless communications. His current activities include the development of high data rate systems for wireless access, as well as research toward long-range links.

Jonas Hilt (jonas.hilt@hhi.fraunhofer.de) received his diploma in electrical engineering from Berlin University of Applied Sciences in 2009. In 2009 he joined the Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, where he is an electrical engineer in the department of photonic networks and systems. His current activities are in the area of design and realization of embedded systems and visible light communication systems (VLC).

Ronald Freund (ronald.freund@hhi.fraunhofer.de) received the Dipl.-Ing. degree and the Dr.-Ing. degree in electrical engineering from Technical University of Ilmenau (TUI), in 1993 and 2002, respectively. In 2013 he received an MBA degree from RWTH Aachen. Since 1995 he has been with the Heinrich Hertz Institute in Berlin, where he is currently leading the Department of Photonic Networks and Systems.

The available optical frontends need further development. Investigating the use of multiple colors and of fully software-defined digital signal processing will allow intervention at all protocol layers. There is a great deal of research opportunity for heterogeneous Li+WiFi networks.

Provisioning 1 Gb/s Symmetrical Services with Next-Generation Passive Optical Network Technologies

Rafael Sánchez, José Alberto Hernández, Julio Montalvo García, and David Larrabeiti

The authors focus on delivering symmetrical 1 Gb/s access to residential users with a target temporal guarantee at the least cost using next-generation PON technologies. They compare four NG-PON standard access technologies, GPON, XGPON, WDM-PON, and the emerging TWDM-PON, from technical and economic perspectives.

ABSTRACT

Service providers spend billions upgrading their broadband access networks to the latest access standards. Fiber has become the technology of choice in the medium and long term, thanks to its speed, reach, and future-proofness. A differential advantage of fiber over other broadband access technologies is that it makes it possible for operators to deliver symmetric-rate services. Most of today's commercial offers based on regular PON range from 10 to 100 Mb/s of committed information rate, and higher rates are advertised as peak rates with unspecified guarantees. In this article we focus on delivering symmetrical 1 Gb/s access to residential users with a target temporal guarantee at the least cost using next-generation PON technologies. We compare four NG-PON standard access technologies, GPON, XGPON, WDM-PON, and the emerging TWDM-PON, from technical and economic perspectives. The study shows that if a service provider wants to keep up with the growing user traffic pattern in the long run, only TWDM-PON can provide 1 Gb/s nearly guaranteed at a moderate cost with respect to the fully dedicated 1 Gb/s point-to-point connection available in WDM-PON technologies.

INTRODUCTION

At present, 1 Gb/s downstream Internet access services are offered by some service providers in the United States, Europe, and Asia even though the number of fiber subscribers (12.4, 22, and 93 million, respectively), fiber market maturity, and penetration rate (10, 10–50, and 45–70 percent, respectively) are very different across the continents [1]. This 1 Gb/s service is being offered as a peak data rate with different levels of guarantee in addition to a minimum multi-megabit-per-second committed information rate. While basic services may not require such a high rate, other factors like user experience enhancement, the increasing amount of connected devices at home, and low latency requirements for interactive gaming and other coming applications (UHD 3D immersive gaming and video conferencing, cloud computing, infrastructure as a service, etc) are expected to boost the demand for symmetric 1 Gb/s access capacity with certain quality of service (QoS) guarantees in the near future.

Deploying 1 Gb/s symmetrical services with optical fiber is expensive due to the high investment costs associated with civil works. Some service providers may opt to take maximum advantage of their existing twisted-pair copper infrastructure in the design. This strategy leads to fiber to the cabinet (FTTC) and fiber to the node (FTTN) deployments, combining fiber with very high rate digital subscriber line version 2 (VDSL2) [2]. However, this configuration also involves costs of installation, powering, and maintenance of intermediate active devices, as well as additional delay, and hence, installing fibers up to the customer premises, either residential or business (FTTH/FTTB), seems to be the best long-run approach to keep up with bandwidth and latency requirements of future applications.

There is passive optical network (PON) technology available to provide 1 Gb/s services to end users, and a number of next-generation PON (NG-PON) standards to be completed very soon. This article aims to compare gigabit PON (GPON), XGPON, and wavelength-division multiplexing (WDM)-PON standards with the new time-shared WDM (TWDM)-PON approaches concerning the provisioning of 1 Gb/s symmetrical connectivity to residential customers. Such a comparison addresses both technological and economic aspects, with the aim to provide a reference for network operators willing to migrate to the next-generation access services. A number of questions are investigated throughout this article: Which kind of FTTH technology is most suitable to offer 1 Gb/s symmetrical services? Can 1 Gb/s be guaranteed 100 percent of the time? Which parameters must be considered in the network design? What is the cost per user associated with each technology? A greenfield scenario deployment of a dense area with 5000 users is considered to answer all these questions.

This article is organized as follows. The following section provides a taxonomy of FTTH access protocols that are capable of supporting 1 Gb/s symmetrical services. After that, we quickly review the basic methodology used in capacity planning with oversubscription, often used by network operators. Then we make a technical and economic comparison of four access protocols including capital expenditures (CAPEX) and operational expenditures (OPEX). The final

The views expressed in this article are those of the authors and do not reflect the opinion of the authors' employers, specifically in terms of architectural and engineering design or experience.

Rafael Sánchez is with Google Inc.; José Alberto Hernández and David Larrabeiti are with Universidad Carlos III de Madrid; Julio Montalvo is with Telefónica Investigación y Desarrollo.

0163-6804/16/$25.00 © 2016 IEEE

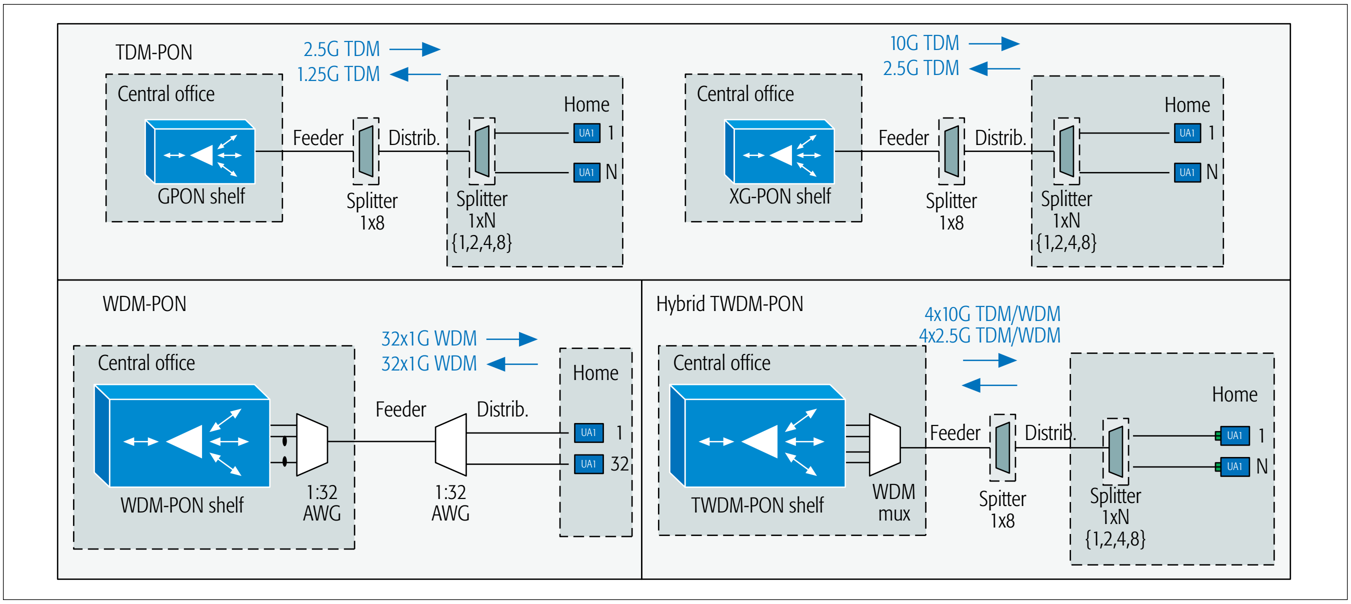


Figure 1. Taxonomy of PON fiber access protocols.

section concludes this article with a summary of its main results along with future work worth investigation.

Taxonomy of Fiber Access Protocols to Provide 1 Gb/s Symmetrical Services

According to the FTTH Council [3], an *access protocol* is "*a method of communication used by the equipment located at the ends of the optical paths to ensure reliable and effective transmission and reception of information over the optical paths*." The physical fiber topology that connects the operator's premises and subscriber's premises, also called an optical distribution network (ODN), can be point-to-point, point-to-multipoint (often referred to as PON), or ring, although hybrid ring-tree topologies can also be found in relevant research works [4]. The design of access protocols is conditioned by the type of underlying topology. This article focuses on protocols for PON topologies, currently the most widely deployed. Figure 1 shows a taxonomy of the PON fiber access protocols under study: time-division multiplexing (TDM-PON), WDM-PON, and a hybrid version, TWDM-PON.

TDM-PON

This technology uses a shared point-to-multipoint approach with one or two wavelengths in the downstream direction (from a central office, CO, to users) and one wavelength in the upstream (from users to a CO). TDM-PON uses a 1:*N* passive splitter/combiner to divide the optical signal among all users in the downstream direction and aggregate the users' data in the upstream direction. The optical line terminal (OLT) uses a dynamic bandwidth assignment (DBA) algorithm to arbitrate access to the shared channel in the upstream direction, avoid collisions, assign bandwidth to the users, and provide QoS for different types of flows.

For example, GPON (International Telecommunication Union — Telecommunication Standardization Sector, ITU-T, G.984) uses the 1490 nm wavelength at 2.5 Gb/s for downstream data traffic (optionally, the 1550 nm wavelength can be used to carry RF video separately), and the 1310 nm wavelength at 1.25 Gb/s for upstream traffic. Recent enhancements like XG-PON (ITU-T G.987) offer 10G/2.5G in the down- and upstream direction, respectively. Besides, there are also symmetrical TDM-PON standards like 2.5G/2.5G GPON or 10G/10G (XG-PON2), but these are not considered in this article due to the lack of deployments.

WDM-PON

In this case, a single wavelength is redirected to an end user from the central office via a passive wavelength router located in the outside plant (OSP). In this case, the power splitter/combiner is replaced by a wavelength selective filter, usually an array waveguide grating (AWG), thus setting up a single wavelength with symmetric bandwidth between each user and the central office. Unlike TDM-PON, WDM-PON provides a *dedicated* point-to-point connection between users and the CO, that is, there is no bandwidth sharing between users. Advantages of WDM- over TDM-PON are scalable bandwidth, long reach (given the low insertion loss of filters, optional amplification), troubleshooting [5], security (users do not see other users' traffic), and the possibility to individually adapt bit rates on a per-wavelength basis.

There are several flavors of WDM-PON technologies available in the market, each with a different implementation technology: injection-locking, tunable lasers, wavelength reuse, and coherent detection [6]. This article considers only the AWG-based injection-locking WDM-PON flavor with 1:32 splitting ratio (1:64 AWG still not commercially available), specified in the standard ITU-T G.698.3.

TWDM-PON

This technology was selected as the primary solution for the NG-PON stage 2 (NG-PON2) project of the Full Service Access Network (FSAN) community, and is currently standard-

	GPON	XG-PON	TWDM-PON	WDM-PON
Standard	ITU-T G.984	ITU-T G.987	ITU-T G.989	ITU-T G.698.3
Availability	In market	In market	In trial	In market
Feeder rate (C_{DL}/C_{UL})	2.5G/1.25G	10G/2.5G	40G/10G	32G/32G
Security	No	No	No	Yes
Outside Plant	Splitter	Splitter	Splitter with WDM mux	AWG
Price	Lower	Medium	Medium	Higher
Power budget (dB)	28 (B+)	35 (E2)	38.5	15

Table 1. Summary of features for PON technologies.

ized (ITU-T G.989 series, completed in October 2015). TWDM-PON takes one step forward with respect to XGPON, leveraging the research and development effort of the PON industry on this technology. Essentially, TWDM-PON increases the aggregate PON rate by stacking multiple XGPONs on different pairs of wavelengths, which yields an aggregate $N \times 10$ Gb/s downstream and $N \times 2.5$ Gb/s upstream. In a prototype shown in [7], $N = 4$, and each TWDM-PON optical network unit (ONU) is equipped with colorless transmitters and receivers operating at 10 Gb/s downstream and 2.5 Gb/s upstream. As in TDM-PONs, bandwidth is shared across several subscribers. This solution is called hybrid since it combines the flexibility of TDM-PONs with the increased capacity of WDM technology.

The advantages of TWDM-PON over pure WDM-PON are its high fanout and "graceful evolution" capability, since it is compatible with older TDM-PON versions, like GPON and XGPON, allowing coexistence within the same ODN.

Table 1 provides a summary of the main features of the four PON technologies under study. Based on a number of real deployments and market interest, the next section studies the suitability of GPON, XGPON, AWG-based WDM-PON, and TWDM-PON to provide 1 Gb/s symmetrical services to residential customers. Such suitability is quantified from both technical and economic perspectives in a hypothetical green field deployment.

Capacity Planning

This section considers the capacity planning for each PON branch, following the architecture of Fig. 1. As noted, GPON, XGPON, and TWDM-PONs have a first fixed splitting stage, 1:8, and a second one, 1:N, that can be configured ($N \in \{1, 2, 4, 8\}$). This section studies how many users can coexist on the same PON branch sharing its bandwidth so that they experience 1 Gb/s symmetrical service most of the time. The analysis is performed only for the uplink direction since it is a more limiting factor than the downlink case.

GPON, XG-PON, and TWDM-PON with Oversubscription

Most packet-switched telecommunication services rely on the concept of oversubscription; the access network is not an exception. Capacity planning based on oversubscription works because of the empirical observation that only a small portion of subscribers are simultaneously active at a given random instant [8, 9]. Network designers leverage this fact to provide access to a large number of users at a moderate expense of resources. Essentially, the $b_{peak} = 1$ Gb/s bandwidth cannot be guaranteed to all users during 100 percent of the time, but only a portion of it.

Now, let n_{tot} refer to the maximum number of users physically attached to the same PON branch. As noted from Fig. 1, the total number of users can take the values $n_{tot} \in \{8, 16, 32, 64\}$ depending on the second splitting stage. This range of n_{tot} only applies to GPON, XGPON, and TWDM-PON technologies since for WDM-PON deployments, we consider $n_{tot} = 32$ fixed (Fig. 1).

Let n_{act} refer to the random variable that considers the number of active users at a given random time. Clearly, $0 \le n_{act} \le n_{tot}$. For simplicity, we consider that every user can be active with probability q, and that all users are uncorrelated and have the same behavior, that is, they are active with probability q or idle with probability $1 - q$. In other words, n_{act} follows a binomial distribution, $n_{act} \sim B(n_{tot}, q)$.[1] As observed in many measurement studies, the value of q is very small for residential users.

Concerning bandwidth, let us define b as the rate observed per individual user in the PON branch, as follows:

$$b = \frac{C_{UL}}{n_{act}}$$

where C_{UL} is the upstream capacity of each NG-PON technology (Table 1). Clearly, b is a discrete random variable that depends on the number of active users: the higher the value of n_{act}, the lower the bandwidth rate experienced per user. In addition, network operators can limit the bandwidth rate experienced by users to b_{peak} when the number of active users is small (i.e., when $b > b_{peak}$). On the contrary, when all users are active ($n_{act} = n_{tot}$), all users are guaranteed at least a minimum rate of $(C_{UL})/(n_{tot})$. In light of this, the random variables b and n_{act} are related as follows:

$$P\left(b \ge \frac{C_{UL}}{k}\right) = P(n_{act} < k), \text{ with } n_{act} \sim B(n_{tot}, q) \tag{1}$$

meaning that, when k users are active, the uplink capacity C_{UL} is equally shared among them.

In general, it is very unlikely to have many active users when q is sufficiently small. This allows network operators to leverage statistical multiplexing gains. Network designers often use the term *oversubscription ratio* o to refer to the maximum carried traffic divided by the maximum bandwidth capacity *promised* to the users, in other words:

$$o = \frac{C_{UL}}{n_{tot} b_{peak}}$$

Finally, let β refer to the probability that b_{peak} is guaranteed to the users in the oversubscription model. Clearly, b_{peak} is guaranteed when no more than $n_{act}^{(max)}$ users are active, namely:

$$n_{act}^{(max)} = \left\lfloor \frac{C_{UL}}{b_{peak}} \right\rfloor \tag{2}$$

[1] It is worth remarking that the probability density function (PDF) of the binomial distribution $B(n, q)$ follows: $P(X = k) = \binom{n}{k} q^k (1 - q)^{n-k}$, $k = 0, 1, \ldots, n$.

Thus, β equals the probability that no more than $n_{act}^{(max)}$ users are simultaneously active; in other words:

$$\beta = P(n_{act} \le n_{act}^{(max)}).$$

Thanks to the properties of the binomial distribution, β can also be thought of as the percentage of time in which b_{peak} is guaranteed.

Numerical Example and Analytical Results

Consider a GPON ($C_{UL}^{(GPON)}$ = 1.25 Gb/s) with q = 0.15 (i.e., 15 percent activity per user) and n_{tot} = 32 users, that is, the second splitting stage is 1:4. First of all, the maximum number of active users in order to guarantee b_{peak} = 1 Gb/s is $n_{act}^{(max)}$ = 1 user, that means, one active user at most (two active users would share 1.25 Gb/s). Following the Binomial distribution, the average number of active users is: $E(n_{act}) = n_{tot}q = 4.8$ users, and the average bandwidth is $E(b) = 327$ Mb/s.[2]

In the unlikely event that all users are active, that is, $n_{act} = n_{tot}$, which occurs with probability

$$P(n_{act} = 32) = q^{32} = 4.3 \cdot 10^{-27},$$

the bandwidth experienced per active user is only b = 39 Mb/s. This is the minimum absolute guaranteed bandwidth 100 percent of the time.

Now, since most users are idle most of the time, the next stage is to see the probability that only $n_{act}^{(max)}$ = 1 user is active in the PON branch, thus receiving b_{peak} bandwidth. Following the binomial distribution, the probability of having 1 active user or less in the PON is only 3.7 percent.

Now, consider that the operator's requirement is that all users must receive b_{peak} = 1 Gb/s at least β = 20 percent of the time. Then the value of n_{tot} can be no larger than 18 total users, since $P(n_{act} \le 1) = 0.22$ when $n_{act} \sim B(n_{tot} = 18, q = 0.15)$ but $P(n_{act} \le 1) = 0.198$ when $n_{act} \sim B(n_{tot} = 19, q = 0.15)$. Since $n_{tot} \le 18$, the maximum split ratio in the second stage must be at most 1:2 ($n_{tot} = 8 \times 2 = 16$ total users per PON branch). In this case, the average bandwidth experienced by users is now $E(b)$ = 637 Mb/s.

In the case of XG-PON, when $C_{UL}^{(XG\text{-}PON)}$ = 2.5 Gb/s, b_{peak} = 1 Gb/s is guaranteed when there are no more than $n_{act}^{(max)}$ = 2 active users in the PON branch. For the same β = 20 percent criteria as before and q = 15 percent, the maximum number of users in the PON branch rises to $n_{tot} \le 27$. Again, the maximum split in the second stage is 1:2 (16 users at most), which yields an average bandwidth rate $E(b)$ = 1.27 Gb/s, limited to b_{peak} = 1 Gb/s.

Figure 2 shows the cumulative distribution function (CDF) of b for GPON with different split ratios (Eq. 1) along with the average bandwidth rate $E(b)$. As shown, cases 1:64 and 1:32 provide very small percentages where 1 Gb/s is guaranteed (3.67 and 0.04 percent, respectively) and small values of average bandwidth.

Furthermore, Table 2 shows the average rate $E(b)$ observed and the percentages of time β where b_{peak} is guaranteed for all NG-PON technologies and different split ratios. The values of TWDM-PON have been computed taking into account that a stack of four XG-PON technologies is shared among n_{tot} users. In other words, we have computed the $E(b)$ and b values for an XG-PON with $(n_{tot})/4$ users.

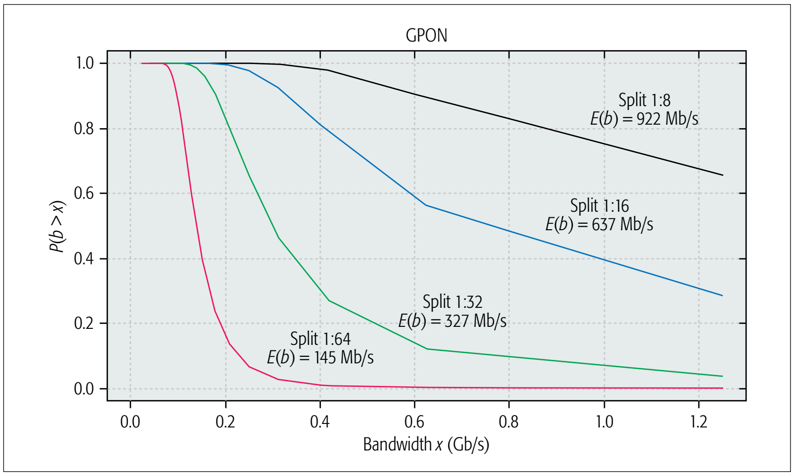


Figure 2. GPON: CDF of b and average bandwidth for different split ratios, q = 15 percent.

When q = 15 percent, XG-PON significantly improves the results of GPON providing 1 Gb/s rate at least 50 percent of the time for the split ratios 1:8 and 1:16. TWDM-PON provides 1 Gb/s most of the time for split ratios 1:32 and below. When large user activity periods are expected (e.g. q = 50 percent), only TWDM-PON with 1:8 and 1:16 split ratios can provide 1 Gb/s bandwidth for a substantial percentage of time.

Finally, it is worth remarking that WDM-PON provides a dedicated point-to-point connection between each user and the OLT with 1 Gb/s guaranteed 100 percent of the time for n_{tot} = 32 users regardless of user activity q.

Economic Study for an Urban Area

This section studies the total cost of ownership (TCO), including both CAPEX and OPEX, required for the deployment of a hypothetical green field urban scenario with 5000 users. Only those FTTH technologies capable of achieving 1 Gb/s symmetrically for a minimum of β = 20 percent of the time have been considered (q = 15 percent assumed). For example, GPON 1:16 is selected because it achieves 28.4 percent (higher than 20 percent), while XGPON 1:32 only achieves 12.2 percent (lower than 20 percent) and therefore is not considered (Table 2). Oversubscription factors beyond the feeder fiber (i.e. from the OLT toward the metro) are not considered.

The calculus of CAPEX is based on commercial prices available from selected undisclosed vendors, complemented with pricing information and network considerations from [10]. Cost of equipment not commercially available yet (TWDM-PON) is derived from market costs of components. Figure 3 shows the resulting cost per user in such a green field deployment relative to the cost per user of the most expensive technology, WDM-PON in this case. The cost includes the following factors.

Central Office: The cost of core cards of the OLT shelves, one-time software licenses, and everything necessary for in-service operation. The cost of uplink transceivers, which is dependent on split ratio, packet loss, and demand distribution, is not included. The reader can find

[2] The average bandwidth rate perceived by the users is computed as

$$E(b) = \frac{\sum_{k=1}^{n_{tot}} \frac{c}{k} P(n_{act} = k)}{1 - P(n_{act} = 0)}$$

which weights the rate perceived by the users (for the cases where at least one user is active) multiplied by their probability.

	1:8	1:16	1:32	1:64
$E(b)$, β	$q = 15$ %			
GPON	922 Mb/s, 65.7%	637 Mb/s, 28.4%	327 Mb/s, 3.7%	145 Mb/s, 0.04%
XGPON	1000 Mb/s, 89.5%	1000 Mb/s, 56.1%	654 Mb/s, 12.2%	290 Mb/s, 0.2%
TWDM	1000 Mb/s, ~100%	1000 Mb/s, 98.8%	1000 Mb/s, 89.5%	1000 Mb/s, 56.1%
WDM-PON	–	–	1000 Mb/s, 100%	–
$E(b)$, β	$q = 50$ %			
GPON	369 Mb/s, 3.5%	168 Mb/s, ~0%	80 Mb/s, ~0%	40 Mb/s, ~0%
XGPON	738 Mb/s, 14.5%	337 Mb/s, ~0%	162 Mb/s, ~0%	79 Mb/s, ~0%
TWDM	1000 Mb/s, ~100%	1000 Mb/s, 68.7%	738 Mb/s, 14.5%	337 Mb/s, ~0%
WDM-PON	–	–	1000 Mb/s, 100%	–

Table 2. Bandwidth comparison between the four NG-PON technologies: average bandwidth and percentage of time where b_{peak} =1 Gb/s is guaranteed.

an uplink analysis based on Monte Carlo simulations in [9]. For TWDM-PON, the cost of the WDM mux is also included here.

OLT: The cost of OLT line cards for each technology. OLT line cards are equipped with 16 ports for GPON, 4 ports for XG-PON/TWDM-PON, and 1 port for WDM-PON.

ONT: The lowest cost of commercially available units equipped with at least four Gigabit Ethernet ports toward the user. In the case of a TWDM-PON ONT, since it is not commercially available yet, the cost is derived from market costs of components of the product.

Passive-Street Cabinet: The cost of the cabinet, splitters, or AWG where appropriate (both first and second stage costs are included), and the cost of splicing the fibers. For GPON, XGPON, and TWDM-PON, the two-stage splitting architecture of Fig. 1 is considered, following [10]. That is, a first fixed 1:8 split stage, placed at the street cabinet, followed by a second variable split stage (1:1, 1:2, 1:4, and 1:8), which is placed at the bottom of the building. In the case of WDM-PON, a 1:32 AWG is assumed and is located at the street cabinet;

Feeder and Distribution Segment: The cost of digging and preparing the trench, manholes, and finer deployment in each segment. As seen in Fig. 1, for GPON, XGPON, and TWDM-PON, feeder fiber is the fiber between the CO and the first 1:8 split, and distribution fiber is the fiber between the first and second splits; while for WDM-PON, feeder fiber is the fiber between the CO and the 1:32 AWG, and distribution fiber is the fiber between the AWG and the ONT. Following [10], the length of the feeder segment in an urban area is assumed to be 850 m, whereas the length of the distribution segment is 80 m. Cost of digging and preparing the trench for an urban area has been assumed USD 120/m;

In-House Segment: The cost of the optical distribution frame (ODF), patch cable, and fiber access terminal in the basement.

Concerning OPEX, only first-year costs are considered, including system support and energy consumption, as a markup of the active (4 percent) and passive (1 percent) infrastructure [10, 11]. Since they are considered as a percentage, OPEX costs are uniformly distributed over the CAPEX costs. System support considers the technical and maintenance support required for the installed equipment. Energy consumption represents the yearly cost of energy (in watts) consumed by the equipments.

As expected, the largest part of the CAPEX lies in the physical infrastructure [12] (in-house segment, distribution segment, street cabinet, feeder segment, and CO), which represents between 50 and 80 percent of the total investment. All technologies under consideration are deployed with a single fiber in the feeder segment, and a single fiber between the remote node and the ONT. Thus, the main difference in terms of TCO corresponds to the CO, OLT,passive-street cabinet (splitter or AWG), and ONT.

Other observations include:

- The shared cost of the OLT should decrease as the split ratio increases. However, in all TDM-PON and hybrid options, the TCO for 1:16 is slightly more expensive than in the 1:8 case. This arises as a penalty for choosing a fixed 1:8 first stage, which means that extra 1:2, 1:4, and 1:8 splitters have to be dimensioned for higher split ratios.
- ONTs are cheaper in GPON due to electronics managing less bandwidth; XG-PON ONTs come next, followed by TWDM-PON and WDM-PON.
- The cost of the passive-street cabinet in TWDM-PON 1:16 is slightly higher than in WDM-PON. This is due to the assumption of a 1:8 split at the first stage for all TDM-PON and TWDM-PON technologies [10]. In this case, for example, for 32 users, a single 1:32 AWG is enough for WDM-PON but would require 4 × 1:8 in the first split + 8 1:2 for TWDM-PON 1:16, that is, although the cost of a 1:32 AWG is much more expensive than the cost of a single 1:16 power splitter, the topology under consideration actually compares 1 × 32 AWG against 12 (4 + 8) power splitters (in the case of TWDM-PON 1:16).
- GPON is the cheapest technology with 1:8 split ratio, and is capable of providing 1 Gb/s for a large portion of the time. However, GPON does not scale up when q increases (Table 2).
- The cost per user of XGPON 1:8 and 1:16 is very similar to TWDM-PON 1:32 and 1:64, respectively, and also provides very similar performance. This is a consequence of the fact that TWDM-PON stacks four XG-PONs.
- TWDM-PON with 1:8 and 1:16 split ratios provide 1 Gb/s nearly 100 percent of the time when q = 15 percent with a substantial cost reduction with respect to WDM-PON, which is the most expensive flavor. However, it is worth remarking that WDM-PON provides 1 Gb/s guaranteed 100 percent of the time regardless of user activity q.
- The high cost of WDM-PON is mainly due to the electronics at the OLT (one laser per user is required) and the lower shelf density (256 users per shelf). The OLT and CO costs dominate in this technology.

Summary and Discussion

This article has compared four different flavors of fiber access protocols capable of offering 1 Gb/s symmetrical services for residential users. In particular, GPON, XGPON, WDM-PON, and the emerging TWDM-PON technologies with dif-

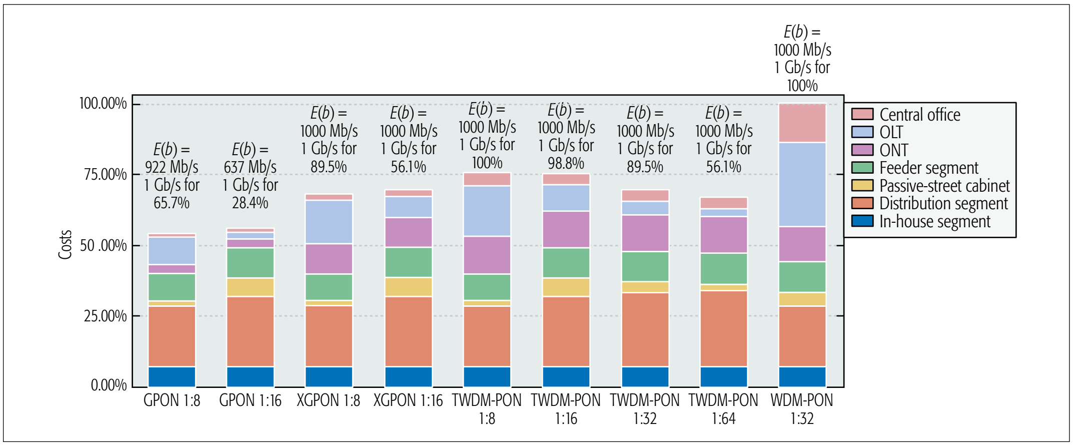


Figure 3. Details of TCO (CAPEX and OPEX) for FTTH options for providing 1 Gb/s symmetrical (bandwidth values for $q = 15$ percent).

The shared cost of the OLT should decrease as the split ratio increases. However, in all TDM-PON and hybrid options, the TCO for 1:16 is slightly more expensive than in the 1:8 case. This arises as a penalty for choosing a fixed 1:8 first stage, which causes that extra 1:2, 1:4 and 1:8 splitters have to be dimensioned for higher split ratios.

ferent split ratios have been analyzed for a green field deployment of 5000 users in a typical urban area. Market prices of either available commercial equipment (GPON, XGPON, and WDM-PON) or prototypes (TWDM-PON) have been used.

The results show that GPON 1:8 and 1:16, XGPON 1:8 and 1:16, TWDM-PON, and WDM-PON are good candidates to enable 1 Gb/s symmetrical services for residential users in terms of both cost and performance for next-generation optical access. However, as the user activity pattern increases, both GPON and XGPON will become insufficient. Only TWDM-PON and WDM-PON can guarantee 1 Gb/s at high levels of user activity (for a fraction of time typically used in design today in the case of TWDM-PON).

Other services than residential (business services and wireless backhaul), which require higher bandwidth, lower latency and physical separation of traffic (for security purposes) than residential scenarios, may require the use of dedicated point-to-point connectivity with absolute bandwidth guarantees, in other words, WDM-PON. In light of this, WDM-PONs with bandwidth provisioning beyond 1 Gb/s have been proposed [13], some supporting up to 10 Gb/s.

Acknowledgments

The authors would like to acknowledge the support of the Spanish projects CRAMnet (grant no. TEC2012-38362-C03- 01) and TIGRE-5CM (grant no. S2013/ICE-2919) in the development of this work. In addition, part of the cost prices in the development of the economic models come from results of the EU FP7 project DISCUS (ICT-318137).

References

[1] A.-K. Hatt *et al.*, "Creating a Brighter Future," press conf. at FTTH Conf., Feb. 2014, White Paper.

[2] T. Rokkas, D. Katsianis, and D. Varoutas, "Techno-Economic Evaluation of FTTC/VDSL and FTTH Roll-Out Scenarios: Discounted Cash Flows and Real Option Valuation," *IEEE/OSA J. Opt. Commun. and Net.*, vol. 2, no. 9, Sept. 2010, pp. 760–72.

[3] FTTH Council, "FTTH Council – Definitions of Terms, Version 4.0," White Paper, Feb. 2015.

[4] J. Prat *et al.*, "Results from EU Project Sardana on 10G Extended Reach WDM PONs," *Proc. OFC/NFOEC*, Mar. 2010, pp. 1–3.

[5] R. Sánchez, J. A. Hernández, and D. Larrabeiti, "Troubleshooting PON Networks Effectively with Carrier-Grade Ethernet and WDM-PON," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. S7–13.

[6] R. Huelsermann, K. Grobe, and D. Breuer, "Results from EU FP7 Project OASE on Next-Generation Optical Access," *Proc. Photonic Networks*, May 2013, pp. 1–8.

[7] Y. Luo *et al.*, "Time- and Wavelength-Division Multiplexed Passive Optical Network (TWDM-PON) for Next-Generation PON Stage 2 (NGPON2)," *IEEE/OSA J. Lightwave Tech.*, vol. 31, no. 4, 2013, pp. 587–93.

[8] J. Segarra, V. Sales, and J. Prat, "Access Services Availability and Traffic Forecast In PON Deployment," *Proc. J. Int'l. Conf. Transparent Optical Networks*, June 2011, pp. 1–6.

[9] S. Lambert *et al.*, "Energy Efficiency Analysis of High Speed Triple-Play Services in Next-Generation PON Deployments," *Computer Networks*, vol. 78, no. 0, 2015, Special Issue: Green Communications, pp. 68–82.

[10] J. Schneir and Y. Xiong, "Cost Analysis of Network Sharing in FTTH/PONs," *IEEE Commun. Mag.*, vol. 52, no. 8, Aug. 2014, pp. 126–34.

[11] L. Valcarenghi *et al.*, "Energy Efficiency in Passive Optical Networks: Where, When, and How?," *IEEE Network*, vol. 26, no. 2, Nov. 2012, pp. 61–68.

[12] T. Rokkas *et al.*, "Economics of Time and Wavelength Domain Multiplexed Passive Optical Networks," *IEEE/OSA J. Opt. Commun. Net.*, vol. 2, no. 12, Dec. 2010, pp. 1042–51.

[13] Z. Al-Qazwini and H. Kim, "10-Gbps Single-Feeder, Full-Duplex WDM-PON using Directly Modulated Laser and RSOA," *Proc. OFC/NFOEC*, Mar. 2012, pp. 1–3.

Biographies

Rafael Sánchez holds an M.Sc. (2008) from Universidad Carlos III de Madrid, Spain, and is a telecommunications engineer (1996) at Polytechnic University of Valencia. Since 1996, he has been involved in multiple networking projects in areas like optical networks (SDH/DWDM), IPTV, digital transmission, fiber access (WDM-PON), and carrier Ethernet in companies like Lucent Technologies, Nortel, and LG-Nortel. Currently, he works for Google in the enterprise division, and is pursuing a Ph.D. degree in telematic engineering at Universidad Carlos III.

José Alberto Hernández completed his five-year degree in telecommunications engineering at Universidad Carlos III de Madrid in 2002, and his Ph.D. degree in computer science at Loughborough University, Leicester, United Kingdom, in 2005. He has been a senior lecturer in the Department of Telematics Engineering at that university since 2010, where he combines teaching and research in the areas of optical WDM networks, next-generation access networks, metro Ethernet, energy efficiency, and hybrid optical-wireless technologies. He has published more than 75 articles in both journals and conference proceedings on these topics. He is a co-author of the book *Probabilistic Modes for Computer Networks: Tools and Solved Problems*.

Julio Montalvo (1979) holds an M.Sc. degree in telecommunication engineering (2003) from the Technical University of Madrid and a Ph.D. degree in electrical, electronics and robotics engineering (2008) from Universidad Carlos III de Madrid (Extraordinary Doctorate Prize). He is currently the author of one granted patent (no. 103868 at Portuguese INPI, 2007) and more than 40 refereed publications in scientific journals (14 in JCR) and conferences. He is the Editor and author of Chapter 2 of the book *Optical Transmission* (Springer). Currently he is with Telefónica Research & Development as a technology specialist in the Fixed Access and Home Network Direction (Global CTO). He is Telefónica's delegate in the Broadband Forum, ITU-T, and FSAN.

David Larrabeiti is a professor of switching and networking architectures at Universidad Carlos III of Madrid. Since 1990 he has participated in EU-funded research projects related to next-generation networks and protocols. In 2009–2010 he was a visiting researcher at Stanford University under Spanish mobility grant PR2009-0221. He is UC3M's principal investigator at the BONE network of excellence on optical networking. His current research interests include the design of hybrid electro-optical packet switches and multipoint optical communications.

Migration Strategies for FTTx Solutions Based on Active Optical Networks

Kun Wang, Anders Gavler, Carmen Mas Machuca, Lena Wosinska, Kjell Brunnström, and Jiajia Chen

AON, one of the most deployed fiber access solutions in Europe, needs to be upgraded in order to satisfy the ever growing bandwidth demand driven by new applications and services. Meanwhile, network providers want to reduce both capital expenditures and operational expenditures to ensure that there is profit coming from their investments.

Abstract

AON, one of the most deployed fiber access solutions in Europe, needs to be upgraded in order to satisfy the ever growing bandwidth demand driven by new applications and services. Meanwhile, network providers want to reduce both capital expenditures and operational expenditures to ensure that there is profit coming from their investments. This article proposes several migration strategies for AON from the data plane, topology, and control plane perspectives, and investigates their impact on the total cost of ownership

Introduction

Optical fiber communication, as a future-proof technology, has its unique advantages of delivering ultra-high capacity. It has been widely deployed in the telecommunication core and aggregation networks for several decades. Fiber to the X (known as FTTx, where x stands for home, building, curb, node, etc.) has also started all over the world. There are more than 93 million FTTx subscribers in Asia, 12 million in the United States, and 20 million in Europe [1]. The most common FTTx solutions deployed today are active optical networking (AON) and time-division multiplexing (TDM) passive optical networking (PON). AON, also known as active Ethernet, has been standardized [2], and is the most deployed FTTx solution in Europe. As of mid-2012, it represented 78 percent of market share [3], already giving it high technology penetration in Europe.

There are two variants of AONs, point-to-point (PtP) and active star (AS), both of which are based on active Ethernet switches. PtP is also referred to as homerun (shown in Fig. 1a), where each subscriber has a dedicated fiber connection between the residential gateway (RG), or optical network terminal (ONT) in fiber to the home (FTTH), and the Ethernet switch with an optical line terminal (OLT) located in the central office (CO). Unlike the PtP, the AON AS has a point-to-multipoint topology, employing active remote nodes (RNs) connecting the CO and multiple households (Fig. 1b). An RN can be located either in a cabinet or inside the building (e.g., a basement of a multi-dwelling unit). The Ethernet switch at an RN aggregates the traffic from a group of subscribers, and is connected by a feeder fiber to another Ethernet switch at the CO. Two or more feeder fibers can be deployed to provide resilience, but the amount of fibers used in the AON AS architecture can be significantly reduced compared to the PtP case. Figure 1c shows fiber to the building/curb (FTTB/C) based on AON AS. The optical signals are terminated at the RN, which is connected to the households via legacy copper cables.

AON deployments have already begun to offer 1 Gb/s per subscriber [4]. However, the earlier and more common deployments of AON solutions offer lower data rates (e.g., up to 100 Mb/s per ONT). Besides serving residential users, AON can support backhaul/fronthaul (Xhaul) for mobile networks, and broadband services for business users, as shown in Fig. 1d. The emerging services, including ultra-high-definition TV, video conferencing, cloud services, fourth/fifth generation (4G/5G) mobile Xhaul, and so on, are gradually eating up the bandwidth of existing networks and driving capacity demands beyond 100 Mb/s. Therefore, proper migration strategies toward solutions capable of delivering the new demanding services are urgently required. Furthermore, the incentive of operators/providers for migration is also related to their willingness to achieve total cost of ownership (TCO) savings by reducing operational expenditure (OPEX).

Different next generation optical access (NGOA) network technologies have been extensively investigated in the past. Wavelength-division multiplexing (WDM) is widely recognized as a promising technology to increase the bandwidth in FTTx. Reference [5] compared the cost and performance of different types of WDM-based PONs, including time and wavelength-division multiplexing (TWDM)-PON, wavelength-routed WDM-PON, and ultra-dense WDM-PON. It has been shown that for high bandwidth demands (beyond 500 Mb/s per customer), WDM technology becomes most efficient for capacity upgrade. Recently, the International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) also approved the second next generation passive optical network (NG-PON2) standard [6], where the primary technology is TWDM-PON. A complete cost evaluation of network migration starting from GPON to TWDM-

Kun Wang is with KTH Royal Institute of Technology and Acreo Swedish ICT; Lena Wosinska, and Jiajia Chen (corresponding author) are with KTH Royal Institute of Technology; Carmen Mas Machuca is with Technische Universität München. Anders Gavler and Kjell Brunnström are with Acreo Swedish ICT.

0163-6804/16/$25.00 © 2016 IEEE

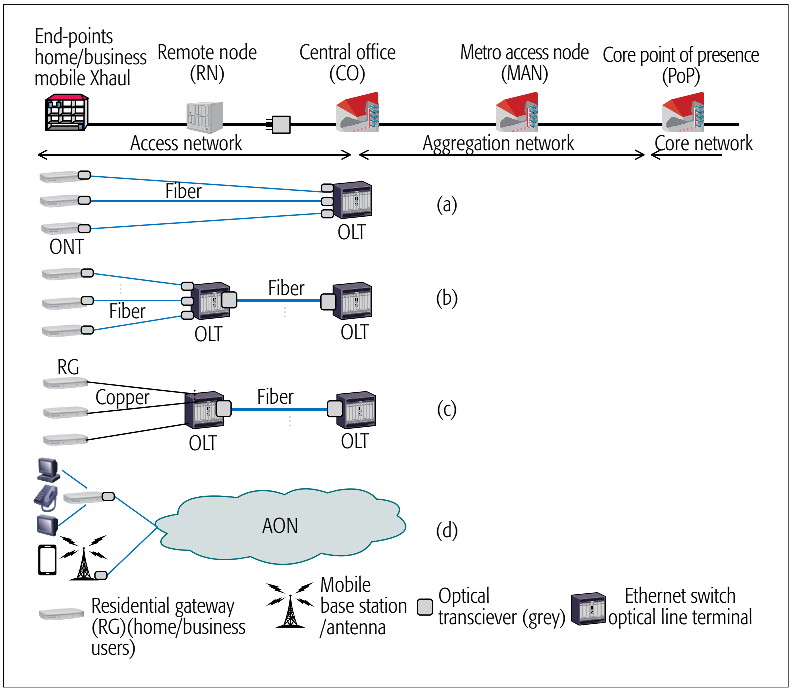


Figure 1. Current AON solutions for FTTx: a) AON PtP, FTTH; b) AON active star, FTTH; c) AON active star, FTTB/C/N; d) use cases.

PON was carried out in [7], in which the result shows that migrating to TWDM-PON is the best option thanks to its high sharing rate while providing high bandwidth on a per-user basis. Meanwhile, node consolidation has been considered as an important trend for access network migration driven by the operators/providers because of a high potential for the TCO savings. References [5, 7] indicate clear cost advantages of node consolidation due to better utilization of aggregation networks. All of the aforementioned studies have a strong focus on the evolution of data plane technologies. However, the other important aspects of network architecture, such as topology and control plane, have not been addressed. In recent years, there has been increasing interest in the concept of software defined networking (SDN). It separates the control plane from traditional network equipment, and migrates toward logically centralized control plane architecture, according to the Open Networking Foundation (ONF) [8]. Many vendors and operators/providers have foreseen great advantages of using SDN to simplify network control and management (C&M). For example, [9] investigates the applicability of SDN to a gigabit-capable PON-based FTTH network. AON is fully based on IP/Ethernet technology, so the applicability of SDN to AON is quite straightforward. However, the impact on TCO and benefits from migration toward SDN enabled by AON are rarely studied.

A comprehensive view of network architecture should include various aspects, such as data plane, control plane, and network topology. In contrast to the existing work, we take into account all three aspects, concentrate on AON, and systematically investigate possible migration strategies from current AON deployment to NGOA networks.

The remainder of this article is organized as follows. In the following section, several cost related aspects are listed, which later on are used as key parameters for high-level evaluation of the considered migration strategies. We then address migration strategies from three different perspectives: data plane, topology, and control plane. Finally, conclusions are drawn.

Network Migration Costs

From the network operators'/providers' perspective, besides the demands of upgrading the network capacity to satisfy the bandwidth demand of emerging services and expanding the network coverage to accommodate more customers, the other major goal of network migration is to achieve higher profits. Therefore, when planning a network migration, operators need to evaluate both the required new investment (capital expenditure, CAPEX) and the potential OPEX saving.

CAPEX

Migration from an old platform/technology to a new one inevitably requires new investment. In order to benefit from the migration and achieve maximal return on investment, operators/providers want to minimize the discounted payback period, while keeping CAPEX as low as possible. CAPEX refers to any costs related to the infrastructure and equipment that need to be purchased and installed before the network becomes operable. The major CAPEX for the access network can be divided into three categories:
- Fiber infrastructure
- Network equipment
- Residential gateway

Fiber infrastructure in the access network embraces all fiber related costs such as duct, fiber cable, trenching, splicing, installations, power splitters, and wavelength filters. It is normally the most expensive part of FTTx deployment, especially when trenching is required [5]. Therefore, the key factor to minimize the CAPEX of network migration is to reuse the existing fiber infrastructure as much as possible.

The investment in next generation network equipment refers to the costs of any active equipment to be placed in the metro-access node (MAN), CO, as well as RN (e.g., Ethernet switches, OLTs). In some cases, WDM filters are integrated together with OLTs; therefore, the cost of these passive components are considered as a part of the OLT cost.

The RG related cost is one of the most important parts of the CAPEX [10]. Migration to the next generation network may require the replacement of all RGs at subscribers' premises due to the upgrades of data plane technology and capacity. In that case, the replacement of RGs is also one of the key aspects of network migration.

OPEX

Another aspect that should be evaluated when considering the gain of a migration strategy is the OPEX reduction of the new network compared to the legacy platform. OPEX refers to any costs required for the operation of the network. The major OPEX components are:
- Energy consumption
- Service provisioning
- Fault management and maintenance

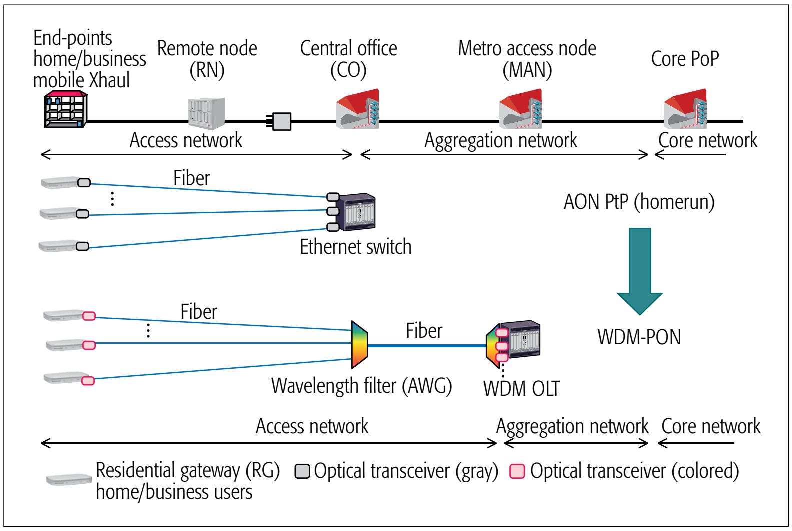


Figure 2. Migration from AON PtP to WDM-PON with the node consolidation approach.

From a network operator/provider point of view, the OPEX related to energy consumption refers to the electricity bill for powering and cooling network equipment. The energy consumption of RG is excluded here because it is usually paid by the subscribers

Service provisioning (SP) is the cost associated with any activities related to adding, changing, and cancelling customer services (e.g., network and service configurations, fiber patching at different locations for connecting a new customer or new services, provider change, user moves). Many factors have impact on the SP cost, such as the required fiber and equipment, the possibility of remote configuration, human resources, and travelling, needed to connect a new customer or change services.

The OPEX related to fault management (FM) is the cost associated with the detection and reparation of any failure in the network including both equipment and infrastructure. The FM process includes failure detection, help desk, opening a trouble ticket, reparation of the failure, travelling to the failure locations, and the required human resource. Maintenance comprises all tasks required to keep the network up and running. This includes software and hardware upgrades, personnel inspections, performance monitoring, inventory management, and so on.

Data Plane Migration

In this section, we investigate migration strategies considering the node consolidation approach from the data plane perspective, taking into account the characteristics of existing AON deployments.

Node Consolidation

The motivation for node consolidation is to reduce the number of COs so that the costs associated with those nodes(e.g., housing, energy, and maintenance) can be saved. All network equipment will therefore be moved to the MAN, allowing many more end users and serving larger areas than the current CO.

Figure 2 shows a proposed migration path from current AON PtP to WDM-PON [5] where WDM technology is recommended to avoid costly additional trenching over a long distance in an aggregation network. In the case of node consolidation, a number of COs are supposed to be closed down to save costs; therefore, all active network equipment has to be moved into a MAN. Wavelength filters (arrayed waveguide gratings, AWGs) are placed in the location of traditional COs to replace the AON PtP switches. The AWGs are passive components and can be installed either underground or in a cabinet, where electrical power is not needed. The AWG multiplexes the wavelength coming from a number of customers into one feeder fiber, which connects to the WDM-PON OLT (Ethernet switch with another AWG and colored optical interfaces). Every user connected to the AWG is assigned a dedicated wavelength from the OLT.

For an existing AON AS, a proper data plane migration option can be toward a fully passive solution, such as TWDM- PON, which is defined by ITU-T ([6]) as a primary technology for NG-PON2. In this case, the active RNs are replaced by passive power splitters, while the Ethernet switches at the old CO location are replaced by AWGs (Fig. 3). Therefore, both RNs and COs can potentially be eliminated to support node consolidation.

Costs Evaluation

CAPEX: From the fiber infrastructure perspective, both WDM-PON and TWDM-PON have the advantage of reusing existing fiber infrastructure from legacy AON, so a large part of CAPEX can be saved. The fiber connection between CO and MAN may also require additional investment if it is not available from the legacy network. However, thanks to WDM technology, the additional amount of fibers required for network upgrade can be reduced. They can be installed in existing ducts, and thus the huge trenching costs can be avoided. There is some fiber splicing and reconnection work involved at COs and RNs. The migration path from AON PtP to WDM-PON can be realized by reconnecting fibers at COs, while migration from AON AS to TWDM-PON requires fiber reconnection at both RNs and COs. AWGs have to be installed at COs and MANs for both migration paths, and active RNs need to be replaced by power splitters.

Both migration paths require investment in network equipment such as OLTs (including optical transceivers) at MANs. At the customers' side, replacement of RGs/ONTs is required in order to adapt to WDM or TWDM technologies. When replacing the active equipment at the CO and RN with passive devices, the cost of closing down the sites and moving the active equipment to MANS may be substantial. Regarding coexistence, passive solutions cannot coexist with legacy AON, and therefore cannot run simultaneously on the same fiber infrastructure in the case of both presented data plane migration paths (i.e., from P2P to WDM-PON and from AON active star to TWDM-PON).

OPEX: The aim of the node consolidation approach is to save OPEX by reducing the number of COs.

Reducing the number of COs will probably not significantly change the overall energy consumption [11]. The energy that network and cooling equipment consume is mainly dependent on the number of customers and the amount of traffic carried by the equipment. This is because node consolidation only reduces the number of access nodes, while the amount of active network equipment is unchanged, as it is either moved to the MAN or replaced by new equipment at the MAN. Therefore, a large portion of CO power consumption is shifted to the MAN.

In the current AON PtP networks, adding or changing customers' subscriptions requires technicians to travel to the COs, manually add/remove fibers, and install new switches if needed. In the AON AS case, technicians need to visit RNs (sometimes both RNs and COs) to perform the tasks. There are a large number of distributed COs and RNs in the network. Therefore, SP in current networks may involve a high cost in human resources. When AON PtP and AS are migrated to the consolidated WDM-PON and TWDM-PON, respectively, only passive components are at COs and/or RNs. The technicians are normally placed at the MAN locations,[1] and there is no need for human involved work at COs. Consequently, the costs associated with travelling and human resource can be reduced significantly.

The decommission of traditional COs can reduce the overall effort for the FM, such as maintenance and administration of the building infrastructure, including cost of renting, cleaning, gardening in the outside area if any, heating, renovations, insurances, and so on. WDM-PON and TWDM-PON do not require any active network equipment in the field, which simplifies maintenance of the network. Consequently, FM and maintenance processes are less time consuming; hence, the cost of human resources can be saved. However, under some circumstances, COs cannot be completely closed down because they are still used for other purposes (e.g. telephone networks, content delivery networks [CDNs], or regulatory reasons). In such cases, the maintenance and other operational cost reduction may be limited. Another important issue in a node consolidated network is the node failure. Due to a large number of subscribers connected to a single node, any failure (e.g., power supply fault at a MAN) can simultaneously affect many customers. Therefore, efficient FM and resiliency mechanisms are required to avoid service interruption for a large amount of customers. This can potentially be addressed through topology migration, introduced next.

Topology Migration

In contrast to tree, mesh/ring topology has better connectivity and offers better resiliency and traffic locality (which is defined as the ability to keep traffic locally in order to offload metro/core network [12]) so that the network performance and quality of experience (QoE) of end users can be greatly enhanced. Some operators/providers have already started building their fiber access network with mesh/ring topology (e.g., [13]).

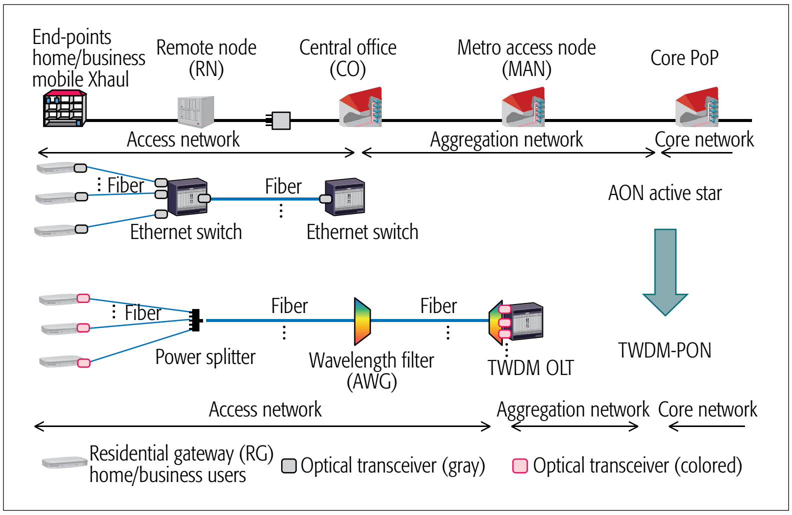


Figure 3. Migration from AON active star to TWDM-PON with node consolidation approach.

Mesh/Ring Topology

A legacy telecommunication network is designed to deliver traditional Internet services that cannot match current and future service requirements (e.g., CDN, 4G/5G Xhaul). Therefore, network operators/providers are trying to build or redesign networks in order to address this problem, while minimizing the TCO by sharing network equipment, fiber infrastructure, housing, and maintenance.

The increasing bandwidth demand from the endpoints (including fixed broadband users and cells for radio access networks) combined with high customer density leads to a huge amount of data traffic toward the aggregation and core network. It increases core network load and energy consumption, and consequently may degrade the network performance. This problem becomes more severe in the case of bandwidth-demanding video content distribution. Local caching of content [12] has been proposed to address this problem. Mesh/ring topology better suits the CDNs, taking advantage of the distributed nodes that are close to the end users, and making use of high connectivity to share caches among neighboring nodes.

The resilience requirement is driven by the increasing demand for a reliable and high-quality broadband infrastructure, particularly important for business customers, mobile Xhaul, public services, healthcare, and so on. From the resiliency point of view, mesh/ring topology is better than PtP/star/tree. If failures occur, alternative disjoint paths can easily be found in most cases, and hence the impact of the failures can be reduced. In contrast, in a star/tree topology, the failure of a centralized node can have a very high impact due to a large number of connected subscribers.

Figure 4a illustrates AON migration from tree to mesh/ring topology. The mesh/ring topology migration keeps all existing AON nodes and fiber connections (solid lines in the figure), and adds more connections between the nodes so that the new network topology becomes mesh/ring. The dashed lines in the figure show examples of the potential links that can be added. The migration toward mesh/ring topology not only

[1] Private communication with Deutsche Telekom within the EU-FP7 OASE project, www.ict-oase.eu.

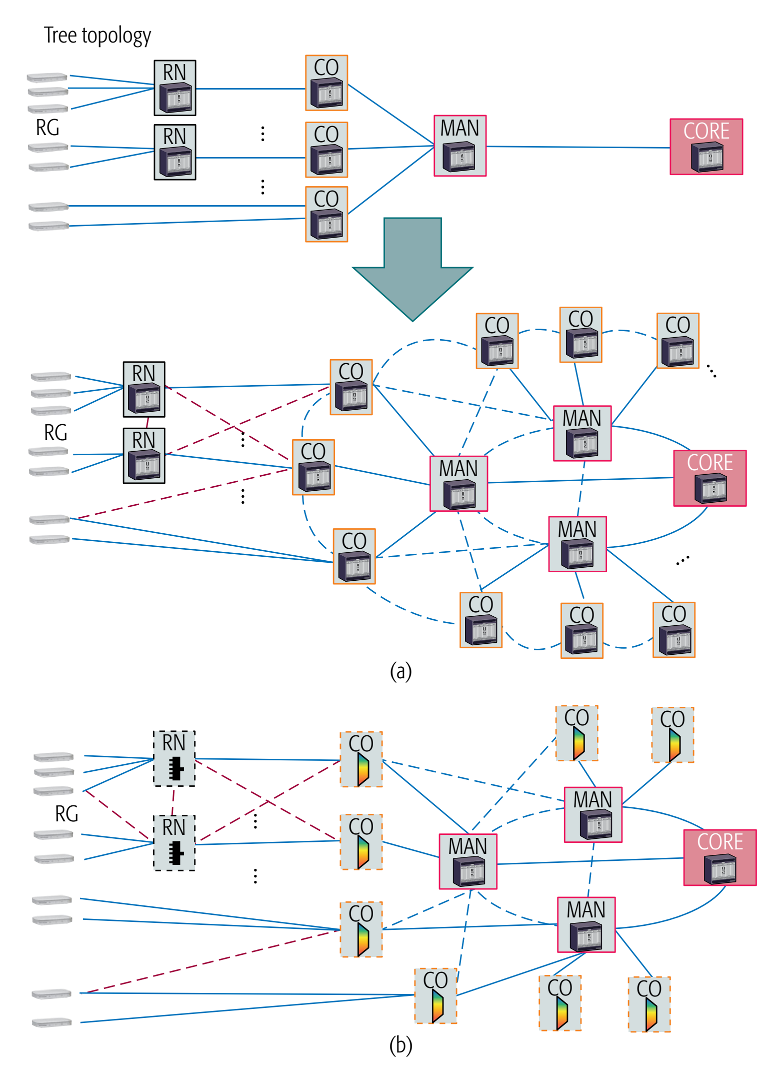


Figure 4. Migration from tree to the mesh/ring topology (where dashed lines refer to fiber links added for topology evolution): a) AON in mesh/ring topology; b) WDM /TWDM PON in mesh/ring topology.

improves network resilience by adding node and link protection, but also enhances traffic locality. Furthermore, from the end users' point of view, the QoE can be improved by traffic locality (e.g., shorter downloading time, smooth video playout). Figure 4b shows an example of combining data plane and topology migrations, that is, node consolidated WDM/TWDM-PON with mesh/ring topology. Besides interconnecting MANs in the mesh/ring, the additional links (e.g., the fibers between the RNs, and between the RGs and COs/RNs) can be added for resilience purposes. In some of these cases, the housing of RNs and COs does not exist, as only passive components are located.

Costs Evaluation

CAPEX: The CAPEX of the mesh/ring topology migration mainly involves adding new fiber connections between the nodes. It can be costly especially when the additional trenching is required. Good network planning can help to reduce these costs. For example, the migration plan can be coordinated with CDN deployment, where the new connections between neighboring nodes would already be provided around the caching locations. Enhanced reliability performance can be offered in the first place for important customers (e.g., business users, healthcare, mobile backhaul) that are willing to pay extra for higher quality of service.

There are no significant changes required for network equipment and RGs in the case of topology migration. However, in order to have very high reliability performance (including protection for the last mile), redundant equipment and optical interfaces are needed. For some endpoints requiring ultra-high connection availability (e.g., higher than 4 nines), two optical interfaces are necessary in order to have end-to-end protection. For WDM/TWDM-PON there are costs of upgrading the passive splitters and filters from 1:M to 2:M in order to provide feeder fiber protection.

OPEX: The potential benefits of mesh/ring topology come from traffic locality, which can reduce the capacity required and energy consumed at higher aggregation nodes. In turn, the costs for the network operator/provider can be reduced. However, the power consumption in the access network can be higher due to a number of distributed nodes. The links dedicated for resilience can be switched off if no failure occurs. On the other hand, it is at the expense of a longer recovery time.

In a mesh/ring AON, SP processes still require technicians to travel to many distributed nodes and to add/remove links manually, while in a mesh/ring WDM-PON/TWDM-PON the need for SP processes exists only at MANs.

Regarding the FM and maintenance aspects, mesh/ring topology can significantly reduce the impact of failure by inherent resilience. The network will become less sensitive to the length of the reparation time. The additional equipment that needs to be installed for redundance increases the maintenance and FM cost because it can also fail and then have to be repaired/replaced. On the other hand, a service penalty, which is quite often paid to business customers due to disconnection, can be dramatically reduced, since the risk of service interruption is much lower.

Control Plane Migration

In this section we investigate the migration strategy in the control plane, which can be carried out together with the network evolutions in the data plane and topology introduced in the previous sections.

Migration Strategies toward SDN

SDN separates the control plane from the traditional network equipment aiming to simplify network operations, and manage the network in a cost-effective and flexible way [8].

One popular SDN architecture is shown in Fig. 5a, where multiple SDN controllers are used according to different network domains and network services. A network orchestrator on top of all controllers coordinates their activities [14]. This solution may fit better in a large network that includes many nodes and devices. The individual controller at each network domain is only

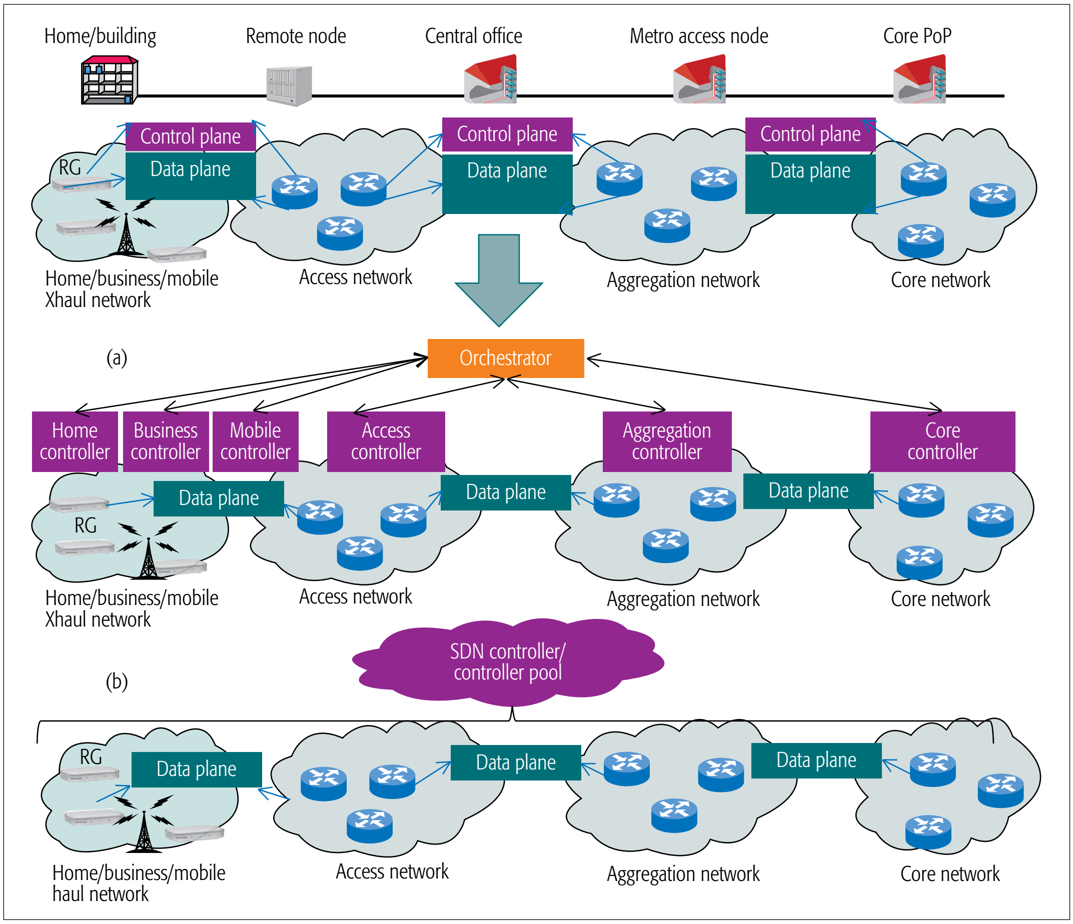


Figure 5. Control plane migration toward an SDN-enabled network: a) option 1; b) option 2.

There are no significant changes required for network equipment and RGs in the case of topology migration. However, in order to have very high reliability performance (including protection for the last mile), redundant equipment and optical interfaces are needed.

responsible for the equipment within its domain. Therefore, it has better performance in terms of security, response time, scalability, and so on. However, it may lose some flexibility and efficiency when it comes to cross-domain application, since it needs coordination among different controllers. The use of a separated controller is good for a network that involves different technologies (e.g., PON in access, IP/Ethernet in core), different service functionalities (e.g., broadband access for residential/business users, mobile Xhaul), or multiple network providers/operators.

An alternative control plane migration toward an SDN-based network is shown in Fig. 5b, where a logically centralized SDN controller is used crossing multiple network domains. A pool of controllers may be used when resilience and scalability are needed. The controller itself also acts as an orchestrator that coordinates network resources across different domains and services. Since AON is fully based on IP/Ethernet technology, which is also the main technology today used in home, enterprise, aggregation, and core networks, it makes the control plane of AON easier to integrate with core/aggregation as a unified controller. The integrated controller has a global view of all network devices and the entire network topology, so it can quickly and efficiently allocate the resources and find an optimal path across the whole network. However, when the network becomes larger (i.e., the number of network devices is high), it may bring issues in latency, security, and so on. Therefore, this solution may be suitable for operators that are running small or medium size networks, but covering access, aggregation, and/or core segments simultaneously. Actually, many existing AONs (e.g., municipal networks) are run by small operators.

Cost Evaluation

CAPEX: From the CAPEX point of view, the major investment of this migration path is purchasing and installing SDN-enabled network equipment (e.g., OpenFlow switches, controller, orchestrator). However, the cost of SDN enabled network equipment is expected to be lower than that of current devices, since the sophisticated control plane and operation system are no longer needed in each individual network device. The controller consists of servers and software, and the cost is relatively low (some open source software is also available [15]). Thanks to the flexibility and programmability of SDN itself, the existing legacy network elements can be reused in the SDN-enabled network. Generally speaking, if migration is only carried out in the control plane toward SDN, all legacy AON devices can be kept, including RG/ONT, since it is an integration of the Ethernet switch/router and optical terminal. However, additional software developments are needed in order to make the legacy device programmable for the SDN controller. For example, legacy switches commonly use certain C&M protocols, such as Simple Network Management Protocol (SNMP). In the SDN controller a software plugin has to be implemented so that the controller can control and reconfigure legacy devices through SNMP. In the OpenDayLight controller community [15], there are many working projects focusing on the development of such plugins. Although the reuse of existing network elements saves the investment in the hardware, replacing legacy equipment with SDN-enabled devices can bring more benefits, such as automatic network topology discovery, a specialized and efficient packet forwarding mechanism, and fast deployment and configuration.

OPEX: AON uses IP/Ethernet as data plane technology, while SDN has good compatibility with IP/Ethernet. Therefore, AON is able to realize simple control plane migration toward an SDN-enabled network. With the help of SDN, AON can optimize network C&M processes and improve the network operation efficiency.

Thanks to SDN programmability, the power consumption of the network can potentially be reduced through dynamic network resource allocation. For example, the controller/orchestrator can route and aggregate the traffic flows to a certain path according to the traffic conditions in the network so that a lower number of nodes are involved, and thus it becomes more energy-efficient than traffic flows distributed over many underutilized nodes. Some of the network elements can be switched off or put in a power saving mode by the controller/orchestrator during periods when there is no or low traffic. The SDN orchestration among different service domains (mobile Xhaul, home/business broadband access) can improve network utilization and energy efficiency not only in the optical transport network, but also in the other network segments (e.g., radio access network) [14].

From the SP perspective, the cost of control plane migration from conventional AON toward an SDN-enabled network mainly involves decommissioning of old C&M in the legacy network elements, upgrading and reconfiguring new SDN controllers, and provisioning new services. In the current networks, it takes a lot of time and human resources when network operators/providers deploy new services or upgrade existing ones, for example, where changes of policies, capacities, and routing rules are needed. The SDN controller provides an interface that allows deployment of applications/services on top of the infrastructure to automatically optimize and quickly initialize new end-to-end services across heterogeneous domains. Operators/providers can freely and easily change the network configurations and routing rules, and dynamically allocate capacity to match the varying customers' traffic demand [14], and consequently the SP costs can be significantly reduced.

The costs of FM and maintenance can also be reduced by migrating toward SDN. The controllers have complete views of the entire network (network elements information, capacities, topologies, traffic load, etc.), which simplifies the network monitoring and fault detection functionality. When a failure happens, the controller/orchestrator can automatically calculate an alternative path and reroute the traffic if available. To repair faulty devices, the required AON technicians can be shared with the core/aggregation network technicians because both AON and the core/aggregation network use the same data plane technology. In addition, because the SDN data plane element has no individual C&M, it is vendor-neutral, which simplifies the installation and repair processes.

The proposed migration strategies have both pros and cons depending on their features. They can be adopted either individually or combined. Operators/providers may choose the migration strategies which fit best to both their current network characteristics and future service/network planning.

Discussion and Conclusion

The aforementioned migration strategies are proposed from the data plane, topology, and control plane perspective. For a network operator/provider, the selected migration strategy can either consider only one of the aspects or is based on a combination of different aspects. In general, one migration strategy can be integrated with the other strategies, so that the benefits of a single migration method can be added on top of the others. For example, in the migration path focused on the data plane (AON towards WDM-PON or TWDM-PON), SDN controllers can be used for both optical WDM/TWDM-based and electrical packet-based network equipment in order to simplify network C&M.

In a combination of mesh/ring topology and SDN, the SDN can utilize the advantage of mesh/ring topology to allocate network resource dynamically in a flexible way according to the traffic locality pattern, while the mesh/ring topology can help the SDN controller with more options (paths) for FM, traffic engineering, routing, and so on.

The migrations toward node consolidation and mesh/ring topology are quite different in nature. The first one aims to reduce the number of access network nodes, while the second one makes use of the distributed nodes to improve network resilience and traffic locality. However, these two migration trends can still be integrated to a certain extent. Interconnecting MANs and passive components in the field in mesh/ring can improve the network resilience. It is extremely important for the node consolidation scenario, where any single failure may affect a huge number of endpoints.

This article proposes several migration strategies for AON from the data plane, topology, and control plane perspectives, respectively, and evaluates these strategies with respect to the key elements of both CAPEX and OPEX. The proposed migration strategies have both pros and cons depending on their features. They can be adopted either individually or in combination. Operators/providers may choose the migration strategies that best fit both their current network characteristics and future service/network planning.

Acknowledgment

This work was supported in part by VINNOVA (The Swedish Governmental Agency for Innovation Systems) through Kista 5G transport lab, and the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement no. 249025 (ICT OASE).

References

[1] A. Hatt *et al.*, "Creating a Brighter Future," *Proc. FTTH Council Europe Press Conf.*, Stockholm, Sweden, 2014; http://www.ftthcouncil.eu/documents/Presentations/20140219PressConfStockholm.pdf

[2] IEEE 802.3ah EFM Std, "IEEE Ethernet in the First Mile," 2005.

[3] V. Chaillou, "Inventory of FTTH/B in Europe," IDATE DigiWorld Inst., 2012; http://www.idate.org/en/News/Inventory-of-FTTH-B-in-Europe_765.html

[4] O. Malik, "In the Netherlands, 1 Gb/s Broadband Will Soon Be Everywhere," 2010; https://gigaom.com/2010/02/19/in-netherlands-1-gbps-broadband-will-soon-be-everywhere/https://gigaom.com/2010/02/19/in-netherlands-1-gbps-broadband-will-soon-be-everywhere/

[5] R. Hlsermann, K. Grobe, and D. Breuer, "Cost and Performance Evaluation of WDM-Based Access Networks," *Proc. OFC/NFOEC*, Anaheim, CA, Mar. 17–21, 2013, Paper NTh3F.4.

[6] ITU-T G.989 Series Recs., "40-Gigabit-Capable Passive Optical Network (NG-PON2)," 2015.

[7] C. Machuca, S. Krauss, and M. Kind, "Migration from GPON to Hybrid PON: Complete Cost Evaluation," *Proc. 14th 2013 ITG Symp. Photonic Networks*, Leipzig, Germany, May 6–7, 2013, pp.1–6.

[8] "Software-Defined Networking: The New Norm for Networks," ONF White Paper, Apr. 13, 2012; https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wp-sdn-newnorm.pdf

[9] N. Bitar, "SDN and Potential Applicability to Access Networks," *Proc. OFC*, San Francisco, CA, Mar. 9–13, 2014, Paper Th1G.4,

[10] C. Bock *et al*., "Techno-Economics and Performance of Convergent Radio and Fibre Architectures," *Proc. 16th Int'l. Conf. Transparent Optical Networks*, Graz, Austria, July 2014, pp. 1–4.

[11] B. Skubic and I. Pappa, "Energy Consumption Analysis of Converged Networks: Node Consolidation vs Metro Simplification," *Proc. OFC/NFOEC*, Anaheim, CA, Mar. 17–21, 2013, paper NM2I.5.

[12] V. Nordell *et al*., "Concurrency and Locality of Content Demand," *2013 Int'l. Conf. Smart Commun. Network Technologies*, vol. 03, June, 2013, pp. 1–5.

[13] A. Broberg, "Public Webinar on: Stokab-Building, Operating and Financing the World's Largest Open Passive Fibre Network," *FTTH Council Europe*, 2013; http://www.ftthcouncil.eu/documents/Webinars/2013/Webinar_16October2013.pdf

[14] A. Rostami *et al*., "First Experimental Demonstration of Orchestration of Optical Transport, RAN and Cloud Based on SDN," *Proc. OFC*, Los Angeles, CA, Mar. 22–26, 2015, Paper Th5A.7.

[15] OpenDaylight SDN Platform, a Linux Foundation Collaborative Project, 2013; http://www.opendaylight.org.

Biographies

Kun Wang received his Master of Science in electrical engineering from KTH Royal Institute of Technology, Sweden, in March 2007. Currently he is working at Acreo Swedish ICT and KTH toward his Ph.D. in next generation optical transport networks. His research interests include FTTx networks, software defined networking, and 5G networks.

Anders Galver received his M.Sc. degree in astro-physics from Stockholm University in 1998. He then joined Ericsson AB, Optical Networks Research Laboratory, and since 2002 he has been with Acreo Swedish ICT. Currently he is group manager of Applied Networking and senior project manager of the multi-layer network architecture research area..

Lena Wosinska [SM] received her Ph.D. degree in photonics and Docent degree in optical networking from KTH Royal Institute of Technology, where she is currently a full professor of telecommunication in the School of Information and Communication Technology (ICT). She is founder and leader of the Optical Networks Lab (ONLab). She has worked in several EU projects and coordinated a number of national and international research projects. Her research interests include fiber access and 5G transport networks, energy-efficient optical networks, photonics in switching, optical network control, reliability and survivability, and optical data center networks.

Kjell Brunnström is a senior scientist at Acreo Swedish ICT AB and an adjunct professor at Mid Sweden University. He is Co-Chair of the Video Quality Experts Group (VQEG). His research interests are in quality of experience for visual media, particularly video quality assessment and display quality.

Carmen Mas Machuca [SM] has been a senior researcher at the Institute for Communication Networks (TUM), Germany, since December 2005. Her main research interests are in the area of converged optical networks and techno-economic studies. She has published more than 100 peer-reviewed papers. She chaired CTTE 2015 and co-chaired RNDM 2015.

Jiajia Chen (jiajiac@kth.se) [SM] received her Ph.D. degree in 2009 and Docent degree in 2015 from KTH Royal Institute of Technology. Currently, she is working as assistant director of the Sino-Swedish joint research center of photonics (JORCEP). She is a co-author of more than 100 papers in optical networking. Her research interests include optical interconnection and transport network technologies.

SUDOI: Software Defined Networking for Ubiquitous Data Center Optical Interconnection

Hui Yang, Jie Zhang, Yongli Zhao, Jianrui Han, Yi Lin, and Young Lee

The authors present a novel SUDOI architecture aiming at the extensive user access from the perspective of heterogeneous cross-stratum and multi-layer networking modes. SUDOI can enable cross-stratum optimization of application and optical network stratum resources, and enhance multiple layer resource integration in ubiquitous data center optical interconnection.

Abstract

Ubiquitous data center optical interconnection is a promising scenario to meet the high burstiness and high-bandwidth requirements of services in terms of the user-access-oriented interconnection between user and data center, inter-data-center, and intra-data-center interconnection. However, in the current mode of operation, the control of the data center and optical network is separately deployed. Enabling even limited interworking among these separated control systems does not provide a mechanism to exchange resource information and enhance the high-level performance requirement of applications. Our previous work implemented cross-stratum optimization of optical network and application strata resources inter-data-center, which allows the accommodation of data center services. In view of this, this study extends to the ubiquitous data center optical interconnection scenario. This article presents a novel SUDOI architecture aimed at extensive user access from the perspective of heterogeneous cross-stratum and multi-layer networking modes. SUDOI can enable cross-stratum optimization of application and optical network stratum resources, and enhance multiple-layer resource integration in the ubiquitous data center optical interconnection. The functional modules of SUDOI architecture, including the core elements of various controllers, are described in detail. The cooperation procedure in user-access-oriented interconnection, multiple-layer resource integration inter-data-center, and intra-data-center service modes is investigated. The feasibility and efficiency of the proposed architecture are also experimentally demonstrated on our OaaS testbed with OpenFlow-enabled optical nodes, and compared to the CSO scheme in terms of blocking probability and resource occupation rate. Numerical results are given and analyzed based on the testbed. Some future discussion and exploration issues are presented in the conclusion.

Introduction

Due to the rapid evolution of high-bit-rate data-center-supported applications, network operators are rethinking the way their networks are controlled to provide interconnection between data centers and users. With the extension of large-scale networks, data center services are typically diverse in terms of required bandwidth and usage patterns. The network traffic supporting such services shows high burstiness and high-bandwidth characteristics. The traditional electrical switching network is inefficient to carry these applications due to the technical bottleneck of bandwidth capacity, energy consumption, and transmission range. To accommodate these services in highly available and energy-efficient flexible connectivity channels, the architecture of optical interconnection and switching among data centers was proposed and demonstrated [1], especially containing elastic optical networks (EONs) [2–3]. Data center interconnection by optical network is a promising scenario to allocate spectrum resources for applications in a highly dynamic, tunable, and efficiently controlled manner.

Additionally, many delay-sensitive services require high-level end-to-end quality of service (QoS) guarantee [4]. This forces the research field of data center interconnection to not be limited to intra-data-center, which further forms the scenario of ubiquitous data center optical interconnection, including between user and data center, inter-data-center, and intra-data-center. In such a scenario, the services guaranteeing QoS not only make significant use of optical network resources in the form of bandwidth consumption, but also relate to application resources (e.g., computing and storage resource) in data centers [5]. Traditionally, the control of data centers and optical networks have been separately deployed in data center interconnect architecture [6]. Depending on the technological heterogeneity and resource diversity, the current architecture does not provide a mechanism to exchange resource information and enhance the high-level performance requirements of these applications for ubiquitous data center optical interconnection.

Software defined networking (SDN) as promising for centralized control has gained popularity by supporting programmability of data center and network functionalities [7–9]. SDN can provide maximum flexibility for operators, unified control over various resources and an abstrac-

Hui Yang (corresponding author), Jie Zhang, and Yongli Zhao are with Beijing University of Posts and Telecommunications; Jianrui Han, Yi Lin, and Young Lee are with Huawei Technologies Co., Ltd.

0163-6804/16/$25.00 © 2016 IEEE

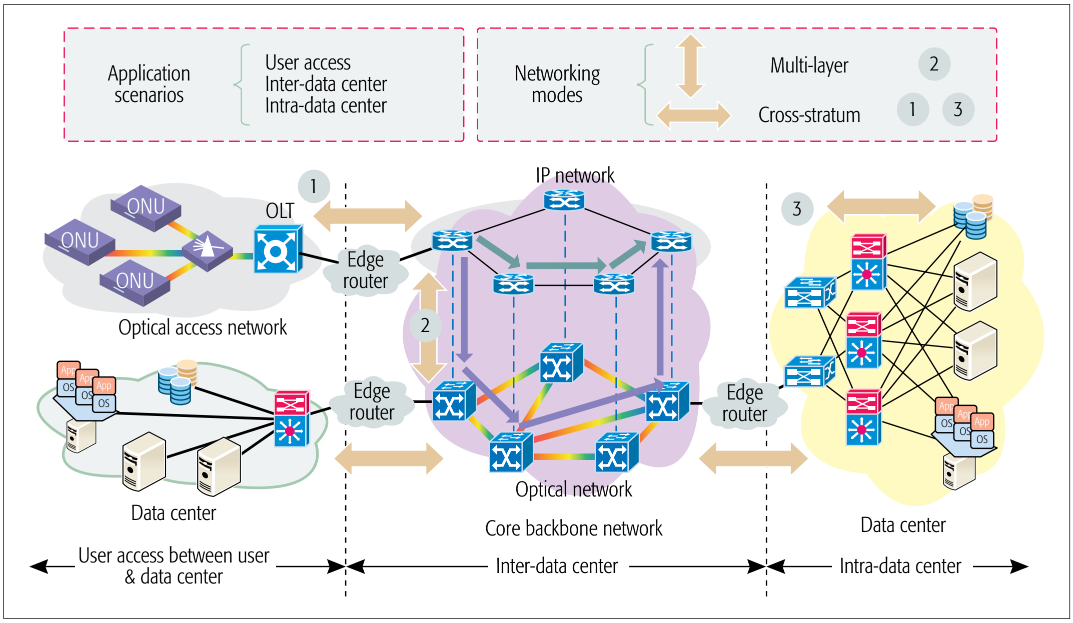


Figure 1. The networking modes and application scenarios for SUDOI.

The interconnection and networking of multiple layers are established along a longitudinal direction, which is called "multi-layer." Based on two networking modes, three scenarios in SUDOI are formed: user-access-oriented interconnection between user and data center, inter-data center, and intra-data center optical interconnection.

tion mechanism of these resources [10, 11], and realize cross-stratum optimization (CSO), which allows global optimization and control across optical network and data center resources in the inter-data-center scenario [12]. This article is not limited to the inter-data center scenario, but also further forms the scenario of ubiquitous data center optical interconnection, including interconnection between the user and data center, and inter-data center and intra-data-center. Different from other projects, this article proposes a software-defined ubiquitous data center optical interconnection (SUDOI) architecture aimed at extensive user access from the perspective of heterogeneous cross-stratum and multi-layer networking modes, by introducing a service-aware schedule scheme. SUDOI can enable cross-stratum optimization of application and optical network stratum resources, and enhance multiple-layer resource integration among various carried granularities in terms of user-access-oriented, inter-data-center, and intra-data-center optical interconnection.

The rest of the article is organized as follows. We describe the SUDOI architecture and relationship among various application scenarios. Functional modules and cooperation procedures for SUDOI in three scenarios and a service-aware schedule scheme are discussed. The SUDOI testbed with OpenFlow-enabled optical devices is presented and discussed with numerical results. Finally, we conclude the article and discuss future research issues.

SUDOI Architecture for Ubiquitous Data Center Interconnection

We present the SUDOI architecture to aim at extensive user access. The networking mode for SUDOI is extended in two directions, as shown in Fig. 1. One is from the perspective of *resource form*. Network and application resources from the access, core network, and data center are interconnected along the east-west direction. It leads to the networking of heterogeneous resources across strata in the *latitudinal direction*, which is established as *cross-stratum*. The other is from the perspective of the *relationship of carrying capacity*. The related entity with small granularity of switching is abstracted as a high-layer network (e.g., IP network), while the entity with large switching granularity is abstracted into a low-layer network (e.g., EON). The interconnection and networking of multiple layers are established along a *longitudinal direction*, which is called *multi-layer*. Based on two networking modes, three scenarios in SUDOI are formed: user-access-oriented interconnection between user and data center, inter-data-center, and intra-data-center optical interconnection.

SUDOI Architecture

The SUDOI architecture for extensive user access is illustrated in Fig. 2. The overlaid IP and optical networks are used to interconnect the distributed data centers, while optical access networks are interconnected and converged into a metro network. Intra-data-center, three levels of optical switches (top-of-rack, aggregation, and core optical switches) are used to interconnect servers. SUDOI consists of four strata/layers: optical access resources, IP resources, optical network resources, and application resources, which are software defined with OpenFlow protocol (OFP) and controlled locally by the optical access controller (OC), IP controller (IPC), transport controller (TC), and application controller (AC) in a unified manner. To control the SUDOI with OFP, an OpenFlow-enabled optical line terminal (OLT) node, IP routers, and optical switches equipped with OFP agents are required, which are referred to as OF-OLT, OF-Router, and OF-OS, respectively, as proposed in [13, 14]. OpenFlow-enabled agent software is embedded to keep the communication between the controller and optical node, which achieves the OFP process and mapping to the physical hardware. The motivations for SUDOI are two-

The flow schedule decides to allocate and adjust the bandwidth to guarantee the QoS, and deploys the SA-FS function. It can interact application usage with AC via an optical-application interface (OAI) and schedule the bandwidth among various OF-OLTs through IPC via an IP-optical interface.

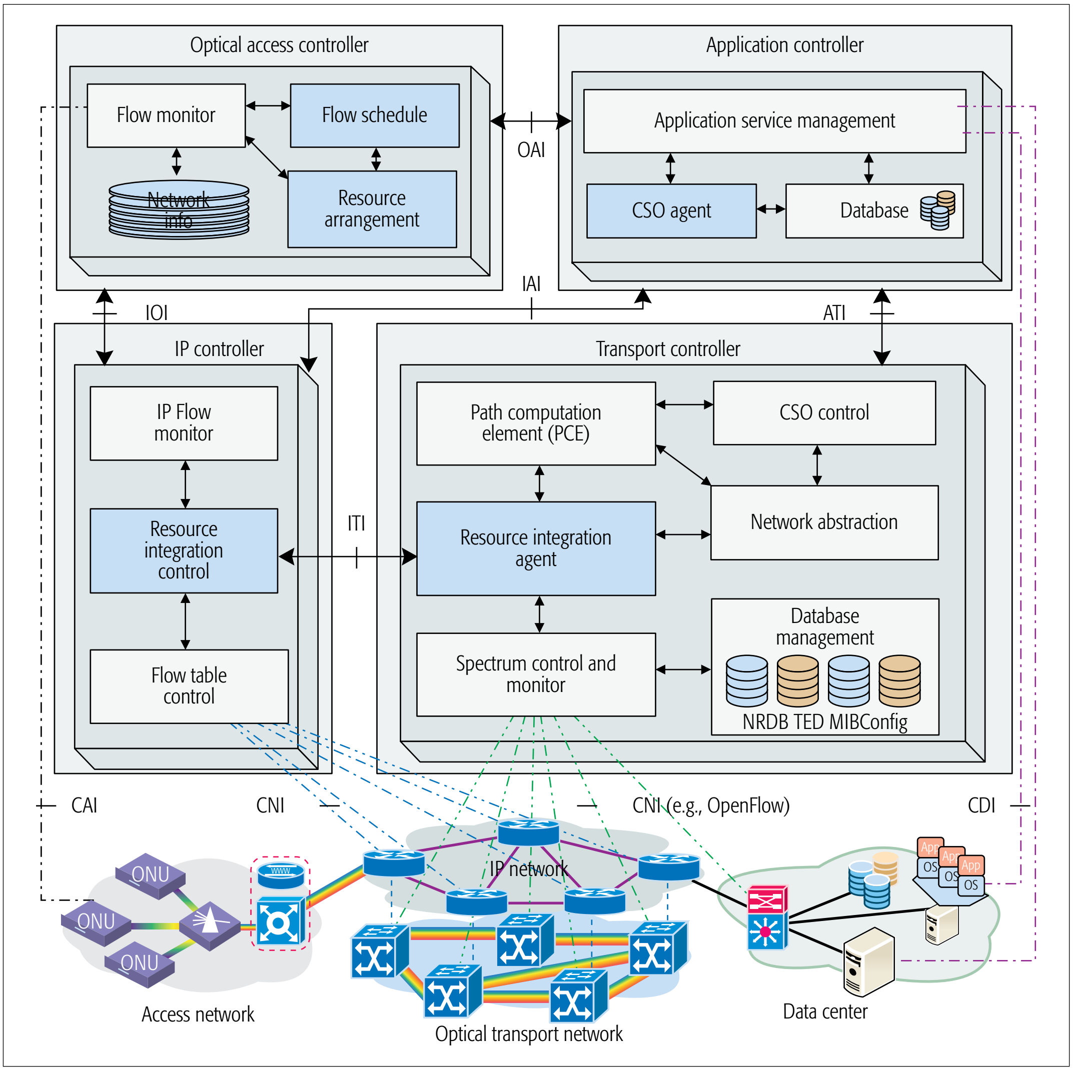


Figure 2. The SUDOI architecture.

fold. First, the SUDOI architecture emphasizes the cooperation between IPC and TC to overcome the interworking obstacles from *multi-layer* overlaid networks, and it effectively realizes *vertical integration*. Second, in order to provide the end-to-end QoS, multi-stratum resources of access, transport network and data center can be merged through controllers' interaction with *horizontal merging*, while achieving global CSO of network and application resources.

Optical Access Controller

The OC is used to maintain and control multiple passive optical network (PON) domains remotely. The dynamic bandwidth assignment among multiple domains is scheduled in a unified way to implement QoS guarantee and resource optimization.

Flow Monitor: The external communication module to interwork the information with OF-OLTs and compile the statistics of flow via resource arrangement module.

Flow Schedule: Decides to allocate and adjust the bandwidth to guarantee the QoS, and deploys the SA-FS function. It can interact application usage with AC via an optical-application interface (OAI) and schedule the bandwidth among various OF-OLTs through IPC via an IP-optical interface (IOI).

Resource Arrangement: Can allocate the access network resources efficiently and configure the corresponding parameters (e.g., DBA and operation, administration, and management, OAM) according to user requirements through OFP.

Network Infobase: Used to conserve the network resource information and update the results when the resource assignment is performed successfully.

IP Controller

The IPC is responsible for analyzing IP flow status with flow resources maintained in an IP network, and performs the multi-layer resource vertical integration.

IP Flow Monitor: Responsible for compiling the status of OF-Routers via flow table control and interworks with the application resources to recognize the data center services through an IP-application interface (IAI).

Resource Integration Control: The core module with SA-FE function, which decides to offload the flow into an optical network after completing the SA-FE and provides resource integration requests to TC via an IP-transport interface (ITI), including source, destination, and QoS parameters.

Flow Table Control: Used to send a flow modification message to offload the flow by updating flow entries of routers, while utilizing the relatively vacant optical network resources.

Transport Controller

The TC exploits optical network information abstracted from a physical network, and performs lightpath provisioning in optical networks accordingly to achieve CSO of network and application resources and multi-layer resource integration.

Resource Integration Agent: The core module for messages scheduling to other modules. It can receive an integration request via ITI and forward it to the path computation element (PCE), returning a success reply containing the lightpath information.

Path Computation Element: Capable of computing an end-to-end network path or route based on a network graph and applying computational constraints.

CSO Control: An engine for the location selection of an application using CSO according to the status of optical network and data center resources. The latter is perceived from a CSO agent in an AC via application-transport interface (ATI). The SA-RP function is deployed in a CSO control module.

Network Abstraction: Can abstract and manage the network topology from physical network through PCE calculation, and provide abstracted resource information to CSO control.

Spectrum Control and Monitor: Can monitor physical network elements and control the spectrum bandwidth and modulation format in the underlying network. The lightpaths are provisioned by controlling all corresponding OF-OSs using OFP.

Database Management Conserves: The lightpath information after the connection setup from the PCE and the abstract topology information from the network abstraction module.

Application Controller

The centralized AC is responsible for monitoring and allocating compute/storage resources in data centers and arranging data center services.

CSO Agent: The communication module to interact with the OC, TC, and IPC through OAI, ATI, and IAI, and provide the application resource utilization periodically or based on an event-based trigger.

Application Service Management: Monitors and maintains virtual application resources obtained from data centers.

Database: Stores the abstract application information from data centers.

Extension of OpenFlow Protocol

In a packet-switched network, OpenFlow abstracts the data plane as a flow entry, which is defined as rule, action, and stats. It represents the packet's characteristic and the action of the switch [8]. For SUDOI control of optical networks, a flow entry of OpenFlow v1.3 in an optical flow table is extended. In SUDOI, the rule is extended as an in/out port (8 bits) and PON label (guard time, bandwidth, time slot, and ONU number occupying 4 × 8 bits space), elastic label (channel spacing, grid, central frequency, and spectrum bandwidth use 4 × 8 bits space) and intra-data-center label (channel space, lambda, and time slot occupy 3 × 8 bits space), which are the main characteristics of the access, transport, and data center optical network. The action of an optical node mainly includes add, switch, drop, and configure to set up a path to the port/label with a specified adaption function (e.g., modulation format), and delete a path to restore the original state of equipment. Various combinations of rule and action are used to realize the control of optical nodes. The statistics function is responsible for monitoring the flow property to provide service provisioning. Note that we extend the feature request and feature reply messages to detect and monitor the status information in the SUDOI scenario. For flow monitoring and detecting an optical node in real time, the controller sends a flow monitor request to each node using an extended *OFPT_FEATURES_REQUEST* message sent periodically via the OFP, while obtaining the flow status information (bandwidth and QoS parameter occupy 2 × 8 bits space) with extended *OFPT_FEATURES_REPLY* message from them.

Technical and Business Context of SUDOI

In a traditional communication ecosystem, the control of access, transport networks, and data centers is usually separately deployed in ubiquitous data center interconnection. The service request can only be exchanged across the interface among those networks. Due to the development of user requirement and company scale, a single business party is forced to extend its types of services to adapt to the environment and promote the competitiveness. Service providers can establish the optical transport network to interconnect their data centers (e.g., Google). Meanwhile, network operators also establish data centers to provide storage and computing resources for their clients (e.g., China Mobile). In the standardization, many network vendors, operators, and service providers focus on the interfaces among the multi-domain controllers, including the southbound and northbound interfaces. The interoperability test among tens of vendors and operators has been addressed by ONF and OIF. If the interfaces among the controllers are presented as the standard, SUDOI can also be performed by different business parties in real business environment.

The interoperability test among tens of vendors and operators has been addressed by ONF and OIF. If the interfaces among the controllers are presented as the standard, SUDOI can also be performed by different business parties in real business environments.

Cooperation Procedure for SUDOI in Various Service Modes

Cooperation among controllers in various service modes is one of the key issues for SUDOI. It helps to achieve the scenarios of user-access-oriented, inter-data-center, and intra-data-center interconnection.

User-Access-Oriented Interconnection Service Mode

The user-access-oriented interconnection service mode between the user and data center could ensure the end-to-end QoS of a user who

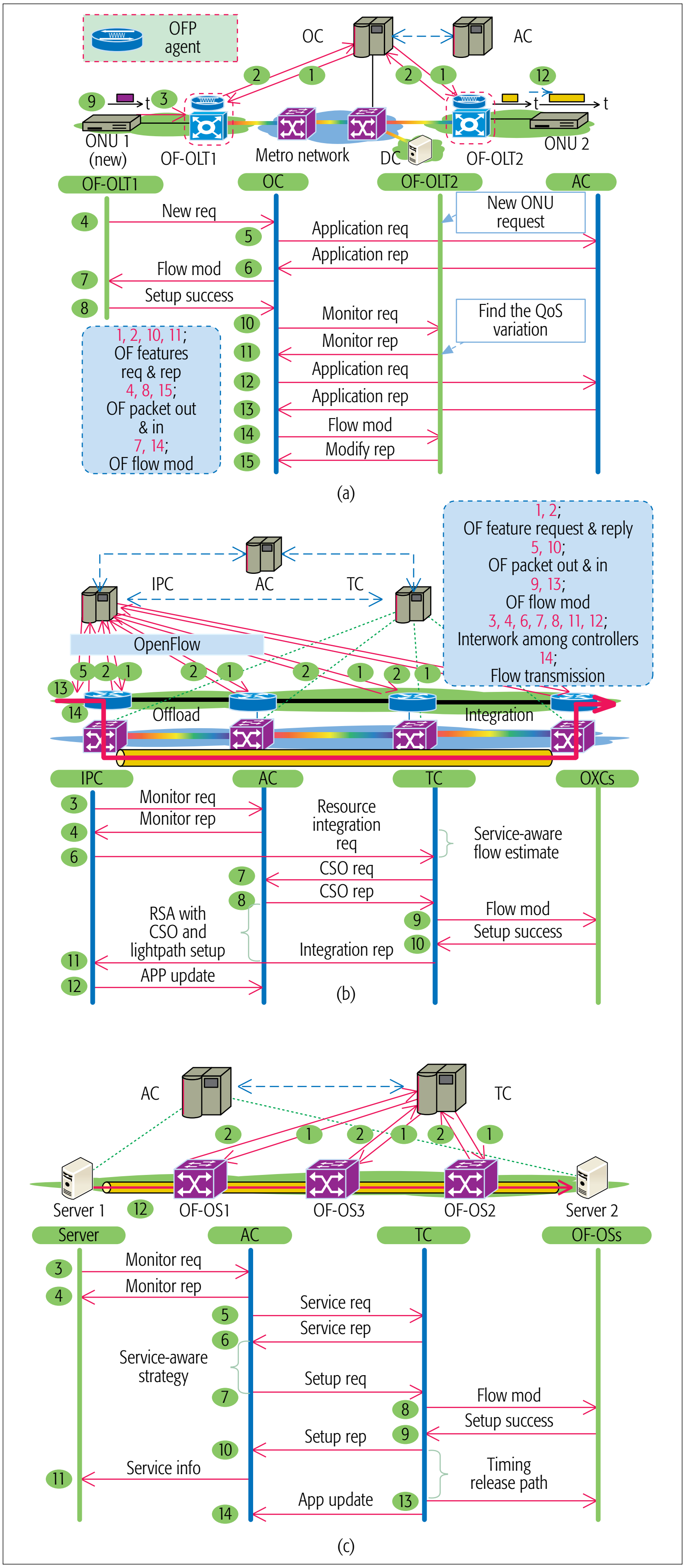


Figure 3. Cooperation procedure in a) user-access-oriented interconnection; b) multiple layer resource integration; c) intra-data center service modes.

requires data center services, which is shown in Fig. 3a. For the flow detecting PON in real time, the OC sends a flow monitor request to each OF-OLT periodically via OFP, while obtaining the flow status information from them (as shown in steps 1–2 in Fig. 3a). We consider two scenarios in the analysis for simplification, that is, new a ONU accesses to the network, and an existing ONU changes the requirement. If a new ONU1 accesses OF-OLT1, the device forwards the request to the OC (steps 3–4). The service-aware flow schedule subschema can be completed in the OC considering the QoS guarantee, current network condition, and data center application usage, which is perceived from the interaction with AC (steps 5–6). Then the OC proceeds to send a new flow entry to arrange the time slot, bandwidth, and corresponding configuration for the new ONU1 (step 7). When the service request is accommodated successfully, the OF-OLT1 can send the setup success reply to the OC (steps 8–9). Once the OC obtains the QoS requirement of the existing ONU2 changed from the monitoring message (steps 10 and 11), it will send the bandwidth granularity variation message to OF-OLT2 for updating flow entry and realize the bandwidth adjustment flexibly (steps 12–14). Then OF-OLT2 responds the modification reply message to the OC when the bandwidth adjustment is completed (step 15).

Multi-Layer Resource Integration Service Mode

The multi-layer resource integration service mode for inter-data-center can utilize IP and optical resources effectively. As shown in Fig. 3b, the IPC detects the status of a flow through interworking with each OF-Router (steps 1–2), while interacting with the data center resources with AC (steps 3–4). Then the IPC evaluates the latest flow status with service-aware flow estimation subschema as the assessment value for the coming flow (step 5). The multi-layer resource integration control can be triggered in IPC when the new estimation has exceeded the threshold, and then sends the request to TC (step 6). The TC computes a path considering CSO of optical and application resources cooperating with AC (steps 7–8), and then proceeds to set up an end-to-end lightpath (step 9). When the TC receives a setup success reply from the last OF-OS (step 10), it responds with the integration reply to IPC with provisioning lightpath and abstracted optical resources information (step 11). After that, the IPC sends a setup message to the router with a buffered packet such that the flow is offloaded to the optical network for utilizing the multi-layer resources effectively (steps 12–14).

Intra-Data-Center Service Mode

The intra-data-center service mode can globally optimize optical network and application resources to accommodate the burstiness of intra-data-center application. In order to enhance the data center service and reduce the realized complexity of interworking, we use one controller (i.e., the TC) to unify the control and scheduling of the transport network and fabric, and manage the optical resources to optimize the resource utilization globally in the scenario of inter- and intra-data-center optical interconnection. In Fig. 3c, the TC monitors the network information of optical nodes by the interaction with OF-OSs (steps 1–2), while the application stratum resources are maintained in AC (steps 3–4). Through interworking application resources, the AC sends the service request to TC to ask for the network resources information (steps 5–6). The service-aware resource provisioning with CSO in TC can be completed to choose the optimal destination node based on various service parameters and network resources utilization, and then responds to the setup request (step 7). The end-to-end lightpath is set up by controlling the corresponding OF-OS (steps 8–9), and TC responds with the setup reply to AC with provisioning lightpath (steps 10-12). Meanwhile, TC records the setup time and service duration time. The release lightpath procedure is reckoned by service time immediately (steps 13–14).

Service-Aware Schedule Scheme in SUDOI

To enhance resource utilization and QoS, we consider the service information and estimate the current resource status to guarantee the optimal accommodation among various kinds of resources. Therefore, the service-aware resource schedule scheme is proposed based on SUDOI, which includes three subschemas in various application scenarios, that is, service-aware flow scheduling (SA-FS), service-aware flow estimation (SA-FE), and service-aware resource provisioning (SA-RP) subschemas.

In the user-access-oriented scenario, the SA-FS subschema decides to allocate and adjust the bandwidth for requests according to QoS priority, flow status, and resource utilization selected from controllers. Note that the SA-FS function is positioned in the flow schedule module of the OC. Under a heavy traffic load scenario, the optical network can offer highly available, cost-effective and energy-efficient connectivity services by provisioning a spectrum path. Compared to the optical network, the IP network is more suitable for supplying small granularity service flows due to its flexibility and convenience. Therefore, based on various network conditions, the SA-FE subschema decides which resource can be assigned for the new traffic flow in the multi-layer resource integration scenario in inter-data center [4], which is deployed in the resource integration control module of IPC. It can analyze the previous flow statuses, estimate the network condition using the resources expectation, and determine whether the new flow needs to be provisioned through the resource integration process. If the value of the new flow returned by the SA-FE exceeds the threshold, the IP resources become scarce in the overloaded IP network. Resource integration is triggered and offloads the flow to the optical network to utilize the relatively vacant optical network resources. In the intra-data-center scenario, the SA-RP subschema realized in the CSO control module of TC could comprehensively consider the status of data center resources (e.g., computing and storage utilization) and optical network resources (e.g., bandwidth and latency) to select the data center destination for requests. Then TC can complete the path computation in the connection and service parameters constraints, and perform spectrum assignment with first fit for the computed path and the lightpath provisioning by OFP.

Experimental Setup and Results Discussion

Experimental Setup

The feasibility and efficiency of the SUDOI architecture are evaluated based on the Optimization as a Service (OaaS) testbed [15], which is shown in Fig. 4. In the data plane, four OpenFlow-enabled optical switches are equipped with commercial reconfigurable optical add/drop multiplexers (ROADMs) using Huawei Optix OSN 6800, while four field programmable gate arrays (NetFPGAs) are deployed as corresponding OF-Routers in the upper layer. Two groups of 10GEPON equipment are deployed on the access side. In each group, two optical network units (ONUs) are interconnected with the OF-OLT by an optical splitter. To emulate an intra-data-center scenario, four optical switches are equipped with FOS, while two cyclic arrayed waveguide grating (AWG) cards are deployed as on the core side. We use Open VSwitch (OVS) as the software OFP agent according to the application programming interface (API) to interact between the controller and optical nodes, and control the hardware through OFP. In addition, the OFP agents are also used to emulate other optical nodes in the data plane to support SUDOI with OFP. Data centers and the other nodes are realized on virtual machines (VMs) created by VMware ESXi V5.1 running on IBM X3650 servers. Since each VM has its own CPU and storage resources, the operating system makes it easy to set up an experimental topology for large-scale extension. For the OpenFlow-based SUDOI control plane, the IPC server is deployed by means of two VMs for flow monitoring and integration control, while CSO and PCE computation, spectrum control, and network abstraction are deployed in the TC server. The AC server is used as the CSO agent to monitor the application resources from data center networks. The OC is deployed in two VMs for flow scheduling and monitoring, and resource allocation. The user plane is deployed in a server and deploys the service information generator to implement batch data center services for the experiments. Note that the AC manages the data center servers and their application resources through the VMware software, which can gather the CPU and storage resources, and configure and control the VMs via internal APIs in the data centers. The TC, OC, and IPC are implemented based on Opendaylight to realize the optical and IP network management in the testbed.

Three Application Scenarios

Based on the testbed, we have verified SUDOI experimentally in three typical scenarios. The signaling procedure in the multi-layer resource integration scenario is presented through a Wireshark capture inserted in IPC and TC, which refers to the related sequence diagram in Fig. 3b. The *features request* message is responsi-

> The service-aware resource schedule scheme is proposed based on SUDOI , which includes three subschemas in various application scenarios: service-aware flow scheduling, service-aware flow estimation, and service-aware resource provisioning subschemas.

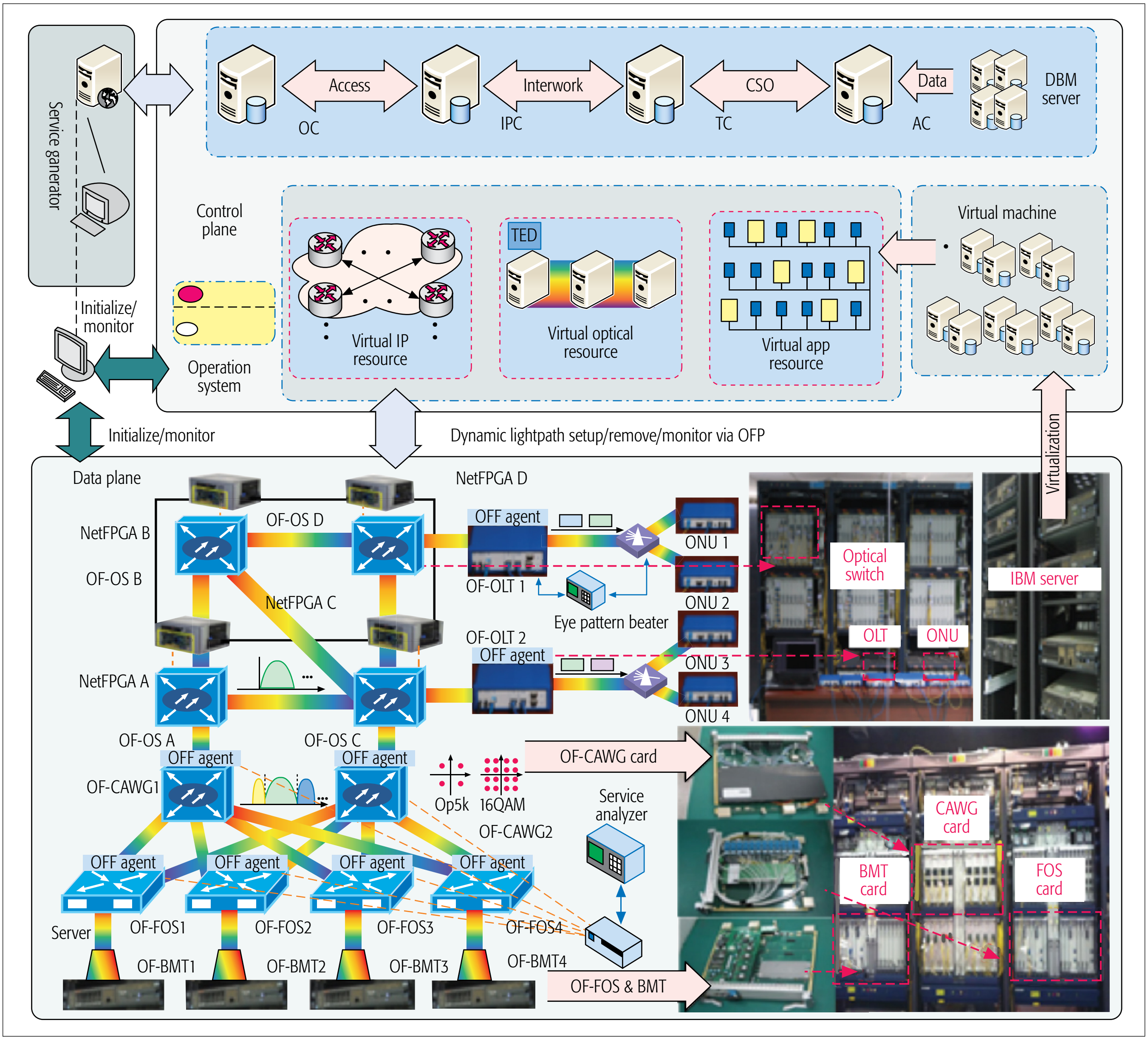
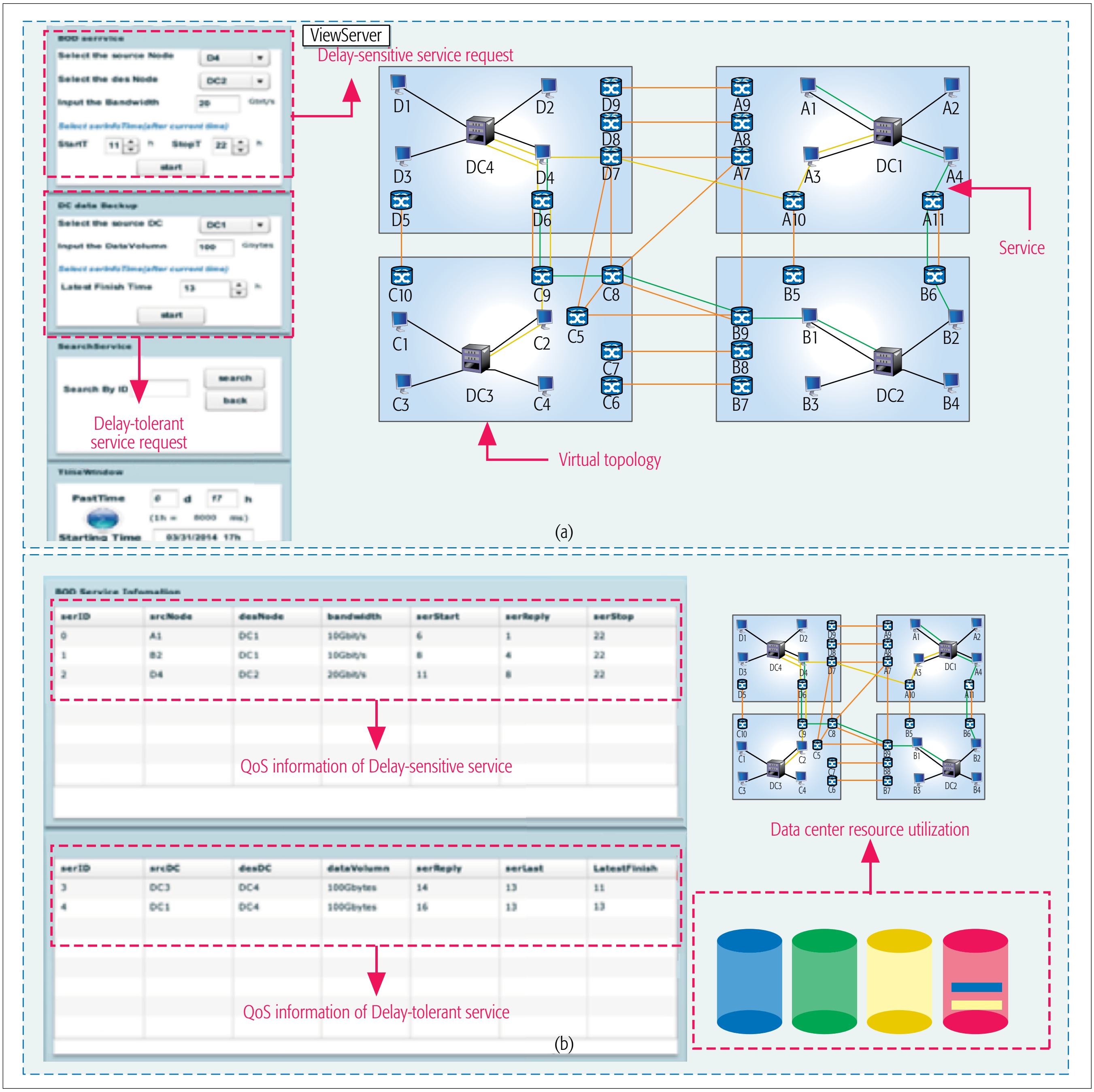


Figure 4. Experimental testbed and demonstrator setup.

ble for SA-FE monitoring by regularly querying OF-Routers about the current flow status (as shown in step 1, Fig. 3b). IPC receives the flow information from OF-Routers via *features reply* (step 2). Then IPC obtains the service usage of application resources through the interworking between the IPC and AC via UDP message (steps 3–4). Here, we use UDP as the interworking message format among various controllers to simplify the procedure and reduce the performance pressure on controllers. Once a new request arrives through *packet in* from OF-Router (step 5), the integration procedure is triggered when the new estimation obtained by SA-FE has exceeded the threshold, and then sends the request to TC (step 6). The TC computes a path considering CSO and then provisions a lightpath to control all corresponding OF-OSs through *flow mod* (step 9). Receiving a setup success reply from OF-OSs via *packet in* (step 10), TC responds to IPC via UDP with provisioning lightpath information (step 11). Then the IPC sends *flow mod* to offload the flow by updating the flow entries of routers (step 13), while updating the application usage with UDP to keep the synchronization (step 12).

We also verify the user-access-oriented mode between user and data center. The Wireshark capture of the message sequence is illustrated and deployed in the OC, which corresponds to the procedures depicted in Fig. 3a. Once a new ONU accesses the OF-OLT for service provisioning, the OC receives a *packet in* message to notice the request from the OF-OLT (step 4), performs an SA-FS subschema, and sends the *flow mod* to add a new flow entry for the new ONU (step 7). If it obtains the QoS of the existing ONU changed from a *features reply* (step 11), the OC sends the bandwidth granularity variation through the *flow mod* to the corresponding OF-OLT for flow entry update (step 14), then receives the status report through *packet in* (step



BOD service

Select the source Node: D4

Select the des Node: DC2

Input the Bandwidth: 20 Gbit/s

Select serInfoTime(after current time)

StartT: 11 h StopT: 22 h

start

DC data Backup

Select the source DC: DC1

Input the DataVolumn: 100 Gbytes

Select serInfoTime(after current time)

Latest Finish Time: 13 h

start

SearchService

Search By ID

search

back

TimeWindow

PastTime: 0 d 17 h

(1h = 8000 ms)

Starting Time: 03/31/2014 17h

BOD Service Infomation

serID	srcNode	desNode	bandwidth	serStart	serReply	serStop
0	A1	DC1	10Gbit/s	6	1	22
1	B2	DC1	10Gbit/s	8	4	22
2	D4	DC2	20Gbit/s	11	8	22

serID	srcDC	desDC	dataVolumn	serReply	serLast	LatestFinish
3	DC3	DC4	100Gbytes	14	13	11
4	DC1	DC4	100Gbytes	16	13	13

Figure 5. The front-end graphical user interface of the testbed: a) topology tab; b) information tab.

15). The result is emphasized for the intra-data-center scenario and illustrates the capture of the message exchange through Wireshark deployed in the TC, which refers to the sequence depicted in Fig. 3c. The TC performs the SA-RP subschema and then provisions the lightpath to control the OF-OS through *flow mod* (step 8). After timing the service time, the path is released by *flow mod* (step 13). Note that the existing OpenFlow messages have the original function. For simplicity, these messages are reused to simplify the implementation in this article. The front-end interface of the testbed for resources visualization is shown partly in Figs. 5a and 5b. It can verify the resource management and QoS enforcement between users and data centers through the core network. Two kinds of data center services are deployed to prove the feasibility of SUDOI: delay-tolerant and delay-sensitive services. We can see that three delay-sensitive services (shown in green) and two delay-tolerant services (shown in yellow) are accommodated with different resources and QoS, which can be presented in the virtual topology of the interface. Figure 5b indicates the current application resource status of data center servers and corresponding QoS information of services, which include detailed path, service bandwidth, and related service time.

Performance Analysis

We also evaluate the performance of the service-aware schedule scheme under the heavy traffic load scenario of SUDOI, and compared

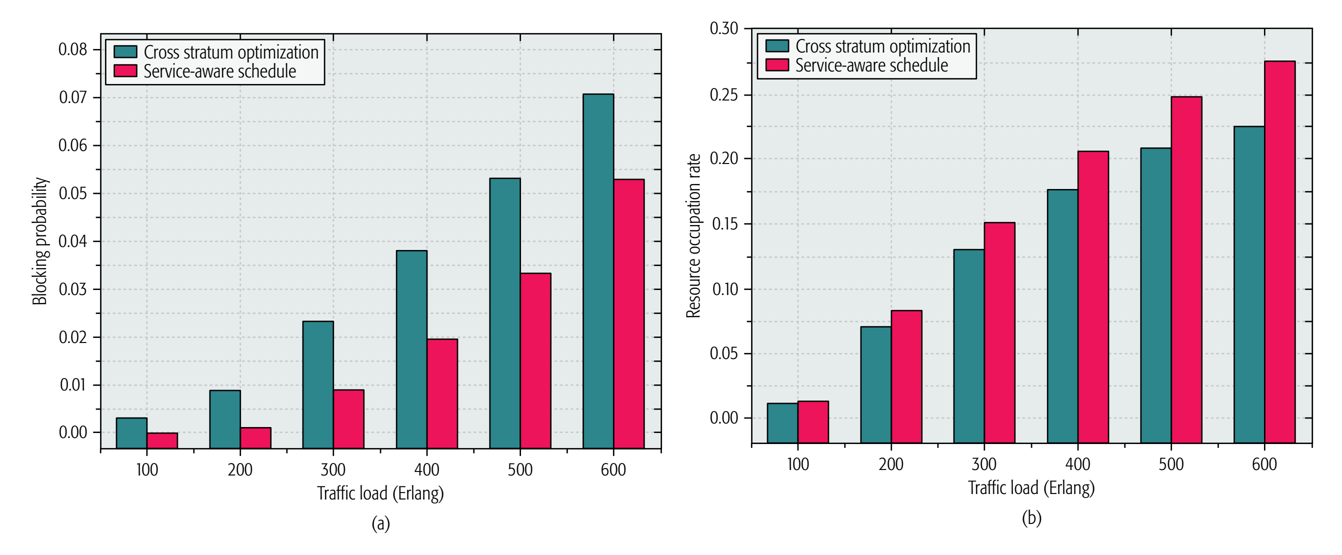


Figure 6. a) Blocking probability; b) resource occupation rate among two schemes under the heavy traffic load scenario.

it with the CSO scheme [15] through VMs. The traffic requests to the data center are established with spectrum randomly from 500 Mb/s to 40 Gb/s, while the service application usage in the data center is selected randomly from 1 to 0.1 percent for each application demand. They arrive at the network following a Poisson process, and results have been extracted through the generation of 100,000 demands per execution. The traffic model depends on the pattern of information transfer in the network. For instance, the arriving traffic is cached in the electrical buffer of an IP router in a connectionless IP network, which follows the self-similarity model. In a realistic scenario of an optical network operator, the traffic demand is assigned from the network management system, which is emulated by the service generator in our experiment. The arrival rate and departing rate of the service follow a Poisson process. Therefore, a Poisson distribution is assigned to the arriving traffic in SUDOI. Figure 6 compares the performance of two schemes in terms of blocking probability and resource occupation rate. The service-aware schedule scheme reduces blocking probability more effectively than the CSO scheme, especially when the network is heavily loaded. The reason is that the service-aware schedule scheme includes three subschemas in the various scenarios of ubiquitous data center optical interconnection. In the user-access-oriented interconnection scenario, the SA-FS realizes the bandwidth allocation and adjustment according to QoS priority, flow status, and resource utilization. In the multi-layer inter-data-center scenario, the SA-FE takes into account the service-aware statistics of the previous flows and integrates multi-layer resources with offloading heavy flows into the optical network. In the intra-data-center scenario, the SA-RP implements global optimization considering both data center application and network resources integrally. The CSO scheme performs destination optimization considering optical network and application resources inside the data center. It leads to various kinds of resources being scheduled on the side edge of the user (e.g., QoS priority) and multi-layer (i.e., IP and optical) resource management not being involved in the CSO scheme. It is hard to realize end-to-end resource optimization. In Fig. 6b, the service-aware schedule scheme outperforms the other scheme in the resource occupation rate significantly. The main reason is that more resources can be occupied when the blocking probability is lower.

Conclusion

To meet the QoS requirement of extensive user access, this article presents a novel SUDOI architecture in ubiquitous data center optical interconnection, which can allow the CSO of application and optical network resources and multi-layer resource integration from the perspective of heterogeneous cross-stratum and multi-layer networking modes, respectively. The functional modules of the architecture and their cooperation procedure in various service modes are described and investigated. The performance of SUDOI is verified on our OaaS testbed for data center services. We evaluate its performance under a heavy traffic load scenario and compare it to the CSO scheme. Numerical results show that SUDOI with a service-aware schedule scheme can utilize optical network and application resources effectively in ubiquitous data center optical interconnection without increasing blocking probability.

Our future works for SUDOI include several aspects. The issue of network survivability of SUDOI could be discussed in the near future. Network virtualization in SUDOI on the OaaS testbed should be studied. Also, the experimental comparison between our architecture and other projects will be performed.

Acknowledgment

This work has been supported in part by NSFC project (61501049, 61271189), the Fundamental Research Funds for the Central Universities (2015RC15), the Fund of State Key Laboratory of Information Photonics and Optical Communications (BUPT), P. R. China, the Science and Technology Project of State Grid Corporation of

China (SGIT0000KJJS1500008), and funded by the State Key Laboratory of Advanced Optical Communication Systems Networks, China.

References

[1] C. Kachris and I. Tomkos, "A Survey on Optical Interconnects for Data Centers," *IEEE Commun. Surveys & Tutorials*, vol. 14, no. 4, Oct. 2012, pp. 1021–36.

[2] M. Jinno *et al*., "Spectrum-Efficient and Scalable Elastic Optical Path Network: Architecture, Benefits, and Enabling Technologies," *IEEE Commun. Mag*., vol. 47, no. 11, Nov. 2009, pp. 66–73.

[3] O. Gerstel *et al*., "Elastic Optical Networking: A New Dawn for the Optical Layer?" *IEEE Commun. Mag*., vol. 50, no. 2, Feb. 2012, pp. S12–S20.

[4] H. Yang *et al.*, "Multi-Stratum Resource Integration for OpenFlow-based Data Center Interconnect [Invited]," *J. Opt. Commun. Netw*., vol. 5, no. 10, Oct. 2013, pp. A240–A248.

[5] T. Szyrkowiec *et al*., "First Field Demonstration of Cloud Datacenter Workflow Automation Employing Dynamic Optical Transport Network Resources Under Openstack and Openflow Orchestration," *Optics Express*, vol. 22, no. 3, Jan. 2014, pp. 2595–2602.

[6] S. Das, G. Parulkar, and N. McKeown, "Why OpenFlow/SDN Can Succeed Where GMPLS Failed," *Proc. ECOC*, paper Tu.1.D.1, Sept. 2012.

[7] M. Channegowda *et al*., "Experimental Demonstration of an Openflow based Software-Defined Optical Network Employing Packet, Fixed and Flexible DWDM Grid Technologies on an International Multi-domain Testbed," *Optics Express*, vol. 21, no. 5, Mar. 2013, pp. 5487–98.

[8] R. Casellas *et al*., "An Integrated Stateful PCE/OpenFlow Controller for the Control and Management of Flexi-Grid Optical Networks," *Proc. OFC/NFOEC,* paper OW4G.2, Mar. 2013.

[9] H. Yang *et al*., "Performance Evaluation of Multi-Stratum Resources Integrated Resilience for Software Defined Inter-Data Center Interconnect," *Optics Express*, vol. 23, no. 10, May 2015, pp. 13,384–98.

[10] L. Liu *et al.*, "OpenSlice: an OpenFlow-Based Control Plane for Spectrum Sliced Elastic Optical Path Networks," *Proc. ECOC*, paper Mo.2.D.3, Sept. 2012.

[11] F. Paolucci *et al*., "OpenFlow-Based Flexible Optical Networks with Enhanced Monitoring Functionalities," *Proc. ECOC*, paper Tu.1.D.5, Sept. 2012.

[12] H. Yang *et al*., "CSO: Cross Stratum Optimization for Optical as a Service," *IEEE Commun. Mag.*, vol. 53, no. 8, Aug. 2015, pp. 130–39.

[13] L. Liu *et al.*, "Field Trial of an OpenFlow-Based Unified Control Plane for Multi-Layer Multi-Granularity Optical Switching Networks," *J. Lightwave Tech.*, vol. 31, no. 4, Feb. 2013, pp. 506–14.

[14] N. Cvijetic *et al*., "SDN and OpenFlow for Dynamic Flex-Grid Optical Access and Aggregation Networks," *J. Lightwave Tech*., vol. 32, no. 4, Feb. 2014, pp. 864–70.

[15] H. Yang *et al*., "Performance Evaluation of Data Center Service Localization Based on Virtual Resource Migration in Software Defined Elastic Optical Network," *Optics Express*, vol. 23, no. 18, SepT. 2015, pp. 23059–71.

Biographies

Hui Yang [M] received his Ph. D degree in communication and information systems from Beijing University of Posts and Telecommunications (BUPT), China, in 2014. He is currently an assistant professor with the Institute of Information Photonics and Optical Communications at BUPT. His main research interests include software defined networks, cross-stratum optimization, data center interconnect, wireless and optical access networks, flexi-grid optical networks, and so on. He has authored or coauthored more than 60 papers related to the previously mentioned research topics in prestigious international journals and conferences, and he is the first author of more than 30 of them. He served as Session Chair of CHINACOM '14, and received the Best Paper Award at NCCA '15. He is also an active reviewer or TPC member for several international journals and conferences..

Jie Zhang is a professor and vice dean of the Institute of Information Photonics and Optical Communications at BUPT. He is sponsored by over 10 projects of the Chinese government; he has published 8 books and more than 100 articles. He also has had seven patents granted. He has served as a TPC member for ACP 2009, PS 2009, ONDM 2010, and so on. His research focuses on optical transport networks, packet transport networks, and so on.

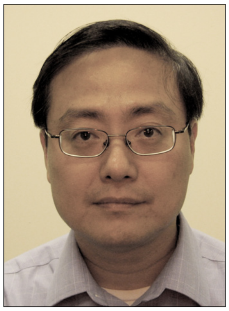
Yongli Zhao is currently a lecturer at the Institute of Information Photonics and Optical Communications at BUPT. He received his B.S. degree in communication engineering and Ph.D. degree in electromagnetic field and microwave technology from BUPT in 2005 and 2010, respectively. He has had more than 100 articles published. His research focuses on wavelength switched optical networks, optical transport networks, packet transport networks, and so on.

Jianrui Han is a system engineer in the Advanced Technology Department of the Wireline Network Business Unit and a leader in providing next generation telecommunications networks at Huawei Technologies Ltd. She joined Huawei in 2001. Her research topics are GMPLS control plane, WSON, and Transport SDN.

Yi Lin is research engineer at Huawei Technologies, Ltd. Co. He received his B.S. degree in electronical information science and technology in 2005 and his M.S. degree in radio physics in 2007 from Sun Yat-Sen University, and joined Huawei in 2007. His main research topic is intelligent control of transport networks, including ASON/GMPLS, PCE, transport SDN, and so on.

Young Lee received his B.A. degree in applied mathematics from the University of California at Berkeley in 1986, his M.S. degree in operations research from Stanford University, California, in 1987, and his Ph.D. degree in decision sciences and engineering systems from Rensselaer Polytechnic Institute, Troy, New York, in 1996. He is currently a principal technologist at Huawei Technologies USA Research Center, Plano, Texas. He leads optical transport control plane technology research and development.

AUTOMOTIVE NETWORKING AND APPLICATIONS


Wai Chen

Luca Delgrossi

Timo Kosch

Tadao Saito

In this 16th issue of the Automotive Networking and Applications Series, we are pleased to present two articles that address information-centric networking for connected vehicles and network engineering for real-time automotive networks.

In the connected vehicle ecosystem, a large amount of information and safety-critical data will be exchanged among vehicles, roadway infrastructures, and pedestrians in a highly dynamic environment characterized by fluctuating wireless links and vehicle mobility. The host-centric IP-address-based model of networking, designed with the end-to-end connectivity principle in mind, is challenged to work in this dynamic roadway environment and is not well suited for the localized nature of many cases of vehicular communications, where the focus is on specific road segments (e.g., the vicinity of a hazard, a point of interest) regardless of the identity or the IP address of any specific vehicle passing by. The first article, "Information-Centric Networking for Connected Vehicles: A Survey and Future Perspectives" by M. Amadeo *et al.*, discusses the applicability of the information-centric networking (ICN) paradigm as a networking solution for connected vehicles. The authors first review core functionalities of ICN and survey related research results on the adaptations and customizations of the baseline ICN architecture to better match dynamic vehicular environments. Through their analysis, the authors show that the native design principles of ICN are well suited for the main features of vehicular ad hoc networks and their applications. The authors conclude with a discussion of the open challenges related to large-scale deployment of ICN and coexistence with other vehicular networking technologies, among others.

With the advances in x-by-wire applications that have strict latency and reliability requirements, formal verification of end-to-end timing constraints on networks has become an important part in the design process of vehicles. The second article, "Network Engineering for Real-Time Networks: Comparison of Automotive and Aeronautic Industrial Approaches" by F. Geyer and G. Carle, first reviews the prevailing network architectures and technologies used by the automotive and aeronautic industries for x-by-wire applications. The authors then present and compare two representative mathematical frameworks, schedulability analysis and network calculus, that are used by each industry to formally verify the end-to-end latency behavior of a network. Based on empirical evaluation results of two use cases, the authors highlight the trade-offs between the two frameworks and provide guidelines on a suitable framework to use depending on the types of network deployed.

We thank all contributors who submitted manuscripts for this Series, as well as all the reviewers who helped with thoughtful and timely reviews. We thank Dr. Osman Gebizlioglu, Editor-in-Chief, for his support, guidance, and suggestions throughout the process of putting together this issue. We also thank the IEEE publication staff, particularly Ms. Charis Scoggins and Ms. Jennifer Porcello, for their assistance and diligence in preparing the issue for publication.

BIOGRAPHIES

WAI CHEN (waichen@ieee.org) received his B.S. degree from Zhejiang University, and M.S., M.Phil., and Ph.D. degrees from Columbia University, New York. He is a chief scientist of China Mobile Research and general manager of the China Mobile Internet-of-Things Research Institute. Previously he was vice president and group director of ASTRI, Hong Kong; and a chief scientist and director at Telcordia (formerly known as Bellcore), New Jersey. While at Telcordia, he led a vehicular communications research program over 10 years in collaboration with a major automaker on automotive networking technologies for vehicle safety and information applications. He has been Principal Investigator of several government funded projects on advanced networking technologies research. He was the General Co-Chair for the IEEE Vehicular Networking Conference (2009–2013) and a Guest Editor for the Special Issue on Vehicular Communications and Networks for the *IEEE Journal on Selected Areas in Communications* (2011). He also served as a Guest Editor for the Special Issue on Inter Vehicular Communication of *IEEE Wireless Communications* (2006), an IEEE Distinguished Lecturer (2004–2006), Co-Chair of the Vehicle-to-Vehicle Communications Workshop (2005–2008) co-located with the IEEE Intelligent Vehicles Symposium, and Co-Chair of the IEEE Workshop on Automotive Networking and Applications (2006–2008) co-located with IEEE GLOBECOM.

LUCA DELGROSSI is manager of the Vehicle-Centric Communications Group at Mercedes-Benz Research & Development North America Inc., Palo Alto, California. He started as a researcher at the International Computer Science Institute of the University of California at Berkeley and received his Ph.D. in computer science from the Technical University of Berlin, Germany. He served for many years as a professor and associate director of the Centre for Research on the Applications of Telematics to Organizations and Society of the Catholic University at Milan, Italy, where he helped create and manage the Master's in Network Economy (MiNE) program. In the area of vehicle safety communications, he coordinated the Dedicated Short Range Communications (DSRC) Radio and On-Board Equipment work orders to produce the DSRC specifications and build the first prototype DSRC equipment as part of the Vehicle Infrastructure Integration (VII) initiative of the U.S. Department of Transportation.

TIMO KOSCH works as a team manager for BMW Group Research and Technology where he is responsible for projects on distributed information systems, including such topics as cooperative systems for active safety and automotive IT security. He has been active in a number of national and international research programs, and serves as coordinator for the European project COMeSafety, co-financed by the European Commission. He is also currently heading the system development for a large German Car2X field test. For more than three years, he chaired the Architecture working group and was a member of the Technical Committee of the Car-to-Car Communication Consortium. He studied computer science and economics at Darmstadt University of Technology and the University of British Columbia in Vancouver with scholarships from the German National Merit Foundation and the German Academic Exchange Service. He received his Ph.D. from the Computer Science Faculty of the Munich University of Technology.

TADAO SAITO [LF] received a Ph. D degree in electronics from the University of Tokyo in 1968. Since then he has been a lecturer, an associate professor, and a professor at the University of Tokyo, where he is now a professor emeritus. Since April 2001 he has been chief scientist and CTO of Toyota InfoTechnology Center, where he studies future ubiquitous information services around automobiles. He has worked in a variety of subjects related to digital communication and computer networks. His research includes a variety of communication networks and their social applications such as ITS. Included in his past study, in the 1970s he was a member of the design group of the Tokyo Metropolitan Area Traffic Signal Control System designed to control 7000 intersections under the Tokyo Police Authority. Now he is Chairman of the Ubiquitous Networking Forum of Japan working on a future vision of the information society. He is also Chairman of the Next Generation IP Network Promotion Forum of Japan. From 1998 to 2002 he was Chairman of the Telecommunication Business Committee of the Telecommunication Council of the Japanese government and contributed to regulatory policy of telecommunication business for broadband network deployment in Japan. He is also the Japanese representative to the International Federation of Information Processing General Assembly and Technical Committee 6 (Communication System). He is an honorary member and fellow of IEICE of Japan.

Call for Papers

IEEE Transactions on Molecular, Biological, and Multiscale Communications

Communications Beyond Conventional Electromagnetism

This journal is devoted to the principles, design, and analysis of signaling and information systems that use physics beyond conventional electromagnetism, particularly for small-scale and multi-scale applications. This includes: molecular, quantum, and other physical, chemical and biological (and biologically-inspired) techniques; as well as new signaling techniques at these scales.

As the boundaries between communication, sensing and control are blurred in these novel signaling systems, research contributions in a variety of areas are invited. Original research articles on one or more of the following topics are within the scope of the journal: mathematical modeling, information/communication-theoretic or network-theoretic analysis, networking, implementations and laboratory experiments, systems biology, data-starved or data-rich statistical analyses of biological systems, industrial applications, biological circuits, biosystems analysis and control, information/communication theory for analysis of biological systems, unconventional electromagnetism for small or multi-scale applications, and experiment-based studies on information processes or networks in biology. Contributions on related topics would also be considered for publication.

Editor-in-Chief

Urbashi Mitra University of Southern California, USA

Associate Editor-in-Chief

Andrew W. Eckford York University, Canada

Submit today!

https://mc.manuscriptcentral.com/tmbmc

Information-Centric Networking for Connected Vehicles: A Survey and Future Perspectives

Marica Amadeo, Claudia Campolo, and Antonella Molinaro

The authors advocate for a paradigm shift from traditional IP-based networking toward the groundbreaking information-centric networking. They scrutinize the applicability of this paradigm in vehicular environments by reviewing its core functionalities and the related work. The analysis shows that information-centric networking is positioned to meet the challenging demands of vehicular networks and their evolution.

ABSTRACT

In the connected vehicle ecosystem, a high volume of information-rich and safety-critical data will be exchanged by roadside units and onboard transceivers to improve the driving and traveling experience. However, poor-quality wireless links and the mobility of vehicles highly challenge data delivery. The IP address-centric model of the current Internet barely works in such extremely dynamic environments and poorly matches the localized nature of the majority of vehicular communications, which typically target specific road areas (e.g., in the proximity of a hazard or a point of interest) regardless of the identity/address of a single vehicle passing by. Therefore, a paradigm shift is advocated from traditional IP-based networking toward the groundbreaking *information-centric networking*. In this article, we scrutinize the applicability of this paradigm in vehicular environments by reviewing its core functionalities and the related work. The analysis shows that, thanks to features like named content retrieval, innate multicast support, and in-network data caching, information-centric networking is positioned to meet the challenging demands of vehicular networks and their evolution. Interoperability with the standard architectures for vehicular applications along with synergies with emerging computing and networking paradigms are debated as future research perspectives.

INTRODUCTION

After years of research and standardization efforts, connected vehicle technologies are almost ready to take off [1]. Primarily conceived to improve driving safety and enable crash prevention through the timely and reliable dissemination of hazard/warning messages among vehicles, connected vehicle technologies are expected to satisfy the ever increasing data appetite of users on wheels entailing vehicle-to-everything (V2X) interactions (Fig. 1). Vehicles exchange data not only with other vehicles (V2V), and the roadside and remote infrastructure (V2R, V2I), but with many other nodes in the vehicle's neighborhood such as the personal communication devices of pedestrians and cyclists, charging stations, and smart grids (e.g., for greener transportation).

Overall, vehicular applications require the distribution of *huge amounts of data among heterogeneous players* under *poor and intermittent connectivity* in high mobility, harsh signal propagation, and sparse roadside infrastructure conditions. The host-centric IP-based protocols of the current Internet, designed with the end-to-end connectivity principle in mind, barely work under such settings. As a matter of fact, stakeholders are open to new networked communication solutions that replace or work alongside IP, for example, by leveraging in-network data caching and localized replication mechanisms to speed up the retrieval of *spatial*- and *time-dependent* contents, to reduce congestion and counteract the intermittent wireless connectivity.

This is where the information-centric networking (ICN) paradigm [2] comes into the picture. ICN originates as a candidate architecture for the future Internet to meet the increasing demand for scalable, reliable, and efficient content distribution. It reverses the traditional IP address-centric philosophy into a content-centric one; this means that a user interested in getting a given content directly uses a "name" to retrieve it without referring to the IP address of the node storing the content.

This innovative approach particularly suits the vehicular ecosystem, which natively privileges the information (e.g., trusted road traffic information relevant to a given incident area) rather than the node identity. Furthermore, with in-network data caching, ICN helps to cope with mobility and sporadic connectivity issues.

As a further strength, unlike the IP solution, which resorted to *patches* for enabling not originally conceived features (e.g., user mobility, security), ICN can be shaped to meet the requirements of future Internet scenarios, as identified by the ICN Research Group of the Internet Research Task Force.[1]

Early works [3, 4], showed the benefits of ICN w.r.t. IP-based solutions in vehicular ad hoc networks (VANETs), and a boom in related proposals has registered in the last couple of years. Such premises motivate this article, the organization of which can be summarized as follows:

- We start with an overview of the ICN paradigm and its potential in vehicular environments.

The authors are with the University Mediterranea of Reggio Calabria.

[1] E. Davies *et al.*, "Request For Comments: 7476. Information Centric Networking: Baseline Scenarios," Internet Research Task Force (IRTF), March 2015.

0163-6804/16/$25.00 © 2016 IEEE

- We scan the recent literature, which promotes ICN and extends its core functionalities to better fit the peculiarities of vehicular communications.
- We identify challenges and debate perspectives for ICN deployment in upcoming VANETs.

Information-Centric Networking

The ICN paradigm was pioneered by the TRIAD project (http://gregorio.stanford.edu/triad/), which defined a *content layer* implementing name-based routing and caching. In more recent years, different ICN architectures have been designed [2], inspired by the aim to better reflect the increasing use of the Internet for information retrieval and dissemination rather than for supporting conversations between pairs of end nodes, like in the original design. Among them, content-centric networking (CCN), originally proposed by Van Jacobson, is under continuous development by research initiatives worldwide, such as the Named Data Networking (NDN) project (http:://named-data.net/).

Within all ICN-based Internet architectures, *information becomes the first-class network citizen*: pieces of information are assigned a *name* and are retrieved without explicitly addressing the hosts/servers that generate or own the information itself. Names uniquely and persistently identify a content (e.g., a movie, a picture, a document, a web page), independent of the location of the producer generating/hosting it, and routers use name-based forwarding rules to retrieve the named content.

This approach has the advantage of securing data instead of the transport channel by embedding authentication and integrity materials in each data packet to make it *self-consistent*. As such, each content packet can be cached by network nodes that make it available for further requests.

The ICN communication model is *connectionless* and *asynchronous* (i.e., consumers and producers exchange data even if not simultaneously connected), and supports *anycast* data retrieval (i.e., the routers forward the request toward any node holding that content). With ICN, *receiver-driven* data exchange is triggered by the consumer sending a request/subscription for a named content; this bans unsolicited data.

Without loss of generality, in the following, we refer to the NDN architecture to describe information-centric content retrieval. NDN is based on two packet types, *Interest* and *Data*, which are used by the forwarding plane of NDN nodes in a two-step process:

- Consumers send out Interest packets specifying the name of the requested content.
- Data packets flow back, carrying the named and secured content units, by following the *traces* left by the forwarded Interests in the nodes.

NDN considers *hierarchical* content names,[2] which appear as user-friendly uniform Resource Identifier (URI)-like identifiers with variable length and variable number of components separated by /. For instance, a picture stored at the UNIRC university server may have the name */unirc/pictures/pictureA.jpg*.

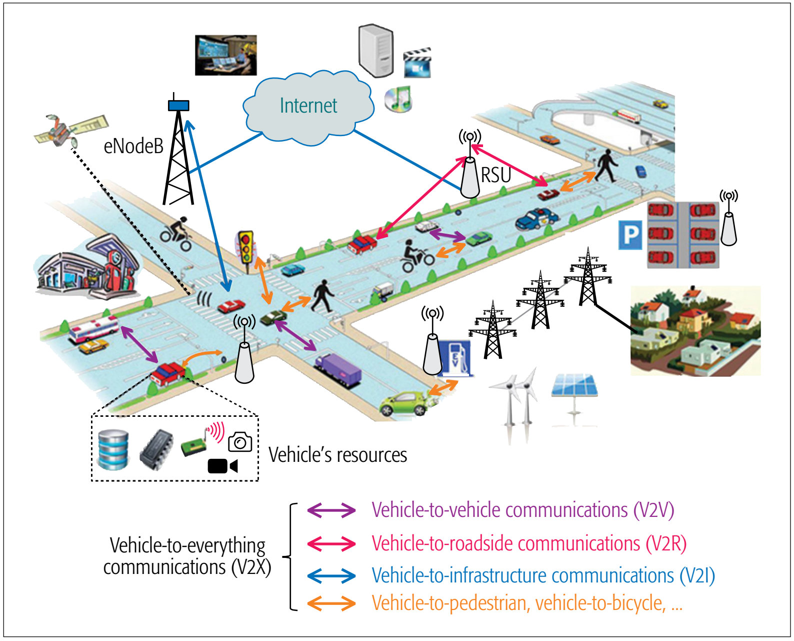

Internet
eNodeB
RSU
P
Vehicle's resources
Vehicle-to-everything communications (V2X)
Vehicle-to-vehicle communications (V2V)
Vehicle-to-roadside communications (V2R)
Vehicle-to-infrastructure communications (V2I)
Vehicle-to-pedestrian, vehicle-to-bicycle, ...

Figure 1. The connected vehicles landscape: from V2V to V2X.

Each NDN node maintains three data structures. A forwarding information base (FIB) routes Interests toward data through name-based lookup. A pending interest table (PIT) keeps state of the forwarded Interest(s) that are not yet satisfied with a returned Data packet. Each crossed node can temporarily cache incoming Data packets in a content store (CS) for faster reply to late requests.

In summary, at Interest reception, a node follows the algorithm in Fig. 2: it first looks in its CS to find a content copy; if a match is not found, it looks in the PIT and, eventually, in the FIB. The example of Interest/Data exchange in Fig. 3 refers to a vehicular environment.

ICN for Connected Vehicles: Motivations

ICN-based VANETs promise enhancements in the areas of *application*, *mobility*, and *security*, as dissected in the following.

Application. Whatever their target, road safety or infotainment, vehicular applications are *information-oriented* in nature: they address a content (e.g., road conditions) and do not care about the producer identity. Generated data are relevant to a given *location* (e.g., points-of-interest notification), and/or to a given *time interval* (e.g., traffic jam warnings expire in a few hours; parking lot availability lasts a few seconds). Finally, generated data are intended for groups of recipients (e.g., ads for parking lots to approaching vehicles).

Through *named data* and routing by name, ICN matches the described vehicular applications' pattern better than the current Internet. Content discovery is simpler because ICN does not need name-to-IP-address resolution, and does not ask for the producer to always be connected.

In addition, ICN simplifies data retrieval from multiple consumers (e.g., map downloading from

[2] In a more general ICN design, *flat* names are also possible, that is, fixed-length identifiers with no semantic structure.

There is a wide consensus on the use of hierarchical naming schemes to effectively match vehicular applications. Hierarchical names are in fact highly expressive and can be easily aggregated under common prefixes to facilitate routing operations and limit the number of FIB entries.

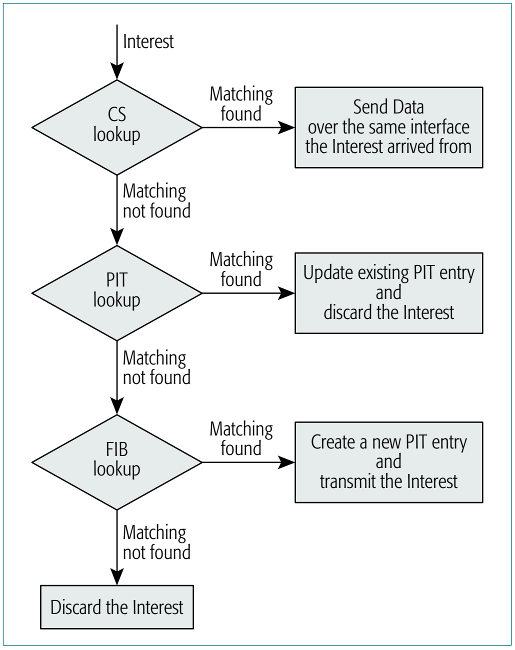

Interest
CS lookup
Matching found
Send Data over the same interface the Interest arrived from
Matching not found
PIT lookup
Matching found
Update existing PIT entry and discard the Interest
Matching not found
FIB lookup
Matching found
Create a new PIT entry and transmit the Interest
Matching not found
Discard the Interest

Figure 2. NDN Interest processing at an intermediate node.

a common roadside unit) by aggregating requests for the same named content in the PIT; it is sufficient to keep track of the Interest incoming interfaces for later Data delivery.

Mobility. IP-based host-centric protocols work awkwardly in mobile environments, and functionality patches, such as Mobile IP, are known to add complexity and perform unsatisfactorily in VANETs [3]. In both the *highway* scenario, where vehicles move at very high speeds, and the *urban* scenario, with signal propagation typically obstructed by buildings, the quality and duration of V2V and V2R links can be adversely affected. In such topologies, classic IP networking operations, like address assignment and path maintenance, become difficult to achieve. Consequently, networking solutions alternative or complementary to IP are also encouraged by standardization bodies in the vehicular application/technology domain [1].

In the Wireless Access in Vehicular Environments (WAVE) stack, the WAVE Short Message Protocol (WSMP), for instance, runs directly over the access layer and supports the single-hop broadcasting of time-sensitive safety data without the need for connection setup operations. In the International Organization for Standardization (ISO)/European Telecommunications Standards Institute (ETSI) architecture for an intelligent transportation system (ITS) station, at the networking and transport layer, besides the IPv6 solution for remote communications, there is room for *geonetworking*, using the geographical position of vehicles for addressing and forwarding purposes, and other (still under definition) networking protocols.

With ICN, the use of named data simplifies mobility support. The *anycasting* and *in-network caching* properties of ICN allow vehicles to retrieve content from the most convenient (typically nearest to the consumer) producer/storage point. This reduces data latency and network traffic. Moreover, a *store-carry-and-forward* mechanism can be supported by ICN, through which a vehicle can serve as a link ("data mule") between disconnected areas and enable communications even under intermittent connectivity.

This is achieved at low cost, thanks to the practically "unlimited" capabilities of vehicles, which do not have energy, processing, or storage constraints.

Security. Due to the *ephemeral* nature of vehicular communications, trustworthiness should be based on data instead of the reputation of providing entities.

ICN natively provides *content-based security*, with protection and trust implemented at the packet level rather than at the communication channel level. Therefore, the setup of a secure connection is no longer required, and the trust in data is decoupled from how/where the data is obtained.

Research Solutions and Opportunities

Although ICN basic mechanisms can be potentially beneficial to address the peculiarities of VANETs identified in the three above-mentioned areas, adequate extensions must be devised to perfectly fit them:

- ICN namespaces matching *applications*' scope, which in turn influence the implementation of in-network *security*
- ICN routing and forwarding strategies, together with in-network caching, effectively managing *mobility* issues

In the following, we scan the representative literature solutions by grouping them according to the main investigated ICN mechanism (i.e., naming and security, routing and forwarding, in-network caching) and identifying the related open issues. Table 1 summarizes the following discussion.

Naming and Security

In the context of VANETs, flexible and expressive naming conventions must be defined to enable applications to retrieve contents that are locally/remotely available or generated on demand. In all cases, content *integrity* and *provenance* must be verified to prevent malicious reporting of fake data, and at the same time, mechanisms are required to protect the user's privacy.

There is a wide consensus on the use of hierarchical naming schemes to effectively match vehicular applications [3, 5–8]. Hierarchical names are in fact highly expressive and can be easily aggregated under common prefixes to facilitate routing operations and limit the number of FIB entries.

In [6] the namespace */traffic/geolocation/timestamp/datatype* is proposed to manage a decentralized floating car data application, where the prefix */traffic* identifies the application, the *geolocation* and *timestamp* components represent the geographical and temporal scope of the content, respectively, and the *datatype* indicates the meaning of the content itself (e.g., vehicle speed). The *geolocation* component is used for *scalable scope-based* content retrieval, as also discussed in [7], where nodes aggregate data at different geolocation granularities (district level, street level, etc.).

The organization of the namespace can be based also on different logical hierarchies. In [5],

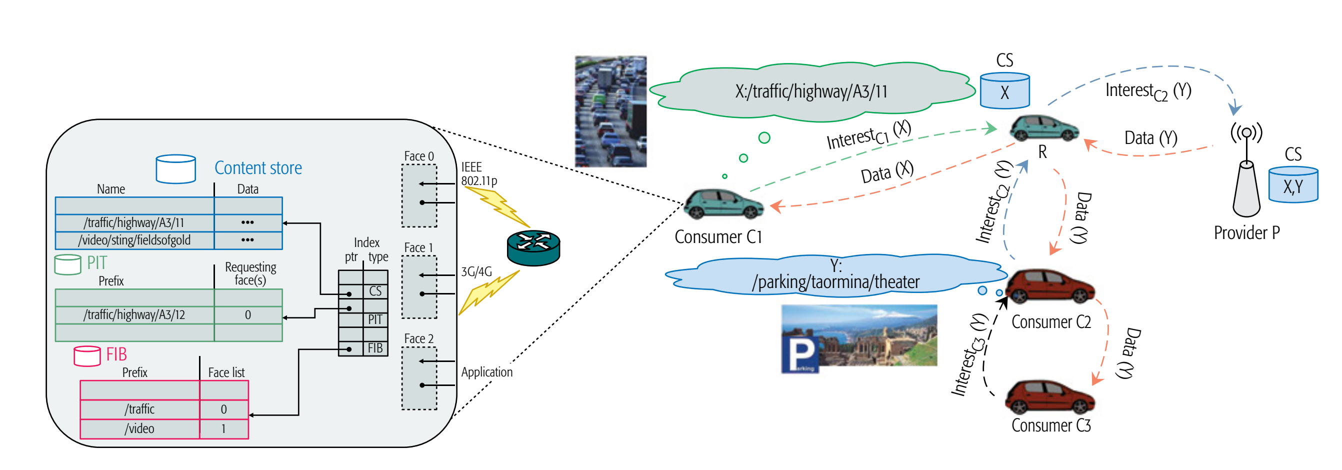


Figure 3. Example of Interest/Data exchange: vehicle R, caching data /traffic/highway/A3/11 (shortened as X), directly replies to consumer C1 without forwarding its Interest to the original provider, P. The Interest for content /parking/taormina/theater (shortened as Y) by C2 is instead forwarded to P. If, in the meanwhile, C2 receives an Interest for content Y from C3; the request is aggregated in the PIT, and the Interest is not forwarded again. When Data Y arrives at R, by following the PIT entries, it is forwarded to C2, which forwards it in its turn to C3.

the namespace is organized as *Category/ServiceName/AdditionalInfo/*, where the main prefix identifies an information category according to content popularity and shareability features. Here, the *category* component (instead of *geolocation*) is used to guide the packets dissemination.

With hierarchical naming, security information (e.g., the publisher signature) is carried in a separate field of the Data packet, thus requiring a public key infrastructure (PKI) for integrity checks. In [3], Data packets collected from vehicles are tagged with their signatures and encrypted using the public key of a reference server. The authors assume that manufacturers record public keys of vehicles and store the public key of the server inside the vehicles before release. A different strategy for an efficient naming and security framework is to capitalize the strengths of hierarchical and flat names, and create hybrid namespaces [9]. Flat names, in fact, enable the use of self-certifying names so that integrity checks run without the need for a PKI, while hierarchical names simplify the prefix-based aggregation.

Although these preliminary research works make the best of ICN names to improve packet delivery (e.g., based on the spatial scope), we are still far from a leading naming solution. If, on one hand, the freedom in the naming design allows researchers to experiment and look for the best performing scheme, on the other hand, the proliferation of different schemes may delay the agreement on common naming conventions and the deployment of large-scale applications. Similarly, some global standardization would be beneficial in the ICN security mechanism, together with an analysis of the possible threats (e.g., distributed denial of service attacks based on Interest flooding), which are still almost uninvestigated in the vehicular environment.

Routing and Forwarding

ICN routing schemes for VANETs can be broadly classified as *proactive* when periodic advertisements from the content providers are needed to keep fresh routing information in the FIB of intermediate nodes, and as *reactive* when the advertisements are not sent in advance and the retrieval is based on Interest flooding.

Flooding-based discovery particularly suits the VANET scenario. It has the advantage of quickly finding the nearest data copy and does not require periodical FIB updates, which can be a heavy (useless) task due to the environment dynamicity and short-lived contents. However, if not properly controlled, flooding may cause network congestion and broadcast storm over the wireless medium. This is why literature solutions improve the forwarding plane with the following strategies:

Collision Avoidance and Packet Suppression. These mechanisms consist of randomizing both Interest and Data sending times, and aborting transmission when detecting that another node has already transmitted the same packet (e.g., in [4, 10–12]).

Selective Flooding. Some works use selective criteria to limit the flooding of the requests due to reactive forwarding. In [5], the route to popular non-shareable/non-cacheable data is proactively stored in the FIBs, while other types of content are searched on demand. In [12], vehicles exchange encounter information; the Interest is only flooded when the producer location is unknown, and only until a relay finds matching location information; when this happens, the Interest is forwarded by geo-routing. A simpler distance-based scheme is deployed in [10], where only the first Interest is flooded to discover the reachable content producer(s), and then subsequent Interests advertise the selected producer identifier and the distance to it so that intermediate nodes forward a request only if they are closer to the provider than the previous sender.

In devising such strategies, the research community mainly benefited from the lessons learned in the past literature for routing in VANETs. But further efforts would be required to account for the unique features of ICN, such as the availability of multiple providers caching the content, and the need to maintain low PIT and FIB sizes and thus limit the lookup delay.

ICN mechanisms	Main proposed solutions	Open challenges
Naming and in-network security	Hierarchical schemes [3, 5–8] Hybrid flat/hierarchical schemes [9] PKI-based integrity checks [3] Self-certifying flat names [9]	–Agreements on common naming and security mechanisms –Analysis of security threats
Routing and forwarding	Collision avoidance and suppression [4, 11] Selective flooding techniques [5, 10, 12]	–Mechanisms to avoid FIB/PIT explosion –Packet prioritization rules –Smart/dynamic outgoing interface selection
In-network caching	Caching of unsolicited contents [6, 8, 9, 11, 12]	–Smart spatial/temporal scope-based caching techniques

Table 1. Summary of the ICN-VANETs relevant literature.

Surveyed solutions are essential in wireless networks with distributed (uncoordinated) access to the medium (e.g., IEEE 802.11p) to reduce the loss rate and congestion. However, today's vehicular networks are moving toward the use of multiple access technologies [1]. By taking advantage of the ICN *layer 2 agnosticism*, a node with multiple radio interfaces can select the most convenient one at a given time, based on measured performances on each of them (e.g., throughput, round-trip time) or collected network information (e.g., density and topology). For instance, in [8], the cellular network is proposed to carry the NDN signaling, while short-range communications are exploited for content distribution. Further enhancements are required to link the design of smart and dynamic selection of the outgoing interface for Interests/Data with the definition of rules for prioritized data transmission (e.g., safety data should have priority over non-safety data).

In-Network Caching

Content caching and replacement policies studied for VANETs, such as pre-fetching and cooperative caching, can be applied or extended to the ICN context. The caching decision may involve contents that vehicles have not requested but that they overhear over the wireless medium. Although vanilla ICN assumes that only solicited data can be cached, it could be useful for a vehicle to store and forward overheard unsolicited contents (e.g., alarms generated by a vehicle in trouble). Many works like [6, 8, 9, 11, 12] extend the ICN data processing in VANETs in this sense.

In addition, there is still room for caching policies that make the best of hierarchical names exposing the *temporal/spatial* scope of the vehicular contents, as preliminarily discussed in [6]. Caching contents out of their spatial scope (e.g., accident warnings beyond the relevance area) as well as caching outdated contents (e.g., traffic jam advertisements from the day before) could be useless. Through ICN naming, vehicles/roadside units (RSUs) can identify the content scope and cache only contents within a specific spatial/temporal range.

So far, however, the benefits of temporal and spatial properties of ICN names in the caching decision have not been clearly supported with quantitative results. We encourage the research community in investigating this promising aspect.

A Look into the Future

We think that, in the path toward a wider acceptance and a large-scale deployment of ICN-based VANETs, further crucial issues need to be addressed, in addition to the enhancements of the ICN mechanisms *per se* undertaken so far in the literature. First, interoperability of ICN with existing and underway connected vehicle standards, technologies, and message sets should be pursued. Then synergies of ICN with the emerging trend of network softwarization should be explored to support the upcoming V2X landscape with the variety of its applications. The latter ones, still under definition, are not necessarily

Challenges	Expected contributions	Potential benefits
Interoperability and coexistence	• Interworking schemes with existing ITS architectures, protocols, and messages • Coexistence solutions with the IP-based core network through proxy functionalities at the edge • Enabling ICN functions (e.g., caching) in nodes of the mobile backhaul and core networks	• Easier and wider ICN penetration • Improved performance for existing ITS standards and technologies
QoS support	• SDN-based centrally controlled name-based forwarding rules • SDN-based wise radio interface selection schemes	• Improved QoS for the end user • Better and flexible resource utilization • Adaptability to dynamic resources and topology changes
Vehicular cloud computing	• Naming schemes addressing cloud resources • Revised forwarding/caching strategies • Simple and effective cloud setup and maintenance operations	• Support for resource-intensive and cooperative apps (e.g., autonomous driving) • Provisioning of value-added services
Big data	• Support for in-network processing operations (e.g., filtering, aggregation) • Novel naming schemes addressing in-network operations	• Reduced network load • Higher scalability
Business models	• Incentive schemes • Assessment of the participation value	• Large-scale participation in content retrieval/cloud services • Novel business opportunities for involved players

Table 2. Future research perspectives for ICN.

related to the ITS arena, for instance, environment monitoring, emergency, and disaster management [13], and go beyond data distribution, hence challenging the capabilities of ICN as originally conceived.

Initial design ideas and hints presented in the following are summarized in Table 2.

Interoperability with reference VANET architectures. We believe that synergies among ICN and ITS standards will provide advantages and accelerate the development in both domains. Although the scientific literature has almost neglected this aspect, it is worth discussing if and how ICN operations can be viewed as components of legacy reference vehicular architectures. The work in [4] was pioneer in proposing to deploy ICN as a replacement (or a complement) of TCP/IP on top of the access layer in the WAVE stack of a vehicular node. Here, we refine our first intuition as illustrated in Fig. 4a, where we propose to locate ICN in the WAVE stack to also encompass some security operations of IEEE 1609.2 at the networking layer.

In the ISO/ETSI stack for an ITS station, we suggest considering ICN to span the networking and transport layer (representing OSI layers 3 and 4), the facilities layer (representing OSI layers 5, 6, and 7; supporting, for instance, message handling and publish-subscribe mechanisms), and the security layer, as illustrated in Fig. 4b. For a full ICN integration, proper interfaces should be clearly defined between the ICN block and existing layers to facilitate interactions without task duplication in different modules.

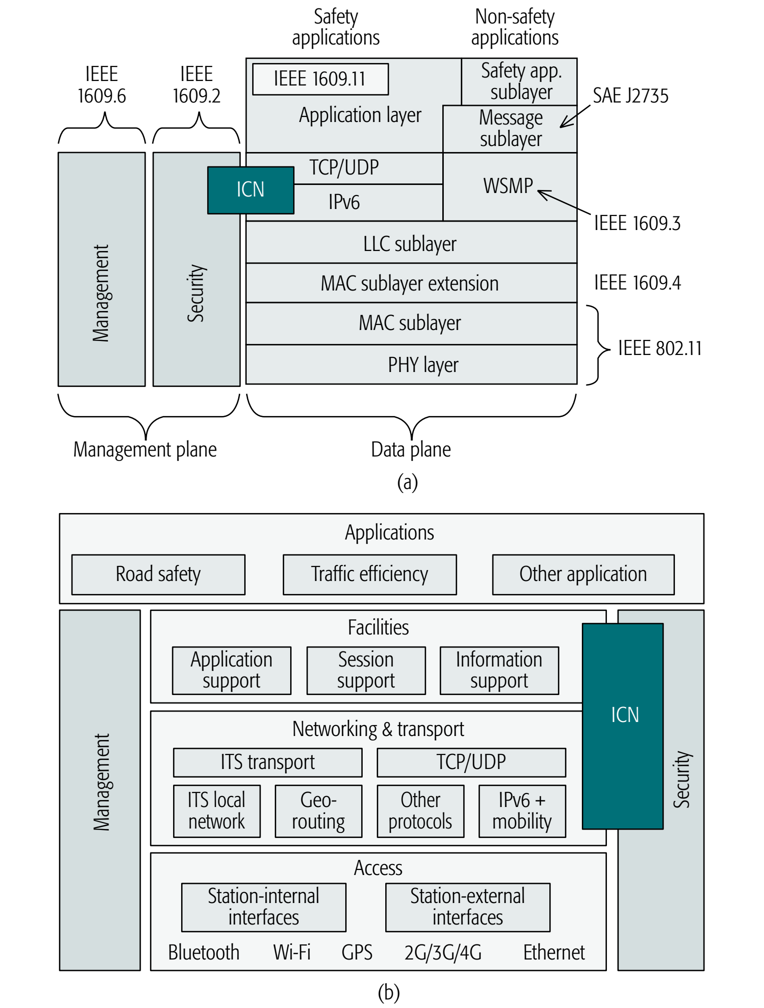


Figure 4. Reference vehicular architectures and ICN: (a) ICN functions in the WAVE architecture (IEEE Std. 1609.0); (b) ICN functions in the ITS-Station architecture (ETSI EN 302 665/ISO 21217).

A means to favor the integration of ICN in the reference architecture could be, in our opinion, to use standard cooperative awareness messages (i.e., CAM in the ETSI architecture, a.k.a. BSM in the WAVE architecture), regularly transmitted by all vehicles, to support ICN routing and forwarding operations, and help to build neighborhood tables without additional overhead.

Coexistence with the IP-based core network. It is quite intuitive that ICN deployment in isolated vehicular network segments could be easier than the replacement of IP in the core network. Indeed, ICN-enabled onboard units could easily be mounted in newly sold cars, and RSUs could be equipped with ICN functions on top of layer 2 (e.g., IEEE 802.11p), hence easily allowing V2V and V2R ICN-based local communications, as preliminarily tested in [14]. This would facilitate wider ICN penetration with incremental graceful upgrades.

Issues may emerge when vehicles generate/request data for/to remote players, reachable via RSUs connected to the Internet or through a cellular interface.

The first case can be addressed from an application-level perspective: we think that similarities between the ICN hierarchical names and the URIs of resources remotely provided/accessed with a RESTful architecture can facilitate such coexistence. For instance, an RSU interfacing ICN vehicular islands and the rest of the Internet could act as a proxy and properly translate URIs into ICN names and vice versa.

The second case of delivery of ICN messages over the cellular interface may raise different concerns. If the advantages in making Internet routers ICN-capable are well known since the dawn of the ICN paradigm, the benefits of letting nodes of the mobile backhaul and core networks understand the semantics of ICN packets and possibly perform ICN operations (e.g., caching) deserve investigation. Findings in such a domain are expected to give insight into the performance of ICN in multi-interface vehicular nodes as well, and to facilitate remote communications.

Quality of service (QoS) support. The delivery requirements of vehicular applications, like short latency and high reliability, have not been considered thus far as an input to ICN operations. ICN mainly works as a *best effort* framework. However, the growing data traffic and strict demands for some future applications (e.g., autonomous driving) pave the way to extend ICN toward QoS support. For instance, software-defined networking (SDN) techniques, recently proposed for VANETs [1], can improve QoS provisioning. With logically centralized network intelligence and state at a controller node, SDN could help to make better decisions based on the combined information from multiple sources, not just individual perception from each node. For instance, SDN-aided ICN forwarding can dynamically decide at which time what type of

> ICN holds much promise for future VANET development, but some challenges still lie ahead before this paradigm can be deployed on a large-scale, co-existing with current and future connected vehicle technologies and standards.

traffic will use which radio interface (e.g., LTE, 802.11p) and configure the forwarding rules (e.g., by injecting the FIBs) for a given type of traffic (e.g., surveillance data in emergency scenarios) according to its name.

Vehicular cloud computing. Vehicles are getting smarter objects, able to share their processing, storage, and sensing resources to support advanced services (e.g., data fusion and processing from different sensors for autonomous driving), by acting as a *local cloud*. In [15], the concept of *vehicular cloud networking* is initially proposed, where such clouds are set up and maintained with the help of ICN. Indeed, through the flexibility of ICN naming, vehicles could request resources by names in a manner that is *agnostic of the physical location* and of the configuration of the system/node that provides them. However, the implications and open questions of ICN-based vehicular clouds are still manifold, since ICN should evolve from a framework delivering contents to a system orchestrating heterogeneous and complex tasks. This means that the semantics of ICN packets, the forwarding, and caching fabric should be re-thought.

Big data. A tremendous amount of data is expected to be generated, delivered, processed, and stored in the upcoming vehicular landscape. ICN already provides some mechanisms to reduce the traffic volume by avoiding request and data packets duplication. Moreover, although not initially conceived to perform in-network data processing operations, ICN could be extended to provide data manipulation *on the fly* (e.g., *filtering* useless data, *aggregating* redundant data) with the advantage of increasing data retrieval scalability and reducing network resources usage. This idea was pioneered by the Named Function Networking (NFN) project (http://www.named-function.net/). NFN allows consumers to also express by name *functions* to be applied over contents; for example, a consumer could request a zipped video file in a specific format (mp4) with the name */name/of/video/codec/mpeg4|/util/compress/zip*. A major challenge of this approach is the design of naming schemes able to identify both functions and data.

Business models. Another crucial aspect to investigate for ICN success in upcoming VANETs will be the definition of incentive mechanisms to motivate car owners to release their on-vehicle resources by caching, processing, and forwarding data in which they may not be interested. A possible approach is to reward vehicles with either monetary incentives or other services in return, like free parking or access to manufacturer/traffic tips. The values of incentives can be determined according to the level of participation (e.g., the volume of forwarded/cached data), the quality of provided resources (e.g., the quality of information may decrease according to the temporal/spatial scope). Overall, agreements and business models are strongly needed between involved stakeholders (e.g., individual drivers, road traffic authorities, content providers, telco operators).

Conclusions

In this article, we have discussed the potential of the ICN paradigm as a networking solution for connected vehicles. The analysis shows that the native design principles of ICN well match the main distinctive features of VANETs and the targeted wide set of vehicular applications. The literature on ICN for VANETs — still at its infancy, due to the age of the ICN topic — designing adaptations and customizations of the baseline ICN architecture to better fit the vehicular environment has been surveyed.

ICN holds much promise for future VANET development, but some challenges still lie ahead before this paradigm can be deployed on a large scale, coexisting with current and future connected vehicle technologies and standards. In the overall balance, the pros outweigh the cons, and encourage the research community to put effort into this timely and increasingly relevant topic.

References

[1] C. Campolo, A. Molinaro, and R. Scopigno, "From Today's VANETs to Tomorrow's Planning and the Bets for the Day After," *Vehic. Commun.*, vol. 2, no. 3, 2015, pp. 158–71.

[2] B. Ahlgren *et al.*, "A Survey of Information-Centric Networking," *IEEE Commun. Mag.*, vol. 50, no. 7, 2012, pp. 26–36.

[3] J. Wang, R. Wakikawa, and L. Zhang, "DMND: Collecting Data from Mobiles Using Named Data," *2010 IEEE Vehic. Net. Conf.*, 2010, pp. 49–56.

[4] M. Amadeo, C. Campolo, and A. Molinaro, "CRoWN: Content-Centric Networking in Vehicular Ad Hoc Networks," *IEEE Commun. Letters*, vol. 16, no. 9, 2012, pp. 1380–83.

[5] Y.-T. Yu *et al.*, "Scalable VANET Content Routing Using Hierarchical Bloom Filters," *IEEE Wireless Commun. and Mobile Comp. Conf.*, 2013, pp. 1629–34.

[6] L. Wang *et al.*, "Data Naming in Vehicle-to-Vehicle Communications," *IEEE INFOCOM Wksps.*, 2012, pp. 328–33.

[7] Z. Yan, S. Zeadally, and Y.-J. Park, "A Novel Vehicular Information Network Architecture based on Named Data Networking (NDN)," *IEEE Internet of Things J.*, vol. 1, no. 6, 2014, pp. 525–32.

[8] A. Bazzi *et al.*, "Cellular Aided Vehicular Named Data Networking," *Int'l. Conf. Connected Vehicles*, 2014.

[9] W. Quan *et al.*, "Social Cooperation for Information-Centric Multimedia Streaming in Highway Vs," *2014 IEEE 15th WoWMoM*, 2014, pp. 1–6.

[10] M. Amadeo, C. Campolo, and A. Molinaro, "Enhancing Content-Centric Networking for Vehicular Environments," *Computer Networks*, vol. 57, no. 16, 2013, pp. 3222–34.

[11] L. Wang *et al.*, "Rapid Traffic Information Dissemination Using Named Data," *Proc. 1st ACM Wksp. Emerging Name-Oriented Mobile Networking Design-Architecture, Algorithms, and Applications*, 2012, pp. 7–12.

[12] Y.-T. Yu *et al.*, "Scalable Opportunistic VANET Content Routing with Encounter Information," *IEEE Int'l. Conf. Network Protocols*, 2013, pp. 1–6.

[13] O. Altintas *et al.*, "Making Cars a Main ICT Resource in Smart Cities," IEEE INFOCOM 2015, *Int'l. Wksp. Smart Cities and Urban Informatics*.

[14] G. Grassi *et al.*, "VANET via Named Data Networking," *IEEE INFOCOM Wksps.*, 2014, pp. 410–15.

[15] E. Lee *et al.*, "Vehicular Cloud Networking: Architecture and Design Principles," *IEEE Commun. Mag.*, vol. 52, no. 2, 2014, pp. 148–55.

Biographies

Marica Amadeo (marica.amadeo@unirc.it) is a postdoctoral researcher at University Mediterranea of Reggio Calabria, Italy. She received a B.S. degree (2005) and an M.S. degree (2008) in telecommunications engineering from the University Mediterranea of Reggio Calabria, and a Ph.D. degree in 2013 from the same university. Her major research interests are in the field of ICN and wireless ad hoc networks.

Claudia Campolo (claudia.campolo@unirc.it) is an assistant professor of telecommunications at University Mediterranea of Reggio Calabria. She received an M.S. degree in telecommunications engineering (2007) and a Ph.D. degree (2011) at the same university. She was a visiting Ph.D. student at Politecnico di Torino (2008) and a DAAD fellow at the University of Paderborn, Germany (2015). Her main research interests are in the field of vehicular networking and future Internet architectures.

Antonella Molinaro (antonella.molinaro@unirc.it) has been an associate professor of telecommunications with the University Mediterranea of Reggio Calabria since 2005. Before, she was an assistant professor with the University of Messina (1998–2001), with the University of Calabria (2001-2004), and a research fellow at the Polytechnic of Milan (1997-1998). She was with Telesoft, Rome (1992–1993), and with Siemens, Munich (1994-1995) as a CEC Fellow in the RACE-II program. Her current research focuses on vehicular networking, information-centric networking, and the future Internet.

Call for Papers
IEEE Communications Magazine
Enabling Mobile and Wireless Technologies for Smart Cities

Background

Due to advancements in communication and computing technologies, smart cities have become a main innovation agenda of research organizations, technology vendors, and governments. To make a city smart, a strong communications infrastructure is required for connecting smart objects, people, and sensors together. Smart city communication involves multiple aggregation and access networks that can be either public or private. The rapid progress in smart cities research is posing enormous challenges in terms of significance, scope, and problem domain. Smart cities rely on wireless and mobile technologies for providing services such as healthcare assistance, security and safety, real-time traffic monitoring, and managing the environment, to name a few. Such applications have been a main driving force in the development of smart cities. These mobile and wireless technologies enable several new services that result in better decision making and actions made by enterprises and governments. Without the appropriate communication networks, it is really difficult for a city to facilitate its citizens in sustainable, efficient, and safer manner/environment. Considering the significance of mobile and wireless technologies in realizing the vision of smart cities, there is a need to conduct research to further investigate the standardization efforts and explore different issues/challenges in wireless technologies, mobile computing, and smart environments.

This Feature Topic focuses on the crossroads between scientists, industry practitioners, and researchers from different domains in wireless technologies, mobile computing, and smart environments. We envision providing a platform for researchers to further explore the domain and explore the challenges. In this Feature Topic, we invite researchers from academia, industry, and government to discuss challenging ideas, novel research contributions, demonstration results, and standardization efforts on enabling mobile and wireless technologies for smart cities. In this Feature Topic we would like to try to answer some (or all) of the following questions: How can mobile and wireless technologies improve the performance and services provided by smart cities? How can one evaluate the impact of mobile and wireless technologies on smart cities services? What are the key mobile and wireless technological challenges that hinder the development of smart cities? How can we standardize the wireless interfaces of devices for communication in smart cities?

Topics of interest include, but are not limited to:
- Resource and network management in smart cities
- Quality of service mechanisms for wireless networks in smart cities
- Integration and coexistence of technologies and networks for smart cities
- Interoperability between heterogeneous networks of smart cities
- Topology and mobility management in smart cities
- Energy-aware wireless protocols and algorithms for smart cities
- Sensing technologies and applications for smart cities
- Wireless networks for smart city surveillance and management
- Experimental network measurements and characterization of smart cities data traffic
- Security and privacy concerns in smart cities

Submissions

Articles should be tutorial in nature, with the intended audience being all members of the global communications technology community. They should be written in a style comprehensible to readers outside the specialty of the article. Mathematical equations should not be used (in justified cases up to three simple equations may be allowed). In general, however, mathematics should be avoided; instead, references to papers containing the relevant mathematics should be provided. Articles should not exceed 4500 words (from introduction through conclusions, excluding figures, tables, and captions). Figures and tables should be limited to a combined total of six. The number of references is recommended not to exceed 15. Complete guidelines for preparation of the manuscripts are posted at http://www.comsoc.org/commag/paper-submission-guidelines. Please send a PDF (preferred) or MSWORD formatted paper via Manuscript Central (http://mc.manuscriptcentral.com/commag-ieee). Register or log in, and go to Author Center. Follow the instructions there. Select "December 2016/Enabling Mobile and Wireless Technologies for Smart Cities" as the Feature Topic category for your submission.

Important Dates

- Submission Deadline: February 29, 2016
- Notification Due Date: June 30, 2016
- Final Version Due Date: September 15, 2016
- Feature Topic Publication Date: December 2016

Guest Editors

Ejaz Ahmed
University of Malaya, Malaysia
imejaz@gmail.com

Ammar Rayes
Cisco Systems, USA
rayes@cisco.com

Wael Guibene
Intel Labs, Ireland
wael.guibene@intel.com

Muhammad Imran
King Saud University, Saudi Arabia
dr.m.imran@ieee.org

Jaime Lloret
Univ. Politecnica de Valencia, Spain
jlloret@dcom.upv.es

Mohsen Guizani
Qatar University, Qatar
mguizani@ieee.org

Guangjie Han
Hohai University, China
hanguangjie@ieee.org

Network Engineering for Real-Time Networks: Comparison of Automotive and Aeronautic Industries Approaches

Fabien Geyer and Georg Carle

With the advent of electronic x-by-wire applications with strict reliability and safety requirements, formal verification of safety-critical networks has become an important step of the design process in the automotive and aeronautic industries.

ABSTRACT

With the advent of electronic x-by-wire applications with strict reliability and safety requirements, formal verification of safety-critical networks has become an important step of the design process in the automotive and aeronautic industries. We first review in this article the different network technologies and architectures used by both industries. We then present and compare the two prevailing mathematical frameworks used by each industry for validating the correct behavior of a network: schedulability analysis and network calculus. Via an empirical evaluation of both methods in two different use cases, we show the strengths and weaknesses of both methods, and derive a simple guideline on which method to use depending on the type of network used.

INTRODUCTION

In the last two decades, distributed embedded electronic applications have become the norm in a large part of the automotive industry. Those applications cover a large set of functionalities with different requirements, ranging from power train and chassis control with hard real-time constraints (like engine control, steering, braking), to passenger entertainment with less stringent constraints. Due to those requirements and the distributed nature of the various electronic control units (ECUs) implementing those functions, the validation of end-to-end timing constraints on networks has become an important part of the design process of a car. In order to guarantee the dependability of those real-time applications, formal methods are generally used to verify end-to-end timing constraints.

In this article we provide an overview of the various formal methods that have been put forward and used by the automotive industry. We cover two important elements of the formal verification: the specification of timing properties of information exchange on the network (i.e., packets and flow descriptions) and the underlying mathematical models used for computing end-to-end performance.

As the challenges described earlier are also present in the aeronautic industry, we compare established methods in the car industry to methods used for the design of hard real-time safety-critical distributed avionic systems. This comparison between the two industries highlights fundamental differences in terms of technology and choice of formal methods used. This comparison is especially relevant now that the automotive industry is extending its use of Ethernet, from initially only infotainment purposes to now hard real-time functionalities, a technology that has been used and put to the test by the aeronautics industry for more than a decade.

We propose to evaluate the two dominant formal methods used by each industry, that is, *schedulability analysis* and *network calculus*. Both methods are compared empirically in two use cases: a small reference automotive bus architecture and randomly generated larger Ethernet networks with multiple hops. We show that while schedulability analysis provides tighter results than network calculus for the analysis of a single bus, it is the opposite in the case of larger and more complex networks where statistical multiplexing plays an important role.

This article is organized as follows. The following section presents predominant network technologies used by the automotive and aeronautic industries and introduces their respective views on formal modeling of timing properties in networks. The section after that delves into the details of the various mathematical models used by both industries, and compares the two main approaches quantitatively. The article is then concluded.

PREVAILING INDUSTRIAL NETWORKING SOLUTIONS

We present, in this section and in Table 1, the predominant network technologies for real-time networks and applications used by the automotive and aeronautics industries. We focus here on the most commonly used technologies and refer the reader to [1] for a more extensive review.

AUTOMOTIVE INDUSTRY

Networks used in today's cars are usually based on the four following technologies, listed here by increasing order of bandwidth: local interconnect network (LIN), controller area network (CAN), FlexRay, and media oriented systems

Fabien Geyer is with Airbus Group Innovations; Georg Carle is with Technische Universität München.

0163-6804/16/$25.00 © 2016 IEEE

transport (MOST). All these network technologies use the concept of a non-preemptive shared bus, meaning that only one node at a time can transmit messages, and all the other nodes have to wait for the completion of an ongoing transmission before transmitting a new message. Due to the relatively low bandwidths provided by these technologies and the ever growing number of applications in today's cars, multiples buses are generally used with gateways interconnecting them. This principle is illustrated in Fig. 1a. Other technologies such as Ethernet have been used and proposed, but usually not in the context of real-time networks.

These technologies are either based on an event-triggered paradigm (LIN and CAN), where messages are sent sporadically according to external triggers; on a time-triggered paradigm, where messages are sent according to a precise time schedule (MOST); or a hybrid solution mixing both paradigms (FlexRay). In order to formalize the exchanges between the different actors in the automotive industry, the Automotive Open System Architecture (AUTOSAR) set of standards [3] defines methods for the formal description of software and network architectures used in cars. In order to describe the timing behavior of messages and frames on networks, task models have been formalized in the AUTOSAR Timing Extension [3].

The following timing constraints models are defined:

- *Periodic events*, with a single event, a predefined repeating pattern of events, or a burst of events
- *Sporadic events*, corresponding to a periodic event that is not guaranteed to occur
- *Arbitrary events*, used for abstracting events that do not fall in the previous categories, such as captures made on real networks

All these models generally define the maximum message size, the minimum time between two events, and the periodicity and possible jitter of the events. Associated with these timing constraints, the AUTOSAR Timing Extension also defines models to describe event chains in order to model the interaction between different ECUs and the chain of messages they will exchange for a specific application.

Aeronautic Industry

While networks used in today's aircraft use a wide range of network technologies, the most prominent ones are ARINC 429, MIL-STD-1553 ARINC 629, and ARINC 664. Initiated in the late 1970's, the ARINC 429 standard defines a bus with a single emitter node and multiple receivers. The MIL-STD-1553 standard is comparable from a functional point of view to LIN as a master node regulates all communications on a 1Mb/s bus. Due to the poor scalability in terms of bandwidth, cost and cabling effort of ARINC 429, the ARINC 629 and 664 standards have been developed in order to multiplex communications, either by using a shared bus (ARINC 629) or a meshed network (ARINC 664).

ARINC 664 — also referred to as AFDX (Avionics Full-Duplex Switched Ethernet) — was developed in the mid 2000's in order to provide a deterministic Ethernet network. It has now become widely adopted for safety critical applications such as *fly-by-wire*. It is based on modern Ethernet concepts, that is, packet switching (as opposed to shared buses) on partially meshed networks, as illustrated in Fig. 1b. A unique task model was formalized in AFDX under the name virtual link (VL). It defines the minimum time between two messages, referred to as the bandwidth allocation gap (BAG), the maximum message size, a unique sender, and the static paths to its receivers. The deterministic property of the network is ensured by encapsulating all communications in VLs, and switches drop messages that are not in compliance with the VLs' configuration.

Contrary to the technologies previously cited, AFDX does not use a concept of cooperation for transmitting messages (i.e., network-wide schedule or master node) as it works on the principle of statistical multiplexing as standard Ethernet. Messages are transmitted as they arrive or queued if multiple messages arrive at the same time.

Technology	Bandwidth	Paradigm	Topology
LIN	20 kb/s	Event-triggered: master/slave	Shared bus
CAN	1 Mb/s	Event-triggered: priority arbitration	Shared bus
FlexRay	20 Mb/s	Hybrid event- and time-triggered	Shared bus
MOST	150 Mb/s	Time-triggered	Shared bus
ARINC 429	100 kb/s	Event-triggered: single sender	Dedicated bus
MIL-STD-1553	1 Mb/s	Event-triggered: master/slave	Shared bus
ARINC 629	2 Mb/s	Time-triggered or event-triggered with arbitration	Shared bus
ARINC 664 (AFDX)	100 Mb/s	Event-triggered: statistical multiplexing	Meshed network

Table 1. Predominant automotive and aeronautics network technologies currently deployed for real-time applications.

Formal Verification of Networks

We noted in the introduction that hard real-time guarantees for communications in both cars and aircraft are needed in order to ensure the dependability of some critical applications. An example of applications where such guarantees are needed is the so-called *drive-by-wire* or *fly-by-wire*, where the commands of the driver or pilot are not translated to the wheels or flight control surfaces via a direct physical connection, but are performed via electronic sensors and actuators. In order to provide safe operation of the car or aircraft, deterministic end-to-end guarantees on the delay between the action of the driver or pilot and the moment the actuator actually performs the according action are needed.

We described in the previous section two different approaches to the network technology and multiplexing concept, non-preemptive shared buses for the automotive industry, and meshed networks with statistical multiplexing for the aeronautics industry. In this section we describe and compare the main methods and models that have been developed to formally verify that a

The methods developed and often used for computing end-to-end performances by the automotive industry for networks generally focus on how to share a bus between different messages from the ECUs. This problem can be generalized to the challenge of sharing a common resource between *N* different tasks.

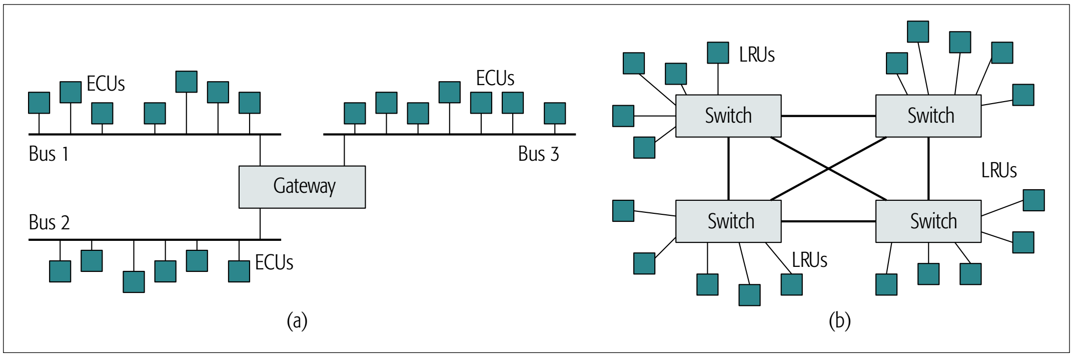


Figure 1. Illustration of typical network topologies used in the automotive and aeronautics industries: a) automotive application where ECUs are connected via buses; b) aeronautics application where line-replaceable units (LRUs) are connected via a meshed network.

network satisfies timing requirements. In this section we focus on hard real-time applications with deterministic guarantees.

Figure 2 presents the basic notions regarding end-to-end delay analysis, where we describe the following three performance indicators. First, the *maximal observed delay* corresponds to the maximal delay measured on a real network during its normal operation. This delay can generally be approximated using simulations. Second, the *exact worst case* corresponds to the theoretical worst case delay that can actually occur in the case when the elements of the network behave within their limits, but in a very specific pattern lead to this worst case. Finally, the *upper bound* corresponds to the bound calculated by an analytical model, which is generally larger than the actual worst case due to approximations, simplifications, or shortcomings of the formal method.

In the case of hard real-time requirements, we are interested in the exact worst case and upper bound metrics as they formally guarantee that delay requirements are met.

We use the notions of *tasks* and *flows* interchangeably in the rest of this article.

Scheduling Analysis of Buses

The methods developed and often used for computing end-to-end performance by the automotive industry for networks generally focus on how to share a bus between different messages from the ECUs. This problem can be generalized to the challenge of sharing a common resource between *N* different tasks, a problem studied since the early days of manufacturing, transportation, and computing.

One early model uses a periodic task model, with T_i being the minimum time between successive messages of task i, and C_i being the maximum time needed to process the task. The response time of task i, noted here as RT_i, is the sum of its processing time and the waiting delay due to the other tasks. Formally, RT_i can be computed using [4]:

$$RT_i = C_i + P_i + \sum_{j \in HP(i)} \left\lceil \frac{RT_i}{T_j} \right\rceil \cdot C_j \qquad (1)$$

with $HP(i)$ denoting the tasks interfering with i, representing tasks of higher priority than task i. If the resource sharing scheme is not preemptable (as, e.g., in CAN), P_i denotes the longest time that any lower-priority message can block the resource. Otherwise, $P_i = 0$. The method described by Eq. 1 often is called *fixed priority schedule analysis* or *deadline monotonic analysis*. With *N* tasks, Eq. 1 leads to a set of *N* equations, and solving them has been shown in [5] to be NP-hard. An illustration of Eq. 1 is presented in Fig. 3.

A common method used for finding a numerical solution to Eq. 1 is to use a fixed point iteration. This leads to a complexity in the order of $\mathcal{O}(k \cdot N^2)$, with *N* being the number of tasks and *k* being the number of iterations needed to reach the fixed point. Practical applications of this method show that *k* grows with the utilization of the network (see [6] for example). The model described in Eq. 1 applies to both event- and time-triggered networks.

While Eq. 1 is used for computing the worst-case response time of a task, the priority ordering of tasks (represented by $HP(i)$ in Eq. 1) cannot directly be derived from Eq. 1. Various methods have been proposed, such as for instance defining the priority ordering as function of the deadline or period ordering, or characterizing the priority assignment task as a mixed integer linear program. We refer to [7] for a broader survey on schedulability analysis and methods to assign priorities.

For concrete application of this method, we refer to [8] for an application on CAN. For FlexRay, a bus mixing time-triggered and event-triggered messages by allocating time-slots for each types of messages in the so-called *static* and *dynamic segments*, the method described earlier in Eq. 1 still can be used via some modifications, such as presented for instance in [9] in the general case. We refer to [10] for an application of this principle to FlexRay.

While the model described in Eq. 1 is not of common practice for the study of avionic networks, it recently has been used in [6] for AFDX. In case of networks where tasks are distributed among multiple buses interconnected with gateways as presented in Fig. 1a, one frequently used technique is to treat each bus separately as presented in [6].

Network Calculus and Bounds

While the previous section focused on methods on how to share a common resource between multiple agents, the methods used for avionic networks and AFDX are different due to the use of meshed topologies and statistical multiplexing. In order to formally verify that end-to-end delays are below their prescribed deadlines in such a network, the *network calculus* framework is currently being used by the aeronautics industry. We refer to [11] for the mathematical theory behind this framework. The task model used in network calculus is based on the concept of an arrival curve, which specifies the worst case amount of data that can arrive in a given time interval $[s, t]$. Formally, this is described as

$$R(t) - R(s) \leq \alpha\,(t - s),\ \forall 0 \leq s \leq t \qquad (2)$$

with $R(t)$ describing the amount of data arriving at time T, and α a function describing the constraint. Similarly, processing nodes in the network (i.e., queues or buses) are modeled using the concept of a service curve, which specifies the guaranteed amount of data that is processed during a given time interval $[s, t]$. Formally, this is described as

$$R^*(t) - R(s) \geq \beta\,(t - s),\ \forall 0 \leq s \leq t \qquad (3)$$

with $R(s)$ describing the amount of data arriving at time s, $R^*(t)$ the amount of processed data at time t, and β a function describing the constraint.

Classically, α and β are defined as affine functions, characterizing the maximum bandwidth that can be generated by a flow and the bandwidth offered by a link. In this case, the arrival curve is usually called a *token bucket*, and the service curve is called *rate latency*. Both types of curves are illustrated in Fig. 4.

In addition to the curve definitions in Eqs. 2 and 3, mathematical operations referred to as *convolution* and *deconvolution* are defined and formalized under the so-called (min,+) calculus. Those operations can be used to aggregate flows, compute the departure curve of a flow traversing a server, or simplify a network of servers to a smaller one. Using Eqs. 2 and 3 and this algebra, two end-to-end performance bounds can be derived. The latency bound is defined as the maximal horizontal deviation between a service and an arrival curve. The buffer bound is defined as the maximal vertical deviation between the two curves. Those two bounds are illustrated in Fig. 4.

In the case of networks with more than one hop and where multiplexing plays an important role (e.g., Ethernet and AFDX), the notion of *grouping* has been developed to tighten the bounds in [12]. It defines the property that when flows share the same path, they become multiplexed after their first common hop, and their arrival curve can be redefined.

The various operations described in the (min,+) algebra can be implemented efficiently, as presented in [13]. By restricting arrival and service curves to affine functions, those operations can be implemented efficiently with a complexity of $\mathcal{O}(1)$. While finding the optimal method of applying network calculus to a network is still an open research question, its computational complexity is on the order of $\mathcal{O}(N_{flows} \cdot N_{switches})$.

As noted earlier, network calculus has been used successfully by the aeronautics industry for AFDX, as detailed in [12]. More recently, it has been used to study CAN in [14].

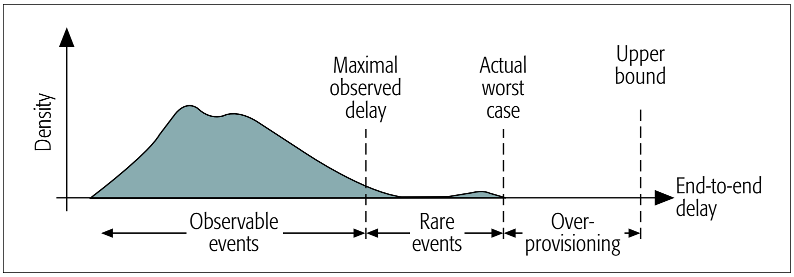


Figure 2. Basic notions regarding end-to-end delay analysis.

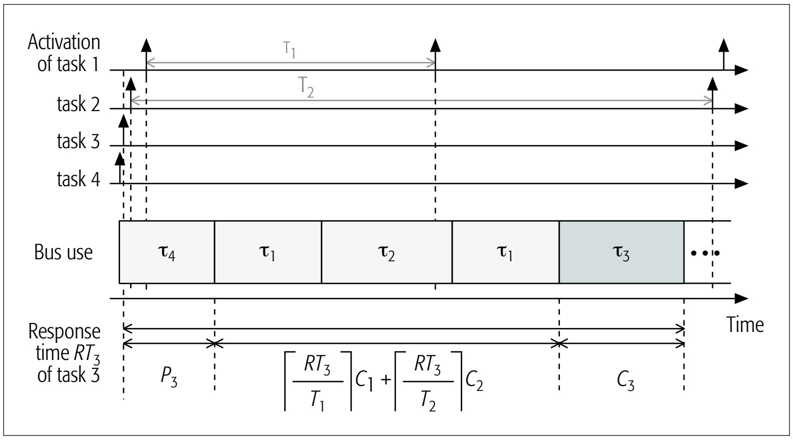

Activation of task 1
T_1
T_2
task 2
task 3
task 4
Bus use
τ_4
τ_1
τ_2
τ_1
τ_3
•••
Time
Response time RT_3 of task 3
P_3
$\left\lceil \frac{RT_3}{T_1} \right\rceil C_1 + \left\lceil \frac{RT_3}{T_2} \right\rceil C_2$
C_3

Figure 3. Illustration of Eq. 1 with four tasks.

Other Methods

The challenge of formally verifying timing constraints on networks has attracted a large body of work, and many methods have been developed apart from the ones presented earlier.

Two notable formal methods should be mentioned, *real-time calculus* [15] and the *optimized trajectory approach* [16]. The goal of both approaches is to gather the strengths of both schedulability analysis and network calculus in order to have a common approach giving tight bounds on both use cases presented in this article.

Model-checking and timed-automata were applied in order to compute the exact worst case behavior of flows, and hence get exact worst case delays as pictured in Fig. 2. While such methods avoid overprovisioning networks, they are usually computationally expensive due to the state-space explosion. Evaluations on avionic networks showed that they are limited to small networks with around 50 flows, as mentioned in [16].

Stochastic methods were also developed in order to take into account the stochastic behavior of flows and hence model more precisely what happens in a real network. These methods are generally able to produce tighter bounds than purely deterministic methods, but they are often more complex to apply due to more sophisticated mathematic concepts. They are also more adapted to so-called *firm real-time* applications (e.g., audio or video streaming) where infrequent deadline misses or packet loss can be tolerated.

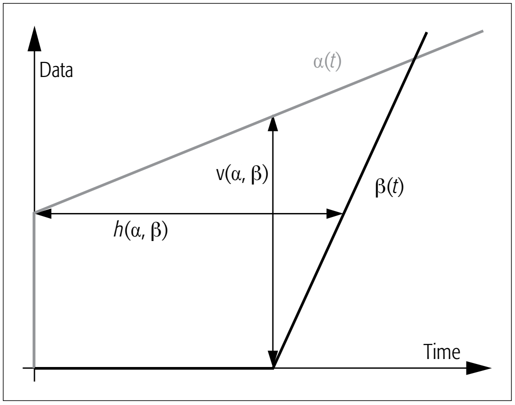


Figure 4. Latency ($h(\alpha, \beta)$) and buffer ($v(\alpha, \beta)$) bounds in network calculus.

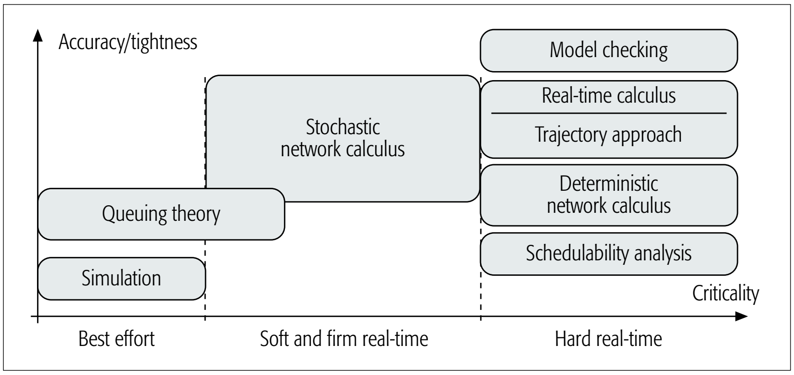


Figure 5. Overview of formal methods for performance evaluation of networks.

An overview of these different methods and their tightness is presented in Fig. 5.

Evaluation and Comparison

We compare both methods presented in this section by evaluating them in two scenarios: first a reference automotive example with a CAN bus, and second randomly generated Ethernet networks where flows traverse multiple hops representing more avionic use cases.

Evaluation of an Automotive CAN Bus: We first focus on the CAN bus scenario described in [8, Table 2]. We chose this example because it should represent a realistic CAN bus, as it is based on a benchmark proposed by the Society of Automotive Engineers (SAE) in the early 1990s to evaluate different multiplexing communications technologies. Also, [8] contains a detailed description of the use case, an explanation of how to apply it to a CAN bus, as well as results of a schedulability analysis as described earlier in Eq. 1.

We refer to [8, 17] for details and numerical results of the schedulability analysis. We follow a modeling approach similar to the one presented in [14] for the network calculus analysis. We use affine functions for the arrival and service curves, as mentioned earlier.

We use the concept of a leftover service curve defined in network calculus in order to use the concept of prioritization of CAN with network calculus. It describes how much bandwidth is left over and the impact in terms of latency flows of high priority have on lower ones.

To get an overview of the latencies we might get in a real network, we also performed simulations of the network using OMNeT++, where we assume no synchronization between the flows. Packets are generated periodically according to the C_i and T_i parameters. We follow a Monte Carlo approach for the simulations, with 36 runs of 5000 s.

The results of the formal verifications from [8], the network calculus analysis, and the simulations are presented in Fig. 6. Error bars for the results of the simulations correspond to a 95 percent confidence interval. We notice that the bounds given by the network calculus analysis are less tight than those of the schedulability analysis, as the priority of the task increases. This comes from the fact that the network calculus analysis made here makes use of a basic fluid model which does not take into account the effect of packetization, as opposed to the schedulability analysis.

Evaluation of Multihop Ethernet Networks: In contrast to the previous evaluation and in order to explore more complex use cases, we now focus on Ethernet topologies where flows traverse multiple hops, as frequently used in the aeronautic industry and illustrated in Fig. 1b. In order to cover a wide variety of use cases, we propose to generate these topologies randomly. The generated network topologies correspond to trees, where leaves are devices (ECUs or LRUs), and internal vertices are Ethernet switches. The traffic is composed of unidirectional and unicast flows, where sources and destinations are taken randomly among the leaves of the tree. Flows follow the periodic task model previously described. The maximum message size of each flow is taken randomly from a uniform distribution between 100 and 1400 bytes. The values used for the period of the messages is taken randomly from the allowed values given by AFDX, that is, 2^k ms with $k \in \{1 \ldots 7\}$. In order to avoid generating topologies where the number of flows and the utilization of the network are linearly correlated, we took different subsets of k such that we generated topologies with few flows and high utilization as well as ones with many flows and low utilization.

Regarding the methodology for computing the end-to-end latency bounds, we used the DiscoDNC tool [18] for the network calculus analysis, and we followed an approach similar to the one presented in [6] for the schedulability analysis. In total, we generated 500 random topologies, with the number of devices varying from 5 to 470, and the number of flows varying from 30 to 4200 and traversing between 1 and 6 hops. The maximum utilized links of each topology varied from 0.09 to 99.9 percent. As opposed to the previous evaluation, there is no priority ordering between flows.

For quantitatively evaluating the gap between the two methods, we use the relative difference between the end-to-end bounds of a flow or, in other words, the difference between the two methods, normalized by the value of the reference method, here network calculus. We noted earlier that one of the strengths of network calculus compared to schedulability analysis is the grouping property, which takes into account the multiplexing feature of Ethernet. Hence, in order to evaluate and quantify the impact of this property, we introduce a metric called the *flow grouping factor* for each physical interface. It characterizes the degree of flow multiplexing and varies between 0 (no multiplexing as in the CAN bus previously studied) and 1 (all flows can be grouped together). This metric is defined as the number of flows that can be grouped, normalized by the number of groups and total number of flows.

The result of the comparison between the two formal methods is presented in Fig. 7. Each point on the figure corresponds to a topology. We first notice that there are no negative values, meaning that for topologies where more than one bus needs to be studied, network calculus will produce on average tighter bounds than the schedulability analysis. This is in contrast to the results presented in Fig. 6, and it can be explained by the fact that there is no priority ordering between flows, and we have at least one hop.

Looking at the overall trend in Fig. 7, there is a linear relationship between the mean flow grouping factor and the mean relative difference between the two methods, as illustrated by the linear regression. We grouped the topology by average number of traversed hops per flow in Fig. 7 and represented the centroid of each group. As the number of hops increases, the difference between the two methods also grows. This is particularly noticeable in areas where groups overlap.

Further comparisons did not lead to a similar tight relationship between a metric of the topology and the relative difference between the two formal methods. For instance, the utilization of the network did not turn out to be a good indicator, as networks with low utilization (i.e., from 5 to 20 percent) and ones with high utilization (i.e., from 80 to 99 percent) showed similar relative differences. This is also shown in our definition of our metric, as it is independent of the bandwidth usage of the flows.

Conclusion of the Evaluations: From the two previous numerical evaluations, we can derive the strengths and weaknesses of both methods. In the case of single bus systems with priorities, bounds derived using schedulability analysis are shown to be tighter than those from network calculus, as presented in Fig. 6. This trend is reversed in the case of networks in which multiple hops are traversed, as shown in Fig. 7, and comes from the fact that the schedulability analysis does not take into account multiplexing as opposed to network calculus. The guideline which can be derived from this is that network calculus should be preferred for networks with multiple hops where multiplexing plays an important role (as illustrated in Fig. 1b), while schedulability analysis should be preferred for single bus networks (as shown in Fig. 1a). This conclusion also helps to understand why the automotive and aeronautics industries each chose different approaches with respect to the formal verification of networks.

Conclusion

In this article we have presented the prevailing network technologies and topology types used by the automotive and aeronautics industries for x-by-wire distributed applications. Due to the hard real-time requirements needed for these safety-critical applications, we have proposed comparing the two dominant approaches used by each industry for formally verifying end-to-end latency requirements: schedulability analysis and network calculus.

Via an empirical evaluation on two use cases, we have deduced that schedulability analysis should be preferred on the single-bus systems often used in the automotive industry, while network calculus should be favored on the larger networks, where multiplexing plays an important role, often seen in the aeronautics industry. As the automotive industry heads toward Ethernet-based networks for real-time applications thanks to recent standardization efforts by the IEEE, this article helps to gain insight from the aeronautics industry, its usage of Ethernet for hard real-time applications in the last decade, and the methods used for formally verifying end-to-end latency requirements.

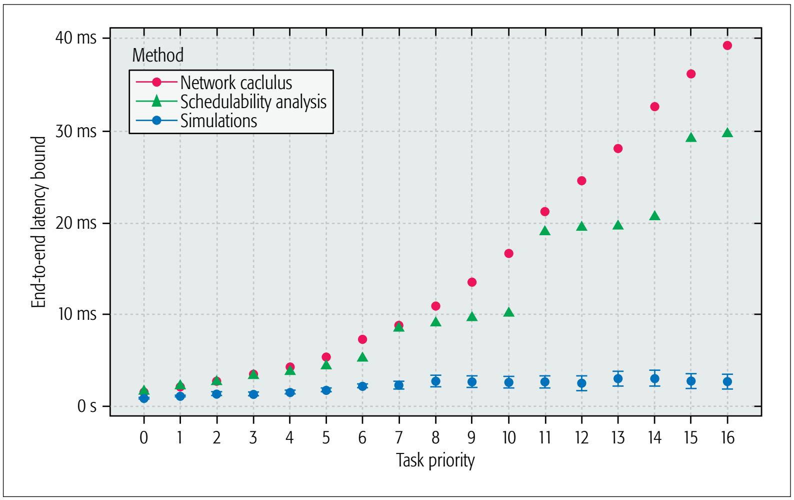


Figure 6. CAN bus evaluation results.

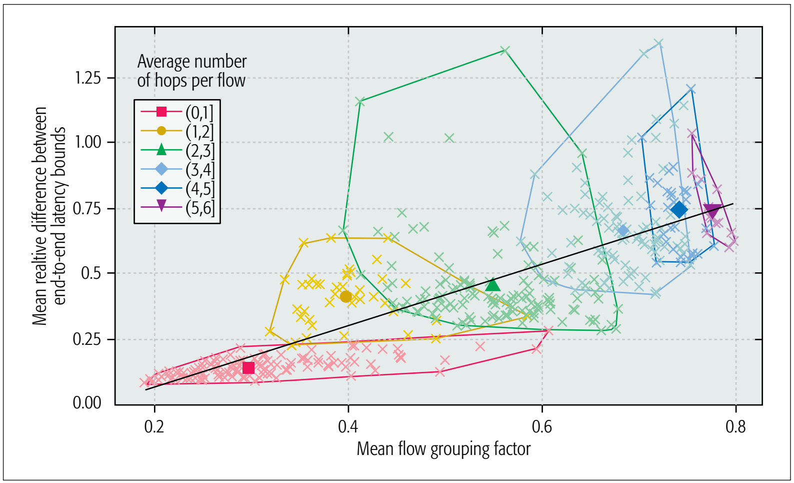


Figure 7. Result of the evaluations on randomly generated Ethernet topologies. Each small cross corresponds to a single topology. The larger points correspond to the centroid of each group.

References

[1] J. Muñoz-Castañer *et al.*, "A Review of Aeronautical Electronics and Its Parallelism with Automotive Electronics," *IEEE Trans. Industrial Elect.*, vol. 58, no. 7, July 2011, pp. 3090–3100.

[2] AUTOSAR Development Cooperation, "AUTOSAR – AUTomotive Open System ARchitecture, Rel. 4.2."

[3] –, "Specification of Timing Extensions, v1.2.0," AUTOSAR Rel. 4.0.3, Sept. 2011.

[4] M. Joseph and P. Pandya, "Finding Response Times in a Real-Time System," *Comp. J.*, vol. 29, no. 5, 1986, pp. 390–95.

[5] F. Eisenbrand and T. Rothvob, "Static-Priority Real-Time Scheduling: Response Time Computation Is NP-Hard," *Proc. IEEE Real-Time Sys. Symp. Comp. Soc.*, Nov. 2008, pp. 397–406.

[6] J. J. Gutiérrez, J. C. Palencia, and M. González Harbour, "Holistic Schedulability Analysis for Multipacket Messages in AFDX Networks," *Real-Time Sys.*, vol. 50, no. 2, Mar. 2014, pp. 230–69.

[7] N. C. Audsley *et al.*, "Fixed Priority Pre-emptive Scheduling: An Historical Perspective," *Real-Time Sys.*, vol. 8, no. 2-3, 1995, pp. 173–98.

[8] K. Tindell, A. Burns, and A. Wellings, "Calculating Controller Area Network (CAN) Message Response Times," *Control Eng. Practice*, vol. 3, no. 8, Aug. 1995, pp. 1163--69.

[9] T. Pop, P. Eles, and Z. Peng, "Holistic Scheduling and Analysis of Mixed Time/Event-Triggered Distributed Embedded Systems," *Proc. 10th Int'l. Symp. Hardware/Software Codesign (CODES)*, May 2002, pp. 187–92.

[10] K. Schmidt and E. G. Schmidt, "Message Scheduling for the FlexRay Protocol: The Static Segment," *IEEE Trans. Vehic. Tech.*, vol. 58, no. 5, June 2009, pp. 2170–79.

[11] J.-Y. Le Boudec and P. Thiran, *Network Calculus: A Theory of Deterministic Queuing Systems for the Internet*, Springer-Verlag, 2001.

[12] J. Grieu, *Analyse et évaluation de techniques de commutation Ethernet pour l'interconnexion des systèmes avioniques*, Ph.D. dissertation, Institut National Polytechnique de Toulouse, Sept. 2004.

[13] A. Bouillard and É. Thierry, "An Algorithmic Toolbox for Network Calculus," *Discrete Event Dynamic Sys.*, vol. 18, no. 1, Mar. 2008, pp. 3–49.

[14] T. Herpel *et al.*, "Stochastic and Deterministic Performance Evaluation of Automotive CAN Communication," *Computer Networks*, vol. 53, no. 8, June 2009, pp. 1171–85.

[15] L. Thiele, S. Chakraborty, and M. Naedele, "Real-Time Calculus for Scheduling Hard Real-Time Systems," *Proc. 2000 IEEE Int'l. Symp. Circuits and Sys.*, vol. 4, May 2000, pp. 101–04.

[16] H. Bauer, J.-L. Scharbarg, and C. Fraboul, "Improving the Worst-Case Delay Analysis of an AFDX Network Using an Optimized Trajectory Approach," *IEEE Trans. Industrial Informatics*, vol. 6, no. 4, Nov. 2010, pp. 521–33.

[17] R. I. Davis *et al.*, "Controller Area Network (CAN) Schedulability Analysis: Refuted, Revisited and Revised," *Real-Time Sys.*, vol. 35, no. 3, Apr. 2007, pp. 239–72.

[18] S. Bondorf and J. B. Schmitt, "The DiscoDNC v2: A Comprehensive Tool for Deterministic Network Calculus," *Proc. 8th Int'l. Conf. Performance Evaluation Methodologies and Tools*, Dec. 2014.

Biographies

Fabien Geyer (fgeyer@net.in.tum.de) is currently with Airbus Group Innovations working on network performance and architectures. He received his M.Eng. in telecommunications from Telecom Bretagne, Brest, France, in 2011 and his Ph.D. degree in computer science from Technische Universität München (TUM), Germany, in 2015. His research interests include formal methods for the performance evaluation and modeling of network architectures and protocols.

Georg Carle (carle@in.tum.de) is a professor at the Department of Informatics of TUM, holding the chair for Network Architectures and Services. He studied at the University of Stuttgart, Brunel University, London, and Ecole Nationale Superieure des Telecommunications, Paris. He received his Ph.D. in computer science from the University of Karlsruhe, and worked as a postdoctoral scientist at Institut Eurecom, Sophia Antipolis, France, at the Fraunhofer Institute for Open Communication Systems, Berlin, and as a professor at the University of Tübingen.

CALL FOR PAPERS
IEEE COMMUNICATIONS MAGAZINE
NEXT GENERATION 911

BACKGROUND

In most parts of the world where a phone is used, users depend on the SOS or emergency service (ES). In North America, this service is accessible by dialing 911 and in most parts of Europe by dialing 112. Since its introduction in the 1950s, the service has adapted to technology changes: from wireline phones with fixed locations to supporting mobile phones. However, the technology is improving faster than the ES infrastructure can keep up. Faster networks coupled with heterogeneous access methods pose a challenge to the evolution of the ES. As if this is not enough, the richer choices available to a user today for communication — video, text messaging, social networking portals like Facebook and Google+, instant messaging, web-based calling, and over-the-top voice over IP applications — make the task of providing ES uniformly ever more difficult. The United States and Europe have decided to approach ES through a clean slate approach. The redesign of the ES in the United States is known as Next Generation 911 (NG911) and in Europe as Next Generation 112 (NG112). The core of NG ES is the Emergency Services IP Network (ESInet). ESInet uses Session Initiation Protocol (SIP) to deliver voice, video, text, and data calls reliably and uniformly to the ES network.

Authors from industry and academia are invited to submit papers for this Feature Topic of *IEEE Communications Magazine* on next generation 911. The Feature Topic scope includes, but is not limited to, the following topics of interest:

- Overview of NG911 implementation efforts
- Technical challenges in implementing NG911 systems
- Status of current implementation of NG911
- Status of NG911 deployment
- Issues in multi-modal NG911 devices
- Impact of social media on NG911 systems
- Societal impacts of the presence (or absence) of such services
- Over-the-top applications and NG911
- Security and privacy of NG911
- Public policy and funding issues with NG911
- Regulatory environment for NG911
- The role of standards (IETF, ITU-T) in NG911

SUBMISSIONS

Articles should be tutorial in nature, with the intended audience being all members of the communications technology community. They should be written in a style comprehensible to readers outside the specialty of the article. Mathematical equations should not be used (in justified cases up to three simple equations are allowed). Articles should not exceed 4500 words. Figures and tables should be limited to a combined total of six. The number of archivable references is not to exceed 15. Complete guidelines for manuscript preparation can be found via the link http://www.comsoc.org/commag/paper-submission-guidelines. Please send a PDF (preferred) or MS-Word formatted paper via Manuscript Central (http://commag-ieee.manuscriptcentral.com). Register or log in, and go to the Author Center. Follow the instructions there. Select "November 2016 / Next-Generation 911".

IMPORTANT DATES

- Manuscript Submission Deadline: March 15, 2016
- Decision Notification : June 30, 2016
- Final Manuscript Due Date: August 31, 2016
- FT Publication Date: November 2016

GUEST EDITORS

Vijay K. Gurbani	Salvatore Loreto	Ravi Subrahmanyan	Carol Davids
Bell Laboratories, Alcatel-Lucent, USA	Ericsson, Sweden	Butterfly Network Inc., USA	Illinois Inst. of Tech., USA
vkg@bell-labs.com	salvatore.loreto@ericsson.com	ravi.subrahmanyan@ieee.org	davids@iit.edu

Massive MIMO: Ten Myths and One Critical Question

Emil Björnson, Erik G. Larsson, and Thomas L. Marzetta

The authors identify 10 myths about Massive MIMO, and explain why they are not true. They also ask a question that is critical for the practical adoption of the technology and which will require intense future research activities to answer properly. They provide references to key technical papers that support their claims.

ABSTRACT

Wireless communications is one of the most successful technologies in modern years, given that an exponential growth rate in wireless traffic has been sustained for over a century (known as Cooper's law). This trend will certainly continue, driven by new innovative applications; for example, augmented reality and the Internet of Things. Massive MIMO has been identified as a key technology to handle orders of magnitude more data traffic. Despite the attention it is receiving from the communication community, we have personally witnessed that Massive MIMO is subject to several widespread misunderstandings, as epitomized by following (fictional) abstract: "*The Massive MIMO technology uses a nearly infinite number of high-quality antennas at the base stations. By having at least an order of magnitude more antennas than active terminals, one can exploit asymptotic behaviors that some special kinds of wireless channels have. This technology looks great at first sight, but unfortunately the signal processing complexity is off the charts and the antenna arrays would be so huge that it can only be implemented in millimeter-wave bands.*" These statements are, in fact, completely false. In this overview article, we identify 10 myths and explain why they are not true. We also ask a question that is critical for the practical adoption of the technology and which will require intense future research activities to answer properly. We provide references to key technical papers that support our claims, while a further list of related overview and technical papers can be found at the Massive MIMO Info Point: http://massivemimo.eu

INTRODUCTION

Massive multiple-input multiple-output (MIMO) is a multi-user MIMO technology where each base station (BS) is equipped with an array of M active antenna elements and utilizes these to communicate with K single-antenna terminals over the same time and frequency band. The general multi-user MIMO concept has been around for decades, but the vision of actually deploying BSs with more than a handful of service antennas is relatively new [1]. By coherent processing of the signals over the array, transmit precoding can be used in the downlink to focus each signal at its desired terminal, and receive combining can be used in the uplink to discriminate between signals sent from different terminals. The more antennas that are used, the finer the spatial focusing can be. An illustration of these concepts is given in Fig. 1a.

The canonical Massive MIMO system operates in time-division duplex (TDD) mode, where the uplink and downlink transmissions take place in the same frequency resource but are separated in time. The physical propagation channels are reciprocal — meaning that the channel responses are the same in both directions — which can be utilized in TDD operation. In particular, Massive MIMO systems exploit the reciprocity to estimate the channel responses on the uplink and then use the acquired channel state information (CSI) for both uplink receive combining and downlink transmit precoding of payload data. Since the transceiver hardware is generally not reciprocal, calibration is needed to exploit the channel reciprocity in practice. Fortunately, the uplink-downlink hardware mismatches only change by a few degrees over a one-hour period and can be mitigated by simple relative calibration methods, even without extra reference transceivers and by only relying on mutual coupling between antennas in the array [2].

There are several good reasons to operate in TDD mode. First, only the BS needs to know the channels to process the antennas coherently. Second, the uplink estimation overhead is proportional to the number of terminals, but independent of M, thus making the protocol fully scalable with respect to the number of service antennas. Furthermore, basic estimation theory tells us that the estimation quality (per antenna) cannot be reduced by adding more antennas at the BS — in fact, the estimation quality improves with M if there is a known correlation structure between the channel responses over the array [3].

Since fading makes the channel responses vary over time and frequency, the estimation and payload transmission must fit into a time/frequency block where the channels are approximately static. The dimensions of this block are essentially given by the coherence bandwidth B_c Hz and the coherence time T_c s, which fit $\tau = B_cT_c$ transmission symbols. Massive MIMO can be implemented using either single-carrier or multi-carrier modulation. We consider multi-carrier orthogonal frequency-division multiplexing (OFDM) modulation here for simplicity, because

Emil Björnson and Erik G. Larsson are with Linköping University; Thomas L. Marzetta is with Bell Labs, Nokia.

0163-6804/16/$25.00 © 2016 IEEE

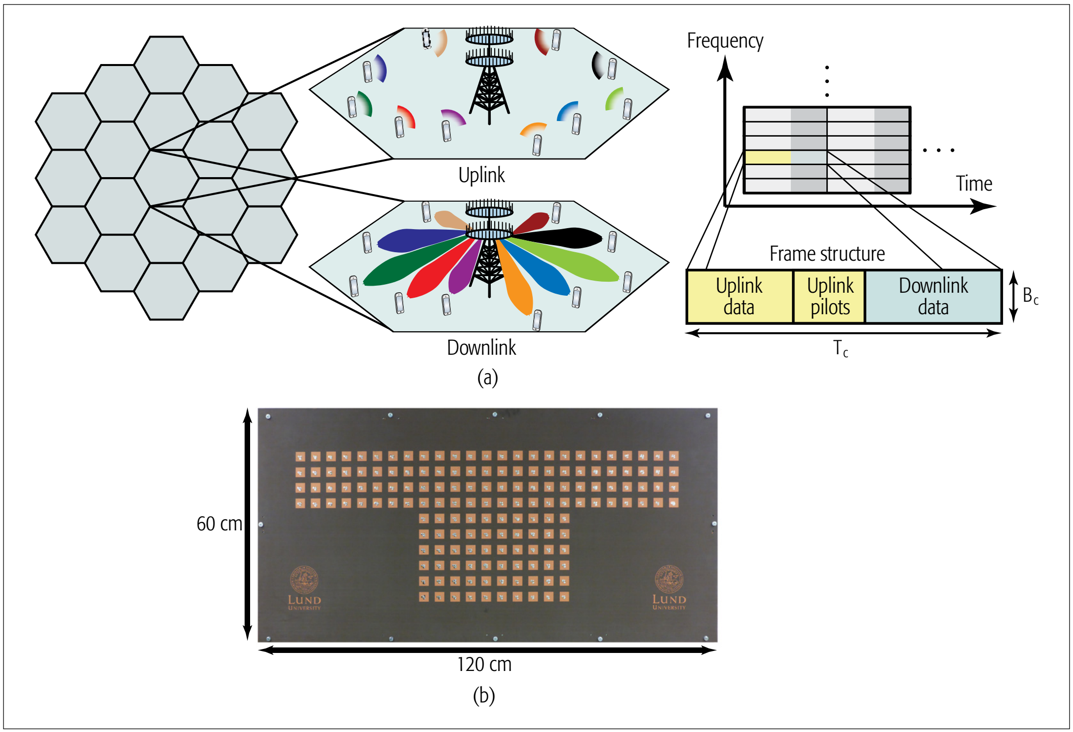


Figure 1. Example of a Massive MIMO system: a) illustration of the uplink and downlink in line-of-sight propagation, where each BS is equipped with M antennas and serves K terminals. The TDD transmission frame consists of $\tau = B_cT_c$ symbols. By capitalizing on channel reciprocity, there is payload data transmission in both the uplink and downlink, but only pilot transmission in the uplink; b) photo of the antenna array of the LuMaMi testbed at Lund University in Sweden [2]. The array consists of 160 dual-polarized patch antennas. It is designed for a carrier frequency of 3.7 GHz, and the element spacing is 4 cm (half a wavelength).

Since the uplink and downlink channels are reciprocal in TDD systems, there is a strong connection between receive combining in the uplink and the transmit precoding in the downlink. This is known as uplink-downlink duality.

the coherence block has a neat interpretation: it spans a number of subcarriers over which the channel frequency response is constant, and a number of OFDM symbols over which the channel is constant (Fig. 1a). The channel coherency depends on the propagation environment, user mobility, and carrier frequency.

Linear Processing

The payload transmission in Massive MIMO is based on linear processing at the BS. In the uplink, the BS has M observations of the multiple access channel from the K terminals. The BS applies linear receive combining to discriminate the signal transmitted by each terminal from the interfering signals. The simplest choice is maximum ratio (MR) combining, which uses the channel estimate of a terminal to maximize the strength of that terminal's signal by adding the signal components coherently. This results in a signal amplification proportional to M, which is known as an array gain. Alternative choices are zero-forcing (ZF) combining, which suppresses inter-cell interference at the cost of reducing the array gain to $M - K + 1$, and minimum mean squared error (MMSE) combining that balances between amplifying signals and suppressing interference.

Receive combining creates one effective scalar channel per terminal where the intended signal is amplified and/or the interference is suppressed. Any judicious receive combining will improve by adding more BS antennas, since there are more channel observations to utilize. The remaining interference is typically treated as extra additive noise; thus, conventional single-user detection algorithms can be applied. Another benefit of the combining is that small-scale fading averages out over the array, in the sense that its variance decreases with M. This is known as *channel hardening* and is a consequence of the law of large numbers.

Since the uplink and downlink channels are reciprocal in TDD systems, there is a strong connection between receive combining in the uplink and transmit precoding in the downlink [4]. This is known as uplink-downlink duality. Linear precoding based on MR, ZF, or MMSE principles can be applied to focus each signal on its desired terminal (and possibly mitigate interference toward other terminals).

Many convenient closed-form expressions for the achievable uplink or downlink spectral efficiency (per cell) can be found in the literature [4–6, references therein]. We provide an example for i.i.d. Rayleigh fading channels with MR processing, just to show how beautifully simple these expressions are:

$$K \cdot \left(1 - \frac{K}{\tau}\right) \cdot \log_2\left(1 + \frac{c_{\text{CSI}} \cdot M \cdot \text{SNR}_{u/d}}{K \cdot \text{SNR}_{u/d} + 1}\right) \quad \text{[bit/s/Hz/cell]} \tag{1}$$

where K is the number of terminals, $(1 - (K/\tau))$ is the loss from pilot signaling, and $\text{SNR}_{u/d}$ equals the uplink signal-to-noise ratio (SNR), SNR_u, when Eq. 1 is used to compute the uplink performance. Similarly, we let $\text{SNR}_{u/d}$ be the downlink

The interest in the Massive MIMO technology has grown quickly in recent years, but at the same time we have noticed that there are several widespread myths or misunderstandings around its basic characteristics. This article inspects ten common beliefs concerning Massive MIMO and explains why they are erroneous.

SNR, SNR_d, when Eq. 1 is used to measure the downlink performance. In both cases, $c_{\mathrm{CSI}} = (1 + 1/(K \cdot \mathrm{SNR}_u))^{-1}$ is the quality of the estimated CSI, proportional to the mean squared power of the MMSE channel estimate (where $c_{\mathrm{CSI}} = 1$ represents perfect CSI). Notice how the numerator inside the logarithm increases proportionally to M due to the array gain and that the denominator represents the interference plus noise.

While canonical Massive MIMO systems operate with single-antenna terminals, the technology also handles N-antenna terminals. In this case, K denotes the number of simultaneous data streams, and Eq. 1 describes the spectral efficiency per stream. These streams can be divided over anything from K/N to K terminals, but we focus on $N = 1$ in this article for clarity in presentation.

Myths and Misunderstandings About Massive MIMO

The interest in Massive MIMO technology has grown quickly in recent years, but at the same time we have noticed that there are several widespread myths or misunderstandings around its basic characteristics. This article inspects 10 common beliefs concerning Massive MIMO and explains why they are erroneous.

Myth 1: Massive MIMO Is Only Suitable for Millimeter-Wave Bands

Antenna arrays are typically designed with an antenna spacing of at least $\lambda_c/2$, where λ_c is the wavelength at the intended carrier frequency f_c. Larger antenna spacings provide less correlated channel responses over the antennas and thus more spatial diversity, but the important thing in Massive MIMO is that each terminal has distinct spatial channel characteristics and not that the antennas observe uncorrelated channels. The wavelength is inversely proportional to f_c, thus smaller form factors are possible at higher frequencies (e.g., in millimeter bands). Nevertheless, Massive MIMO arrays have realistic form factors also at a typical cellular frequency of $f_c = 2$ GHz; the wavelength is $\lambda_c = 15$ cm and up to 400 dual-polarized antennas can thus be deployed in a 1.5 × 1.5 m array. This should be compared to contemporary cellular networks that utilize vertical panels, around 1.5 m tall and 20 cm wide, each comprising many interconnected radiating elements that provide a fixed directional beam. A 4-MIMO setup uses four such panels with a combined area comparable to the exemplified Massive MIMO array.

Example: Figure 1b shows a picture of the array in the LuMaMi Massive MIMO testbed [2]. It is designed for a carrier frequency of $f_c = 3.7$ GHz, which gives $\lambda_c = 8.1$ cm. The panel is 60 × 120 cm (i.e., equivalent to a 53-in flat-screen TV) and features 160 dual-polarized antennas, while leaving plenty of room for additional antenna elements. Such a panel could easily be deployed at the facade of a building.

The research on Massive MIMO has thus far focused on cellular frequencies below 6 GHz, where the transceiver hardware is very mature. The same concept can definitely be applied in millimeter-wave bands as well — many antennas might even be required in these bands since the effective area of an antenna is much smaller. However, the hardware implementation will probably be quite different from what has been considered in the Massive MIMO literature [7]. Moreover, for the same mobility the coherence time will be an order of magnitude shorter due to higher Doppler spread [8], which reduces the spatial multiplexing capability. In summary, Massive MIMO for cellular bands and for millimeter bands are two feasible branches of the same tree, where the former is mature, and the latter is greatly unexplored and possesses many exciting research opportunities.

Myth 2: Massive MIMO Only Works in Rich-Scattering Environments

The channel response between a terminal and the BS can be represented by an M-dimensional vector. Since the K channel vectors are mutually non-orthogonal in general, advanced signal processing (e.g., dirty paper coding) is needed to suppress interference and achieve the sum capacity of the multi-user channel. *Favorable propagation* (FP) denotes an environment where the K users' channel vectors are mutually orthogonal (i.e., their inner products are zero). FP channels are ideal for multi-user transmission since the interference is removed by simple linear processing (i.e., MR and ZF) that utilizes the channel orthogonality [9]. The question is whether there are any FP channels in practice.

An approximate form of favorable propagation is achieved in non-line-of-sight (non-LOS) environments with rich scattering, where each channel vector has independent stochastic entries with zero mean and identical distribution. Under these conditions, the inner products (normalized by M) go to zero as more antennas are added; this means that the channel vectors get closer and closer to orthogonal as M increases. The sufficient condition above is satisfied for Rayleigh fading channels, which are considered in the vast majority of works on Massive MIMO, but approximate favorable propagation is obtained in many other situations as well.

Example: Suppose the BS uses a uniform linear array (ULA) with half-wavelength antenna spacing. We compare two extreme opposite environments in Fig. 2a: non-LOS isotropic scattering (i.i.d. Rayleigh fading) and LOS propagation. In the LOS case, the angle to each terminal determines the channel, and this angle is uniformly distributed. The simulation considers $M = 100$ service antennas, $K = 12$ terminals, perfect CSI, and an uplink SNR of $\mathrm{SNR}_u = -5$ dB. The figure shows the cumulative probability of achieving a certain sum capacity, and the dashed vertical lines in Fig. 2a indicate the sum capacity achieved under FP.

The isotropic scattering case provides, as expected, a sum capacity close to the FP upper bound. The sum capacity in the LOS case is similar to that of isotropic scattering in the majority of cases, but there is a 10 percent risk that the LOS performance loss is more than 10 percent. The reason is that there is substantial probability that two terminals have similar angles [9]. A simple solution is to drop a few "worst" terminals from service in each coherence block; Fig. 2a illustrates this by dropping 2 out of the 12 terminals. In this case, LOS propagation offers similar performance as isotropic fading.

Since isotropic and LOS propagation represent two rather "extreme" environments, and both are favorable for the operation of Massive MIMO, we expect that real propagation environments — which are likely to lie between these extremes — would also be favorable. This observation offers an explanation for the FP characteristics of Massive MIMO channels consistently seen in measurement campaigns (e.g., in [10]).

Myth 3: Massive MIMO Performance Can Be Achieved by Open-Loop Beamforming Techniques

The precoding and combining in Massive MIMO rely on measured/estimated channel responses to each of the terminals and provide an array gain of $c_{\mathrm{CSI}}M$ in any propagation environment [9] — without relying on any particular array geometry or calibration. The BS obtains estimates of the channel responses in the uplink by receiving K mutually orthogonal pilot signals transmitted by the K terminals. Hence, the required pilot resources scale with K but not with M.

By way of contrast, open-loop beamforming (OLB) is a classic technique where the BS has a codebook of L predetermined beamforming vectors and sends a downlink pilot sequence through each of them. Each terminal then reports which of the L beams has the largest gain and feeds back an index in the uplink (using $\log_2(L)$ bits). The BS transmits to each of the K terminals through the beam that each terminal reported to be the best. OLB is particularly intuitive in LOS propagation scenarios, where the L beamforming vectors correspond to different angles of departure from the array. The advantage of OLB is that no channel reciprocity or high-rate feedback is needed. There are two serious drawbacks, however. First, the pilot resources required are significant, because L pilots are required in the downlink and L should be proportional to M (in order to explore and enable exploitation of all channel dimensions). Second, the $\log_2(L)$-bits-per-terminal feedback does not enable the BS to learn the channel responses accurately enough to facilitate true spatial multiplexing. This last point is illustrated by the next example.

Example: Figure 2b compares the array gain of Massive MIMO with that of OLB for the same two cases as in Myth 2:
- Non-LOS isotropic scattering (i.i.d. Rayleigh fading)
- LOS propagation with a ULA

The linear array gain with MR processing is $c_{\mathrm{CSI}}M$, where $c_{\mathrm{CSI}} = (1 + 1/(K \cdot \mathrm{SNR}_u))^{-1}$ is the quality of the CSI (proportional to the mean-squared power of the estimate). With $K = 12$ and $\mathrm{SNR}_u = -5$ dB, the array gain is $c_{\mathrm{CSI}}M \approx 0.79M$ for Massive MIMO in both cases. For OLB, we use the codebook size of $L = M$ for $M \leq 50$ and $L = 50$ for $M > 50$ in order to model a maximum permitted pilot overhead. The codebooks are adapted to each scenario by quantizing the search space uniformly. OLB provides a linear slope in Fig. 2b for $M \leq 50$ in the LOS case, but the array gain saturates when the maximum codebook size manifests itself — this would happen even earlier if the antennas are slightly misplaced in the ULA. The performance is much worse in the isotropic case, where only the logarithmic array gain $\log(M)$ is obtained before the saturation occurs. The explanation is the finite-size codebook, which needs to quantize all M dimensions in the isotropic case since all directions of the M-dimensional channel vector are equally probable. In contrast, an LOS channel direction is fully determined by the angle of arrival, and thus the codebook only needs to quantize this angle.

In summary, conventional OLB provides decent array gains for small arrays in LOS propagation, but is not scalable (in terms of overhead or array tolerance) and not able to handle isotropic fading. In practice, the channel of a particular terminal might not be isotropically distributed,

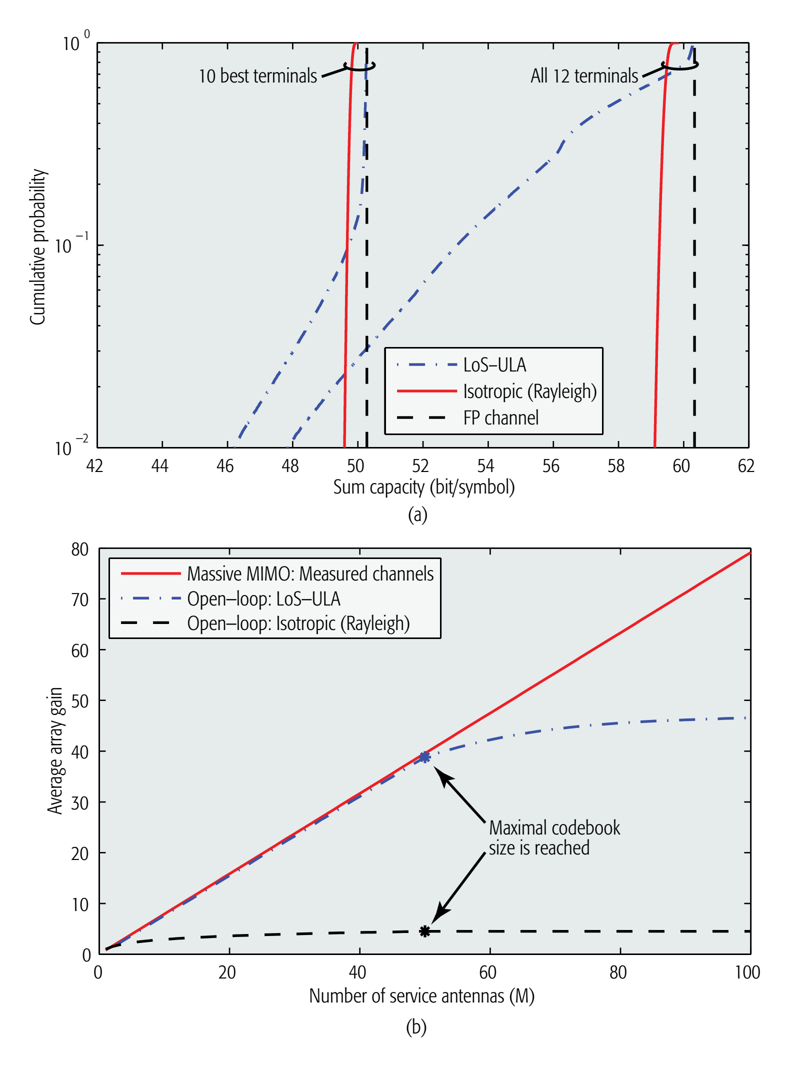


Figure 2. Comparison of system behavior with i.i.d. Rayleigh fading and LOS propagation. There are $K = 12$ terminals and $\mathrm{SNR}_u = -5$ dB: a) cumulative distribution of the uplink sum capacity with $M = 100$ service antennas, when either all 12 terminals or only the 10 best terminals are served; b) average array gain achieved for different number of service antennas. The uplink channel estimation in Massive MIMO always provides a linear slope, while the performance of open-loop beamforming depends strongly on the propagation environment and codebook size.

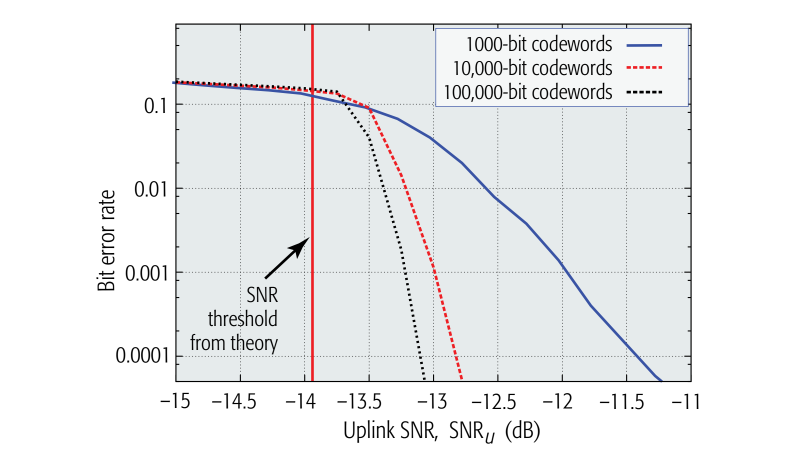


Figure 3. Empirical uplink link performance of Massive MIMO with $M = 100$ antennas and $K = 30$ terminals using QPSK modulation with 1/2 coding rate and estimated channels. The vertical red line is the SNR threshold where zero BER can be achieved for infinitely long codewords, according to the spectral efficiency expression in Eq. 1.

but have distinct statistical spatial properties. The codebook in OLB unfortunately cannot be tailored to a specific terminal, but needs to explore all channel directions that are possible for the array. For large arrays with arbitrary propagation properties, the channels must be measured by pilot signaling as is done in the Massive MIMO protocol.

Myth 4: The Case for Massive MIMO Relies on Asymptotic Results

The seminal work [1] on Massive MIMO studied the asymptotic regime where the number of service antennas $M \to \infty$. Numerous later works, including [4–6], have derived closed-form achievable spectral efficiency expressions (unit: bits per second per Hertz) that are valid for any number of antennas and terminals, any SNR, and any choice of pilot signaling. These formulas do not rely on idealized assumptions such as perfect CSI, but rather on worst case assumptions regarding the channel acquisition and signal processing. Although the total spectral efficiency per cell is greatly improved with Massive MIMO technology, the anticipated performance per user lies in the conventional range of 1–4 b/s/Hz [4]. This is part of the range where off-the-shelf channel codes perform close to the Shannon limits.

Example: To show these properties, Fig. 3 compares the empirical link performance of a Massive MIMO system with the uplink spectral efficiency expression in Eq. 1. We consider $M = 100$ service antennas, $K = 30$ terminals, and estimated channels using one pilot per terminal. Each terminal transmits with quadrature phase shift keying (QPSK) modulation followed by low density parity check (LDPC) coding with rate 1/2, leading to a net spectral efficiency of 1 b/s/Hz/terminal; that is, 30 b/s/Hz in total for the cell. By equating Eq. 1 to the same target of 30 b/s/Hz, we obtain the uplink SNR threshold $\text{SNR}_u = -13.94$ dB as the value needed to achieve this spectral efficiency.

Figure 3 shows the bit error rate (BER) performance for different lengths of the codewords, and the BER curves drop quickly as the length of the codewords increases. The vertical line indicates $\text{SNR}_u = -13.94$ dB, where zero BER is achievable as the codeword length goes to infinity. Performance close to this bound is achieved even at moderate codeword lengths, and part of the gap is also explained by the shaping loss of QPSK modulation and the fact that the LDPC code is optimized for additive white Gaussian noise (AWGN) channels (which is actually a good approximation in Massive MIMO due to the channel hardening). Hence, expressions such as Eq. 1 are well suited to predict the performance of practical systems and useful for resource allocation tasks such as power control (see Myth 9).

Myth 5: Too Much Performance Is Lost by Linear Processing

Favorable propagation, where the terminals' channels are mutually orthogonal, is a property that is generally not fully satisfied in practice; see Myth 2. Whenever there is a risk for inter-user interference, there is room for interference suppression techniques. Nonlinear signal processing schemes achieve the sum capacity under perfect CSI: dirty paper coding (DPC) in the downlink and successive interference cancellation (SIC) in the uplink. DPC/SIC remove interference in the encoding/decoding step by exploiting knowledge of what certain interfering streams will be. In contrast, linear processing can only reject interference by linear projections (e.g., as done with ZF). The question is how much performance is lost by linear processing as compared to the optimal DPC/SIC.

Example: A quantitative comparison is provided in Fig. 4a considering the sum capacity of a single cell with perfect CSI (since the capacity is otherwise unknown). The results are representative for both the uplink and downlink due to duality. There are $K = 20$ terminals and a variable number of service antennas. The channels are i.i.d. Rayleigh fading and $\text{SNR}_u = \text{SNR}_d = -5$ dB.

Figure 4a shows that there is indeed a performance gap between the capacity-achieving DPC/SIC and the suboptimal ZF, but the gap reduces quickly with M since the channels decorrelate — all the curves get closer to the FP curve. Nonlinear processing only provides a large gain over linear processing when $M \approx K$, while the gain is small in Massive MIMO cases with $M/K > 2$. Interestingly, we can achieve the same performance as with DPC/SIC by using ZF processing with a few extra antennas (e.g., 10 antennas in this example), which is a reasonable price to pay for the much relaxed computational complexity of ZF. The gap between ZF and MR shrinks considerably when inter-cell interference is considered, as shown below.

Myth 6: Massive MIMO Requires an Order of Magnitude More Antennas than Users

For a given set of terminals, the spectral efficiency always improves by adding more service antennas, because of the larger array gain and the FP property described in Myth 2. This might be the reason Massive MIMO is often referred to as systems with at least an order of magnitude more service antennas than terminals; that is, $M/K > 10$. In general, the number of service antennas, M, is fixed in a deployment and not a variable, while the number of terminals, K, is the actual design parameter. The scheduling algorithm decides how many terminals are admitted in a certain coherence block, with the goal of maximizing some predefined system performance metric.

Example: Suppose the sum spectral efficiency is the metric considered in the scheduler. Figure 4b shows this metric as a function of the number of scheduled terminals for a multi-cellular Massive MIMO deployment of the type considered in [4]. There are $M = 100$ service antennas per cell. The results are applicable in both the uplink and the downlink if power control is applied to provide an SNR of –5 dB for every terminal. A relatively short coherence block of $\tau = 200$ symbols is considered, and the pilot reuse across cells is optimized (this is why the curves are not smooth). The operating points that maximize the performance for ZF and MR processing are marked, and the corresponding values of the ratio M/K are indicated. Interestingly, the optimized operating points are all in the range $M/K < 10$; thus, it is not only possible to let M and K be at the same order of magnitude, it can even be desirable. With MR processing, the considered Massive MIMO system operates efficiently also at $M = K = 100$, which gives $M/K = 1$; the rate per terminal is small at this operating point, but the sum spectral efficiency is not. We also stress that there is a wide range of K-values that provides almost the same sum performance, showing the ability to share the throughput between many or few terminals by scheduling.

In summary, there are no strict requirements on the relation between M and K in Massive MIMO. If one would like to give a simple definition of a Massive MIMO setup, it is a system with unconventionally many active antenna elements, M, that can serve an unconventionally large number of terminals, K. One should avoid specifying a certain ratio M/K, since it depends on a variety of conditions, such as the system performance metric, propagation environment, and coherence block length.

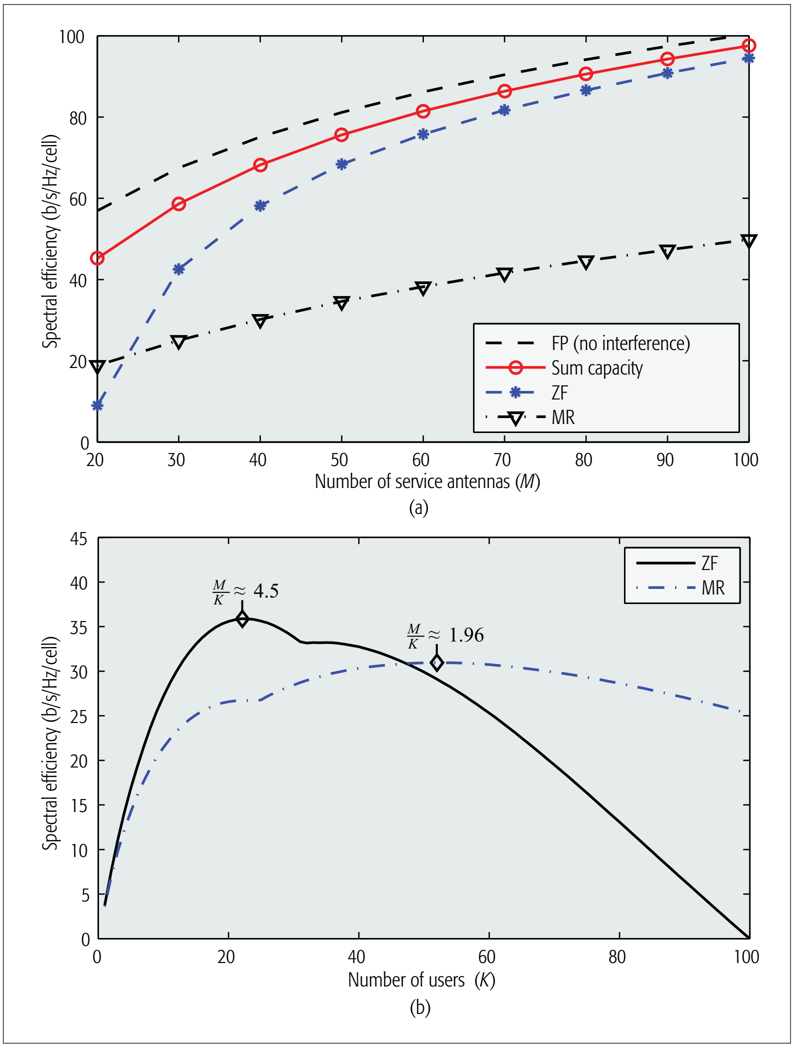


Figure 4. Sum spectral efficiency for i.i.d. Rayleigh fading channels with linear processing: a) the sum capacity achieved by DPC/SIC is compared to linear processing, assuming perfect CSI, no inter-cell interference, and $K = 20$ terminals. The loss incurred by linear processing is large when $M \approx K$, but reduces quickly as the number of antennas increases. In fact, ZF with around $M + 10$ antennas gives performance equivalent to the capacity with M antennas; b) performance in a multi-cellular system with a coherence block of $\tau = 200$ symbols, $M = 100$ service antennas, estimated CSI, and an SNR of –5 dB. The performance is shown as a function of K, with ZF and MR processing. The maximum at each curve is marked, and it is clear that $M/K < 10$ at these operating points.

Myth 7: A New Terminal Cannot Join the System Since There is No Initial Array Gain

The coherent processing in Massive MIMO improves the effective SNR by a factor $c_{\text{CSI}}M$, where $0 < c_{\text{CSI}} \leq 1$ is the CSI quality (see Myth 3 for details). This array gain enables the system to operate at lower SNRs than contemporary systems. As seen from the factor c_{CSI}, the BS needs to estimate the current channel response, based on uplink pilots, to capitalize on the array gain. When a previously inactive terminal wishes to send or request data, it can therefore pick one of the unused pilot sequences and contact the BS using that pilot. The system can, for example, be implemented by reserving a few pilots for random access, while all active terminals use other pilots to avoid collisions. It is less clear how the BS should act when contacting a terminal that is currently inactive; it cannot exploit any array gain since this terminal has not sent a pilot.

This question was considered in [11], and the solution is quite straightforward to implement. Instead of sending precoded downlink signals to the K terminals, the BS can occasionally utilize the same combined transmit power to only broadcast control information within the cell (e.g., to contact inactive terminals). Due to the lack of array gain, this broadcast signal will be $c_{\text{CSI}}M/K$ times weaker than the user-specific precoded signals. We recall that $M/K < 10$ at many operating points of practical interest, which was noted in Myth 6 and exemplified in Fig. 4b. The "loss" $c_{\text{CSI}}M/K$ in effective SNR is partially compensated by the fact that the control signals are not exposed to intra-cell interference, while further improvements in reliability can be achieved

Receiver noise and data signals associated with other terminals are two prime examples of undesired additive quantities that are mitigated by the coherent processing. There is also a third important category: distortions caused by impairments in the transceiver hardware.

using stronger channel codes. Since there is no channel hardening, we can also use classical diversity schemes, such as space-time codes and coding over subcarriers, to mitigate small-scale fading.

In summary, control signals can also be transmitted from large arrays without the need for an array gain. The numerical examples in [11] show that the control data rate is comparable to the individual precoded payload data rates at typical operating points (due to the lack of intra-cell interference and the concentration of transmit power), but the multiplexing gain is lost since one signal is broadcasted instead of precoded transmission of K separate signals.

Myth 8: Massive MIMO Requires High Precision Hardware

One of the main features of Massive MIMO is coherent processing over the M service antennas, using measured channel responses. Each desired signal is amplified by adding the M signal components coherently, while uncorrelated undesired signals are not amplified since their components add up noncoherently.

Receiver noise and data signals associated with other terminals are two prime examples of undesired additive quantities that are mitigated by coherent processing. There is also a third important category: distortions caused by impairments in the transceiver hardware. There are numerous impairments in practical transceivers; for example, nonlinearities in amplifiers, phase noise in local oscillators, quantization errors in analog-to-digital converters, I/Q imbalances in mixers, and non-ideal analog filters. The combined effect of these impairments can be described either stochastically [12] or by hardware-specific deterministic models [13]. In any case, most hardware impairments result in additive distortions that are substantially uncorrelated with the desired signal, plus a power loss and phase rotation of the desired signals. The additive distortion noise caused at the BS has been shown to vanish with the number of antennas [12], just like conventional noise and interference, while the phase rotations from phase noise remain but are not more harmful to Massive MIMO than to contemporary systems. We refer to [12, 13] for numerical examples that illustrate these facts.

In summary, the Massive MIMO gains do not require high-precision hardware; in fact, lower hardware precision can be handled than in contemporary systems since additive distortions are suppressed in the processing. Another reason for the robustness is that Massive MIMO can achieve extraordinary spectral efficiencies by transmitting low-order modulations to a multitude of terminals, while contemporary systems require high-precision hardware to support high-order modulations to a few terminals.

Myth 9: With So Many Antennas, Resource Allocation and Power Control Are Hugely Complicated

Resource allocation usually means that the time-frequency resources are divided between the terminals to satisfy user-specific performance constraints, find the best subcarriers for each terminal, and combat the small-scale fading by power control. Frequency-selective resource allocation can bring substantial improvements when there are large variations in channel quality over the subcarriers, but it is also demanding in terms of channel estimation and computational overhead since the decisions depend on the small-scale fading, which varies on the order of milliseconds. If the same resource allocation concepts were applied in Massive MIMO systems, with tens of terminals at each of the thousands of subcarriers, the complexity would be huge.

Fortunately, the channel hardening effect in Massive MIMO means that the channel variations are negligible over the frequency domain and mainly depend on large-scale fading in the time domain, which typically varies 100–1000 times slower than small-scale fading. This renders the conventional resource allocation concepts unnecessary. The whole spectrum can be simultaneously allocated to each active terminal, and the power control decisions are made jointly for all subcarriers based only on the large-scale fading characteristics.

Example: Suppose we want to provide uniformly good performance to the terminals in the downlink. This resource allocation problem is only nontrivial when the K terminals have different average channel conditions. Hence, we associate the kth terminal with a user-specific CSI quality $c_{\mathrm{CSI},k}$, a nominal downlink SNR value of $\mathrm{SNR}_{d,k}$ when the transmit power is shared equally over the terminals, and a power-control coefficient $\eta_k \in [0, K]$ that is used to reallocate the power over the terminals (under the constraint $\Sigma_{k=1}^{K} \eta_k \leq K$). By generalizing the spectral efficiency expression in Eq. 1 to cover these user-specific properties (and dropping the constant pre-log factor), we arrive at the following optimization problem:

$$
\begin{aligned}
&\underset{\substack{\eta_1,\ldots,\eta_K \in [0,K] \\ \sum_k \eta_k \leq K}}{\text{maximize}} \ \min_k \log_2\left(1+\frac{c_{\mathrm{CSI},k}\cdot M \cdot \mathrm{SNR}_{d,k}\cdot \eta_k}{\mathrm{SNR}_{d,k}\sum_{i=1}^{K}\eta_i+1}\right) \\
&\qquad\qquad\Updownarrow \\
&\underset{\substack{\eta_1,\ldots,\eta_K \in [0,K] \\ \sum_k \eta_k \leq K,\, R\geq 0}}{\text{maximize}} \quad R \\
&\text{subject to} \\
&c_{\mathrm{CSI},k}\cdot M \cdot \mathrm{SNR}_{d,k}\cdot \eta_k \geq \\
&\qquad (2^R-1)\left(\mathrm{SNR}_{d,k}\sum_{i=1}^{K}\eta_i+1\right) \ \text{for } k=1,\ldots,K.
\end{aligned}
\tag{2}
$$

This resource allocation problem is known as max-min fairness, and since we maximize the worst terminal performance, the solution gives the same performance to all terminals. The second formulation in Eq. 2 is the epigraph form of the original formulation. From this reformulation it is clear that all the constraints are linear functions of the power-control coefficients $\eta_1, \ldots, \eta_K$; thus, Eq. 2 is a linear optimization problem for every fixed worst terminal performance R. The whole problem is solved by line search over R to find the largest R for which the constraints are feasible. In other words, the power control

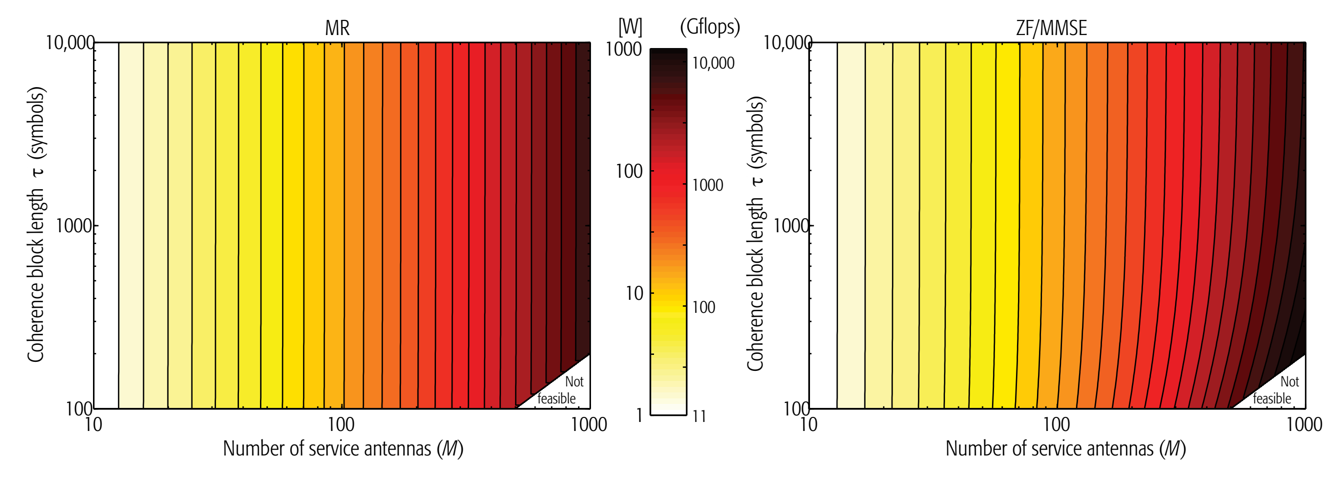


Figure 5. Computational complexity (in flops) of the main baseband signal processing operations in an OFDM Massive MIMO setup: FFTs, channel estimation, precoding/combining of payload data, and computation of precoding/combining matrices. The complexity is also converted into an equivalent power consumption using a typical computational efficiency of 12.8 Gflops/W [15].

optimization is a so-called quasi-linear problem and can be solved by standard techniques (e.g., interior point methods) with low computational complexity. We stress that the power control in Eq. 2 only depends on the large-scale fading; the same power control can be applied on all subcarriers and over a relatively long time period.

To summarize, the resource allocation can be greatly simplified in Massive MIMO systems. It basically reduces to admission control (which terminals should be active) and long-term power control (in many cases a quasi-linear problem). The admitted terminals may use the full bandwidth — there is no need for frequency-selective allocation when there is no frequency-selective fading. The complexity of power control problems such as Eq. 2 scales with the number of terminals, but is independent of the number of antennas and subcarriers.

Myth 10: With So Many Antennas, the Signal Processing Complexity Will Be Overwhelming

The baseband processing is naturally more computationally demanding when having $M > 1$ BS antennas that serve $K > 1$ terminals, compared to only serving one terminal using one antenna port. The important question is how fast the complexity increases with M and K; is the complexity of a typical Massive MIMO setup manageable using contemporary or future hardware generations, or is it totally off the charts?

In an OFDM implementation of Massive MIMO, the signal processing needs to take care of a number of tasks; for example, fast Fourier transform (FFT), channel estimation using uplink pilots, precoding/combining of each payload data symbol (a matrix-vector multiplication), and computation of the precoding/combining matrices. The complexity of these signal processing tasks scales linearly with the number of service antennas, and everything except the FFT complexity also increases with the number of terminals. The computation of a precoding/combining matrix depends on the processing scheme: MR has linear scaling with K, while ZF/MMSE have faster scaling since these involve matrix inversions. Nevertheless, all of these processing tasks are standard operations for which the required number of floating point operations per second (flops) are straightforward to compute [14]. This can provide rough estimates of the true complexity, which also depends strongly on the implementation and hardware characteristics.

Example: To exemplify the typical complexity, suppose we have 20 MHz bandwidth, 1200 OFDM subcarriers, and an oversampling factor of 1.7 in the FFTs. Figure 5 shows how the computational complexity depends on the length τ of the coherence block and on the number of service antennas M. The number of terminals are taken as $K = M/5$, which was a reasonable ratio according to Fig. 4b. Results are given for both MR and ZF/MMSE processing at the BS. Each color in Fig. 5 represents a certain complexity interval, and the corresponding colored area shows the operating points that give a complexity in this interval. The complexities can also be mapped into a corresponding power consumption; to this end, we consider the state-of-the-art digital signal processor (DSP) in [15] which has a computational efficiency of $E = 12.8$ Gflops/W.

Increasing the coherence block means that the precoding/combining matrices are computed less frequently, which reduces the computational complexity. This gain is barely visible for MR, but can be substantial for ZF/MMSE when there are many antennas and terminals (since the complexity of the matrix inversion is then large). For the typical operating point of $M = 200$ antennas, $K = 40$ terminals, and $\tau = 200$ symbols, the complexity is 559 Gflops with MR and 646 Gflops with ZF/MMSE. This corresponds to 43.7 W and 50.5 W, respectively, using the exemplified DSP. These are feasible complexity numbers even with contemporary technology, in particular, because the majority of the computations can be parallelized and distributed over the antennas. It is only the computation of the precoding/combining matrices and the power control that may require a centralized implementation.

In summary, the baseband complexity of Mas-

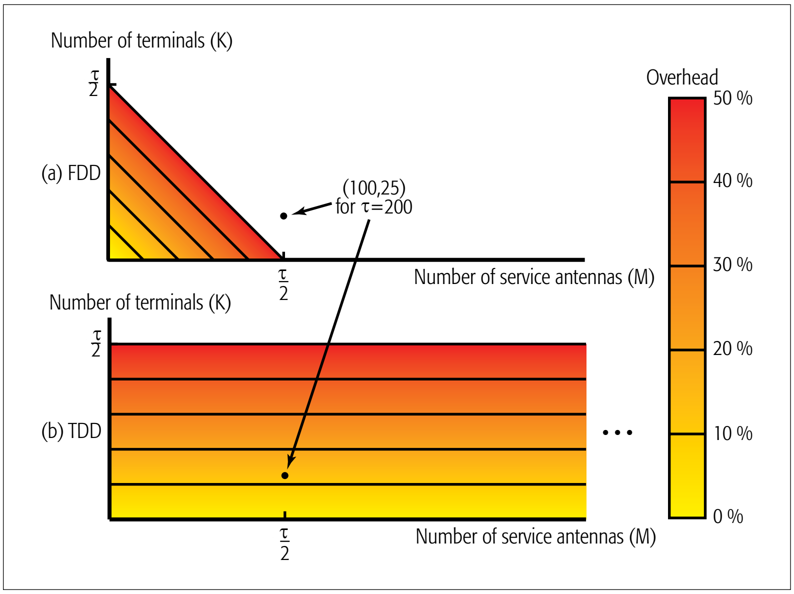


Figure 6. Illustration of the typical overhead signaling in Massive MIMO based on a) FDD; b) TDD operation. The main difference is that FDD limits the number of antennas, while TDD can have any number of antennas. For a coherence block with $\tau = 200$, even a modest Massive MIMO setup with $M = 100$ and $K = 25$ is only supported in TDD operation.

sive MIMO is well within the practical realm. The complexity difference between MR and ZF/MMSE is relatively small since the precoding/combining matrices are only computed once per coherence block — the bulk of the complexity comes from FFTs and matrix-vector multiplications performed on a per symbol basis.

The Critical Question

Can Massive MIMO Work in FDD Operation?

The canonical Massive MIMO protocol, illustrated in Fig. 1a, relies on TDD operation. This is because the BS processing requires CSI, and the overhead of CSI acquisition can be greatly reduced by exploiting channel reciprocity. Many contemporary networks are, however, operating in frequency-division duplex (FDD) mode, where the uplink and downlink use different frequency bands, and channel reciprocity cannot be harnessed. The adoption of Massive MIMO technology would be much faster if the concept could be adapted to also operate in FDD. But the critical question is: can Massive MIMO work in FDD operation?

To explain the difference between TDD and FDD, we describe the related CSI acquisition overhead. Recall that the length of a coherence block is $\tau = B_c T_c$ symbols. Massive MIMO in TDD mode uses K uplink pilot symbols per coherence block, and the channel hardening eliminates the need for downlink pilots. In contrast, a basic FDD scheme requires M pilot symbols per coherence block in the downlink band, and K pilot symbols plus feedback of M channel coefficients per terminal on the uplink band (e.g., based on analog feedback using M symbols and multiplexing of K coefficients per symbol). Hence, it is the $M + K$ uplink symbols per coherence block that is the limiting factor in FDD. The feasible operating points (M, K) with TDD and FDD operations are illustrated in Fig. 6 as a function of τ, and are colored based on the percentage of overhead that is needed.

The main message from Fig. 6 is that TDD operation supports any number of service antennas, while there is a trade-off between antennas and terminals in FDD operation. The extra FDD overhead might be of little importance when $\tau = 5000$ (e.g., in low-mobility scenarios at low frequencies), but it is a critical limitation when $\tau = 200$ (e.g., for high-mobility scenarios or at higher frequencies). For instance, the modest operating point of $M = 100$ and $K = 25$ is marked in Fig. 6 for the case of $\tau = 200$. We recall that this was a good operating point in Fig. 4b. This point can be achieved with only 12.5 percent pilot overhead in TDD operation, while FDD cannot even support it by spending 50 percent of the resources on overhead signaling. It thus appears that FDD can only support Massive MIMO in special low-mobility and low-frequency scenarios.

Motivated by the demanding CSI acquisition in FDD mode, several research groups have proposed methods to reduce the overhead; two excellent examples are [3, 8]. Generally speaking, these methods assume that there is some kind of channel sparsity that can be utilized; for example, a strong spatial correlation where only a few strong eigendirections need to be estimated or the impulse responses are sparse in time. While these kinds of methods achieve their goals, we stress that the underlying sparsity assumptions are so far only hypotheses. Measurement results available in the literature indicate that spatial sparsity assumptions are questionable at lower frequencies (e.g., [10, Fig. 4]). At millimeter-wave frequencies, however, the channel responses may indeed be sparse [8].

The research efforts on Massive MIMO in recent years have established many of the key characteristics of the technology, but it is still unclear to what extent Massive MIMO can be applied in FDD mode. We encourage researchers to investigate this thoroughly in the coming years, to determine if any of the sparsity hypotheses are indeed true or if there are some other ways to reduce the overhead signaling. Proper answers to these questions require intensive research activities and channel measurements.

Acknowledgment

We would like to thank all of our Massive MIMO research collaborators; in particular, Prof. Fredrik Tufvesson and his colleagues at Lund University, who provided the picture of their LuMaMi testbed. The writing of this article was supported by the EU FP7 under ICT-619086 (MAMMOET), and by ELLIIT and CENIIT.

References

[1] T. L. Marzetta, "Noncooperative Cellular Wireless with Unlimited Numbers of Base Station Antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, 2010, pp. 3590–3600.

[2] J. Vieira *et al.*, "A Flexible 100-Antenna Testbed for Massive MIMO," *Proc. IEEE Globecom Wksp. – Massive MIMO: From Theory to Practice*, 2014.

[3] H. Yin *et al.*, "A Coordinated Approach to Channel Estimation in Large-Scale Multiple-Antenna Systems," *IEEE JSAC*, vol. 31, no. 2, 2013, pp. 264–73.

[4] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for Maximal Spectral Efficiency: How Many Users and Pilots Should Be Allocated?," *IEEE Trans. Wireless Commun.*, to appear, http://arxiv.org/pdf/1412.7102.

[5] H. Ngo, E. Larsson, and T. Marzetta, "Energy and Spectral Efficiency of Very Large Multiuser MIMO Systems," *IEEE Trans. Commun*., vol. 61, no. 4, 2013, pp. 1436–49.

[6] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of Cellular Networks: How Many Antennas Do We Need?," *IEEE JSAC*, vol. 31, no. 2, 2013, pp. 160–71.

[7] A. Alkhateeb *et al*., "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," *IEEE J. Sel.. Topics Signal Processing*, vol. 8, no. 5, 2014, pp. 831–46.

[8] A. Adhikary *et al*., "Joint Spatial Division and Multiplexing for mm-Wave Channels," *IEEE JSAC*, vol. 32, no. 6, 2014, pp. 1239–55.

[9] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of Favorable Propagation in Massive MIMO," *Proc. EUSIPCO*, 2014.

[10] X. Gao *et al*., "Massive MIMO Performance Evaluation Based on Measured Propagation Data," *IEEE Trans. Wireless Commun*., vol. 14, no. 7, 2015, pp. 3899–3911.

[11] M. Karlsson and E. G. Larsson, "On the Operation of Massive MIMO with and Without Transmitter CSI," *Proc. IEEE SPAWC*, 2014.

[12] E. Björnson *et al*., "Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits," *IEEE Trans. Info. Theory*, vol. 60, no. 11, 2014, pp. 7112–39.

[13] U. Gustavsson *et al*., "On the Impact of Hardware Impairments on Massive MIMO," *Proc. IEEE GLOBECOM*, 2014.

[14] H. Yang and T. L. Marzetta, "Total Energy Efficiency of Cellular Large Scale Antenna System Multiple Access Mobile Networks," *Proc. OnlineGreenComm*, 2013.

[15] D. Schneider, "Could Supercomputing Turn to Signal Processors (Again)?" *IEEE Spectrum*, Oct. 2012, pp. 13–14.

Biographies

Emil Björnson (emil.bjornson@liu.se) received a Ph.D. degree in 2011 from KTH Royal Institute of Technology, Sweden. He was a joint postdoctoral researcher at Supélec, France, and at KTH Royal Institute of Technology, Sweden. He has been with Linköping University, Sweden, since 2014, and is currently an associate professor. He is the first author of the textbook *Optimal Resource Allocation in Coordinated Multi-Cell Systems*, and received the 2014 Outstanding Young Researcher Award from IEEE ComSoc EMEA, the 2015 Ingvar Carlsson Award, and best conference paper awards in 2009, 2011, 2014, and 2015.

Erik G. Larsson [F'16] (erik.g.larsson@liu.se) is a professor at Linköping University. He has been Associate Editor for several IEEE journals. He is serving as Chair of the IEEE SPS SPCOM Technical Committee in 2015–2016, and has served as Chair of the Steering Committee for *IEEE Wireless Communications Letters* in 2014–2015, and General Chair of the Asilomar SSC Conference 2015. He received the *IEEE Signal Processing Magazine* Best Column Award twice, in 2012 and 2014, and the IEEE ComSoc Stephen O. Rice Prize in Communications Theory in 2015.

Thomas L. Marzetta [F'03] (tom.marzetta@alcatel-lucent.com) received his Ph.D. in electrical engineering from Massachusetts Institute of Technology in 1978He worked for Schlumberger-Doll Research in petroleum exploration and for Nichols Research Corporation in defense research before joining Bell Labs in 1995, where he served as director of the Communications and Statistical Sciences Department within the former Math Center. He is the originator of Massive MIMO, and co-head of the Bell Labs FutureX Massive MIMO project. He is on the Advisory Board of Massive MIMO for Efficient Transmission (MAMMOET), an EU-sponsored FP7 project, and was Coordinator of the GreenTouch Consortium's Large Scale Antenna Systems Project. For his achievements in Massive MIMO he has received the 2015 IEEE W. R. G. Baker Award, the 2015 IEEE Stephen O. Rice Prize, and the 2014 Thomas Alva Edison Patent Award, among others. He became a Bell Labs Fellow in 2014. In May 2015 he received an Honorary Doctorate from Linköping University.

A Cross-Layer Design for a Software-Defined Millimeter-Wave Mobile Broadband System

Yong Niu, Yong Li, Min Chen, Depeng Jin, and Sheng Chen

Aiming to overcome the challenging problems in mmWave networks, such as interference management, spatial reuse, anti-blockage, QoS guarantee, and load balancing, we architecturally borrow the ideas of heterogeneous cloud radio access networks and software-defined networking to propose a software-defined mmWave mobile broadband system via a cross-layer design approach.

ABSTRACT

Heterogeneous networks, which deploy small cells in the mmWave band underlying the macrocell network, have attracted intense interest from both academia and industry. Different from the communication systems using lower carrier frequencies, mmWave communications have unique features, such as high propagation loss, directional communications, and sensitivity to blockage. Aiming to overcome the challenging problems in mmWave networks, such as interference management, spatial reuse, anti-blockage, QoS guarantee, and load balancing, we architecturally borrow the ideas of heterogeneous cloud radio access networks and software-defined networking to propose a software-defined mmWave mobile broadband system via a cross-layer design approach. In this architecture, a centralized controller is introduced by abstracting the control functions from the network layer to the physical layer. Through quantitative simulations in a realistic indoor scenario, we demonstrate the performance advantages of our system in terms of network throughput and flow throughput. This work is the first cross-layer and software-defined design for mmWave communications, which opens up an opportunity for mmWave communications to make a significant impact on future 5G networks.

INTRODUCTION

With explosive growth of mobile traffic, heterogeneous networks (HetNets) with small cells deployed underlying the conventional homogeneous macrocell network have attracted intense interest from both academia and industry [1]. Since the spatial reuse gain of deploying small cells in the carrier frequencies employed in today's cellular systems is fundamentally limited by interference constraints [2], there is increasing interest in deploying small cells in higher frequency bands, such as the millimeter-wave (mmWave) bands, to significantly boost the network capacity of HetNets. With huge bandwidth, small cells in the mmWave band are able to provide multiple gigabits per second transmission rate for a large number of wireless multimedia services such as uncompressed high definition television (HDTV), high-speed data transfer between devices, wireless gigabit Ethernet, and wireless gaming. Moreover, recent developments in transceiver components design have paved the way to practically and economically utilizing the mmWave band, and several standards on mmWave networks have been or are being defined to achieve multi-gigabit rates, for example, IEEE 802.15.3c [3] and IEEE 802.11ad [4].

There are some fundamental differences between mmWave communications and existing communication systems using lower carrier frequencies (e.g., from 900 MHz to 5 GHz). Due to the huge propagation loss, mmWave communications are range-limited and only suitable for local-range communications [5]. Consequently, the most likely scenarios for deploying mm Wave wireless personal area networks (WPANs) are conference rooms, living rooms, and enterprise cubicles [4]. On one hand, in order to overcome huge attenuation, beamforming (BF) has been adopted as an essential technique, and mmWave links are inherently directional [6]. With directional listening and transmitting, third party nodes cannot hear current transmissions, and carrier sensing is disabled, which is referred to as the famous deafness problem. On the other hand, mmWave links are vulnerable to blockage due to their weak ability to diffract around obstacles such as furniture and human bodies. Blockage by a human body penalizes the link budget by about 20 to 30 dB [5].

Due to the rapid growth of mobile data demands as well as to overcome the limited range of mmWave communications, in a practical mmWave mobile broadband system, the number of access points (APs) deployed over both public and private areas increases tremendously. For example, APs must be deployed densely in scenarios such as enterprise cubicles and conference rooms to provide seamless coverage. With APs densely deployed in indoor environments, interference among neighboring basic service sets (BSSs) cannot be neglected, and should be managed efficiently to maximize concurrent transmissions (spatial reuse) [7]. On the other hand, with multiple APs deployed, multi-AP diversity can also be utilized to overcome the blockage problem [8]. With the small coverage area of each AP, user mobility will cause significant load fluctuations in each BSS. Consequently, handover, user association, and resource allocation have to be managed efficiently at each AP in concert

Yong Niu, Yong Li and Depeng Jin are with Tsinghua University; Min Chen is with HuaZhong University of Science and Technology; Sheng Chen is with University of Southampton.

0163-6804/16/$25.00 © 2016 IEEE

with other neighboring APs in order to achieve the desired goals such as mobility management and load balancing.

Clearly, coordination among APs must be explicitly considered in the design of mmWave mobile broadband systems. In traditional wireless networks, distributed coordination is usually exploited. Distributed coordination among APs, however, does not scale well [9], and the latency will increase significantly with the number of APs, which is unsuitable for mmWave communication systems where a time slot only lasts for a few microseconds. Moreover, with distributed coordination it is difficult to achieve the intelligent control mechanisms required for complex operational environments that involve dynamic behavior of accessing users and temporal variations of the communication links. Thus, the traditional distributed network control mechanism may not be suitable for mmWave mobile broadband systems, and it is important and urgent to design a new paradigm at the architecture and system level for mmWave communication systems.

Heterogeneous cloud radio access networks (H-CRANs), as a new paradigm for improving both spectral efficiency and energy efficiency, suppress inter-tier interference and enhance the cooperative processing capabilities through combination with cloud computing [10, 11]. On the other hand, software-defined networking (SDN) advocates separating the control plane and data plane, and abstracting the control functions of the network into a logically centralized controller [9].

In this article, we architecturally borrow the ideas of H-CRANs and SDN to propose a software-defined mmWave mobile broadband system via a cross-layer design approach. In this architecture, we extend the original concept of SDN control by taking the functions of the physical layer into consideration as well, not just those of the network layer. Specifically, by abstracting the control functions of both layers, a logically centralized programmable control plane oriented toward both the network and physical layers is introduced, through which we achieve the fine-grained control and flexible programmability of the system. The interfaces between the control plane and data plane are also defined to facilitate cross-layer control of the control plane. With the characteristics of mmWave communications considered, we overcome the challenging problems in the system such as interference management, spatial reuse, anti-blockage, QoS guarantee, and load balancing, by centralized and cross-layer control. To the best of our knowledge, we are the first to propose a software-defined mobile broadband system for mmWave communications via a cross-layer design approach. This software-defined and cross-layer design opens up an opportunity for mmWave communications to make a significant impact on fifth generation (5G) networks.

The article is structured as follows. We first present a typical indoor scenario of the mmWave mobile broadband system. After offering the system design goal and principle, we present the proposed cross-layer software defined mmWave mobile broadband system, and analyze its open problems and challenges. Then we show the advantages of the centralized control and cross-layer design through a typical application scenario for anti-blockage. Finally, we analyze how much gain we can achieve by leveraging such a cross-layer software-defined design by quantitative simulations.

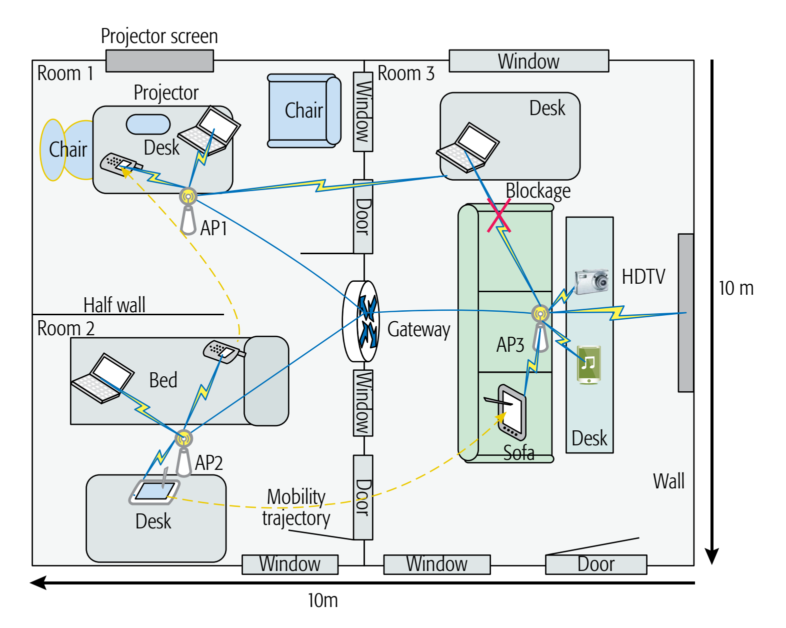


Figure 1. A typical indoor application scenario of the software-defined mmWave mobile broadband system.

System Overview

Typical Indoor Scenarios

As stated in the introduction, the most foreseeable applications for mmWave communications are in the indoor environment. Figure 1 depicts a typical indoor scenario of an mmWave mobile broadband system, where three APs are deployed in the rooms, and they are connected to the gateway via fiber. In this environment, the accessing devices are naturally mobile, the movement trajectories of which are illustrated as yellow dashed lines. In room 3, the line of sight (LOS) path between AP3 and the laptop is blocked by the sofa.

Design Goals

Before presenting the proposed system, we first summarize the design goals for the envisaged software-defined mmWave mobile broadband system.

Interference Management and Spatial Reuse: In mmWave WPANs, directional transmissions enable less interference between links. However, due to the limited communication range, the interference between links cannot be neglected. The interference in an mmWave broadband system can be divided into two portions: interference within each BSS and interference among different BSSs. To enhance network capacity, the interference should be efficiently managed, and concurrent transmissions should be supported among different BSSs as well as within each BSS by global effective interference management and efficient concurrent transmission scheduling.

Robust Network Connectivity: To overcome

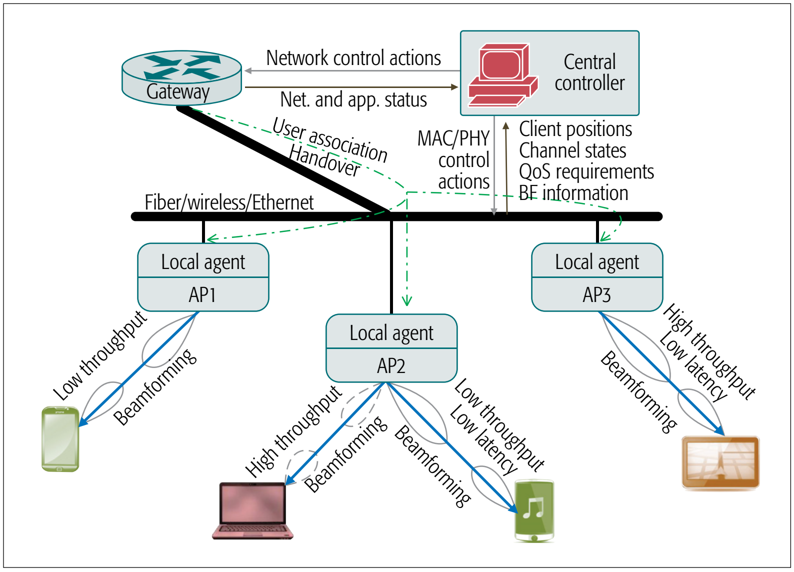


Figure 2. Overview of our proposed architecture for the software-defined mmWave mobile broadband system.

the blockage problem, robust network connectivity should be provided in a cross-layer manner to ensure good user experience. Beam switching to exploit non-LOS (NLOS) transmissions can be applied in the physical layer. Besides, relaying and multihop transmissions in the medium access control (MAC) layer can also be applied. With multiple APs deployed, for devices in the overlapping regions of BSSs, performing handovers in the network layer is also an effective way to overcome blockage through cooperation between APs.

Optimized Load Balance: With a small coverage area, user mobility will cause significant load fluctuations within each BSS. Thus, user association and handovers between APs should be carried out to achieve an optimized load balance through global and intelligent control of all APs in the system.

Flexible QoS Guarantee: The software-defined mmWave broadband system should provide cross-layer QoS guarantees for different kinds of traffic in terms of latency, throughput, and connection reliability. In the physical layer, selection of modulation and coding schemes (MCSs) should be applied to meet different throughput requirements. In the MAC layer, scheduling of flows should be applied to meet the QoS requirement of each flow. In the network layer, handovers between APs should also be performed to ensure the QoS of each flow.

System Architecture

Architecture Overview

In order to achieve the above design goals, we propose a cross-layer software-defined architecture for the mmWave mobile broadband system, as illustrated in Fig. 2. In this architecture, the control functionalities from the physical layer to the network layer are incorporated in a centralized controller. The controller in our architecture encapsulates all the control logic functionality of the network, while the data plane consists of forwarding and wireless communication devices, such as APs and gateway. The controller has two components: the central controller and local agents. The central controller usually resides on the gateway, makes the rules, and controls the behavior of the gateway and APs from a global perspective. Due to the inherent delay between the central controller and each AP [10], there is a local agent residing on each AP to adapt to the rapidly varying network states. The central controller and APs can be connected via wireless, fiber, Ethernet, or any form of backhaul links, which should have short delay to ensure the real-time control of the central controller. To achieve efficient control from the network layer to the physical layer, the measurement interface and control interface between the controller and each data plane device are defined in our architecture. Through the measurement interface, network and application statuses, client positions, channel states, and other state parameters are reported to the controller periodically. On the other hand, through the control interface, control flows in the network layer, MAC layer, and physical layer are resolved and translated to the actions of APs and the gateway.

Interface

The interface between the controller and data plane devices consists of:
- The measurement interface from the data plane devices to the controller
- The control interface from the controller to the data plane devices

Measurement Interface: Through the measurement interface, the controller obtains the local and global network states and application information from each AP and the gateway. Typical state parameters include client positions, channel states of links, BF information of clients, QoS requirements of flows, the number of clients under each AP, and so on. The data plane devices periodically report the state parameters to the controller, which then dynamically updates its local and global network views.

Control Interface: The control interface for the data plane devices resolves the control flows from the controller and translates the control decisions to the actions of each AP or the gateway in the network layer, MAC layer, and physical layer. The control interface adopts the "match-action" strategy, and the control strategies for the network layer and physical layer are quite different.

Network Layer Interface: Handovers, user association, and resource allocation are performed through the control interface in the network layer. For example, forwarding packets from the gateway to the APs is controlled through this interface. Concurrent transmission schedules are also pushed to each AP through this control interface. The control interface operates on a table indexed by "Flow ID". Flow ID is identified based on the function fields in the packet header, such as IP address and MAC address. Specifically, when the gateway receives a packet, it first checks whether this flow matches its control rules. If so, the gateway will execute the corresponding actions. For example, **if** dest IP = xx.xx.xx.xx, **then** forward to AP1.

Physical Layer Interface: The selection of

MCSs according to the channel conditions and traffic patterns, transmission power control, and BF between paired devices are completed through the control interface in the physical layer. Although the physical layer control interface also operates on a table, the match fields, rules, and actions are quite different from those of the network layer control interface. Specifically, when an AP transmits a packet to a client, it first checks whether this flow matches its control rules. If so, the AP executes the corresponding actions. For example, **if** slot = xx and dest IP = xx.xx.xx.xx, **then** direct the beam toward client 2 and transmit at 2 Gb/s.

Controller

Centralized Controller: The central controller controls the data plane devices from the global perspective based on the up-to-date network states obtained via the measurement interface. Given the network states, the controller maintains and updates a global network state database, to which we refer as the "mmWave information center" (MIC). The MIC consists of the following elements.

- **Interference Graph:** A weighted and directional graph where each vertex represents one link, and the weight of each arc is the interference level between the two links. The interference level may be the interference power or other parameters to indicate the interference strength, such as the distance between the transmitter of one link and the receiver of another link, and a binary variable to indicate whether the transmitter of one link is inside the exclusive region (ER) of the receiver of another link [12]. The interference graph can be obtained according to the network state parameters from the measurement interface, such as client positions, BF information of clients, and other physical layer parameters, for example, transmission power, path loss exponent, and the cross-correlation between two links.
- **QoS Graph:** A weighted and directional graph where each vertex represents a client or an AP in the network. The weight of each arc is the QoS requirements of each flow, such as throughput, latency, and connection reliability. The QoS graph can be obtained directly through the measurement interface.
- **Link Quality Graph:** A weighted and directional graph where each vertex represents a client or an AP in the network. The weight of each arc indicates the link quality, which may be the received signal strength at the receiver, the transmission rate that the link can support, or the frequency of the link outage. The link quality graph can be obtained from the transmission rate measurements of the links directly, or inferred from the network state parameters such as client positions and the number of link outages.
- **Flow Statistics:** These are the statistics of the ongoing flows, for example, the number of transmitted packets, the number of queued packets, and the number of flows under each AP.

Based on the MIC, the central controller can achieve effective and efficient radio resource allocation. For example, based on the interference graph, QoS graph, and link quality graph, efficient concurrent scheduling algorithms can be implemented in the central controller to maximize spatial reuse, while effective interference management can be achieved based on the interference graph. Based on the link quality graph and the flow statistics, smoother handovers, and reduced dropped connections and ping-pong can be accomplished. The central controller can also achieve efficient load balancing based on the link quality graph and the flow statistics.

With more APs and users in the system, the workload of the central controller is getting heavy. To address this issue, the central controller in our system can be implemented via a vertical approach with multiple controllers incorporated [14]. The central controller consists of multiple local controllers and one root controller. Each local controller is responsible for managing some APs and users, and the root controller is responsible for coordinating multiple local controllers for global management and control. In this way, more APs and users can be managed by the central controller, and more complete and optimized control functions can be encapsulated in the central controller.

Local Agent: To alleviate the inherent delay from the central controller to each AP, a local agent is deployed at each AP to adapt rapidly to the varying channel conditions and traffic patterns. Since the central controller makes control decisions based on the delayed state information compared to the local agent, the control decisions from the local agent are more timely. When the control decisions from the central controller conflict with those from the local agent, the control decisions from the local agent will be adopted by the control interface. Due to lower complexity of the local agent, the local agent is mainly in charge of delay-sensitive control functions. For example, beam switching to exploit NLOS transmissions in case of sudden blockage is handled at the local agent. The local agent also selects the appropriate MCSs according to the fast varying channel conditions.

Control Overhead

The control overhead in the system mainly consists of three parts: the network state information measurement through the measurement interface, the control decision computation of the control plane, and the control flow forwarding from the control plane to the data plane.

Initially, complete network state information measurements are performed to establish the MIC. With relatively low user mobility, the network states do not change all the time, and remain unchanged for a period. Thus, each kind of measurement is performed periodically to track the variations in the network state. The network state measurements within each BSS are managed by each AP, and the measurements among BSSs are performed via the cooperation of APs under the control of the central controller. Then the network state information is sent to the central controller from the APs via the backhaul links. With the gigabit-per-second transmission rate, this process can be completed quickly. At the same time, to limit the complexity and avoid excessive overhead, some state information cannot be obtained in practice. In this case, some compensation information can be obtained

The control overhead in the system mainly consists of three parts: the network state information measurement through the measurement interface; the control decision computation of the control plane; and the control flow forwarding from the control plane to the data plane.

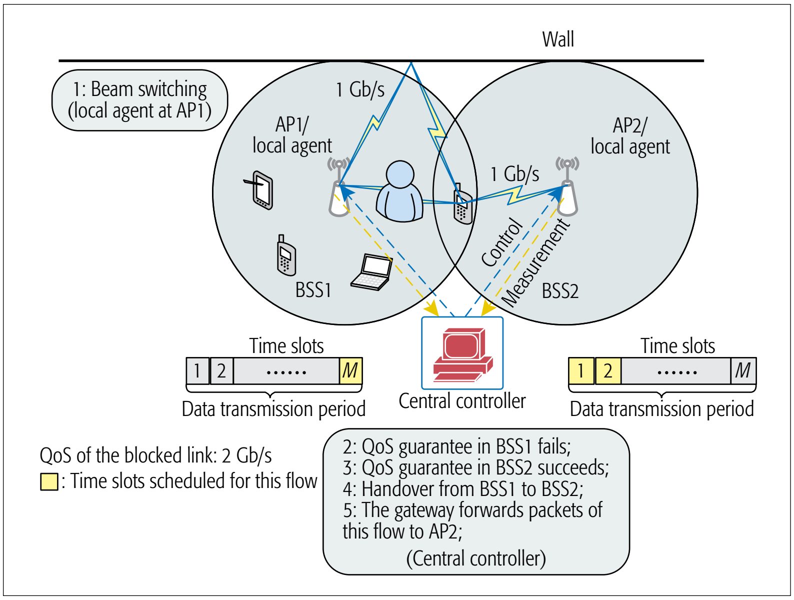

Wall
1: Beam switching (local agent at AP1)
1 Gb/s
AP1/ local agent
AP2/ local agent
1 Gb/s
Control
Measurement
BSS1
BSS2
Time slots
1 2 ...... M
Data transmission period
Central controller
Time slots
1 2 ...... M
Data transmission period
QoS of the blocked link: 2 Gb/s
□: Time slots scheduled for this flow
2: QoS guarantee in BSS1 fails;
3: QoS guarantee in BSS2 succeeds;
4: Handover from BSS1 to BSS2;
5: The gateway forwards packets of this flow to AP2;
(Central controller)

Figure 3. An example of overcoming blockage in the software-defined mmWave mobile broadband system.

instead to estimate the required information. The overhead in the control decision computation of the control plane depends on the efficiency of the algorithms employed in the central controller. Therefore, efficient and effective control algorithms are needed to reduce the control overhead and also achieve good performance. After the control decisions are obtained, the control flows are forwarded from the central controller to the APs via the backhaul links. Then the control flows are distributed by the APs to users within the BSSs. With the gigabit-per-second transmission rate, this process can also be completed in a short time.

Open Problems and Challenges

Measurements

In order for the central controller to obtain an accurate and comprehensive global view of the network, efficient measurement mechanisms are needed. There is already some work addressing this critical issue. For example, a bootstrapping scheme can be executed to obtain up-to-date network topology and node location information [14], while a BF information table that records all the beamforming training results among clients can be established at the AP [15]. However, most of the current work focuses on the network state measurement within a BSS. For clients in the overlapped region of two BSSs, the link quality information and BF information between the clients and the neighboring APs are also required for tasks such as handover and interference management. Since concurrent transmissions should be enabled among different BSSs as well as within each BSS, the interference caused by concurrent transmissions, especially among different BSSs, must be estimated as accurate as possible. On the other hand, with the dynamics due to user mobility considered, the measurement algorithms should be able to track the variations in the network states in as little time as possible to reduce overhead. Therefore, efficient measurement algorithms are open problems that need to be extensively investigated to facilitate the deployment of such an mmWave mobile broadband system.

Algorithms Employed in the Centralized Controller

To achieve the design goals, effective and efficient algorithms for transmission scheduling, load balancing, BF, anti-blockage, and power control are needed. Although there are many works on MAC protocols or scheduling algorithms for mmWave WPANs, most works focus on the scenario of a BSS [7] and do not consider the interference among different BSSs. Also, there are several approaches to overcome blockage, such as beam switching from an LOS path to an NLOS path, relaying [5], and multi-AP diversity [8]. In [8], multiple APs are deployed, and when the link between a client and an AP is blocked, another AP is selected to complete the remaining transmission task. For transmissions between clients, beam switching to an NLOS path is usually a good choice. In some cases, however, the NLOS path is difficult to find, or for high-rate applications, the transmission rate supported by the NLOS path cannot meet the throughput requirement. In this case, relaying is an effective way to overcome blockage and even to improve the throughput [5]. Therefore, every approach has its advantages and shortcomings, and is only efficient in certain circumstances. With global and cross-layer control over the data plane devices, how to combine these approaches and apply them appropriately to ensure robust network connectivity and improve network performance remains an open problem that warrants further investigation.

Dealing with Erroneous State Information

Due to the complexity of the indoor environment and the inherent delay between the central controller and APs, there may be deviation and error in the obtained network state information. In this case, control decisions based on erroneous state information also have deviation and error. Therefore, for some error-sensitive decisions, there should be mechanisms in the centralized algorithms for the central controller to correct errors in the control decisions. After a control decision is made based on some state information, the related network states should be monitored. When the network states do not match the control decision, the central controller should adjust the control decision accordingly, and also re-measure the previous network states the control decision on which was based. Besides, the algorithms in the central controller should be tolerant of the deviation in the state information to some extent to avoid significant degradation in performance.

Application Scenarios

In this section, we present an example where beam switching and handovers are combined to overcome blockage in our proposed software-defined mmWave mobile broadband system.

As illustrated in Fig. 3, a blockage occurs suddenly between AP1 and a mobile phone. In BSS1, this downlink flow is scheduled to transmit in the Mth time slot of the data transmission period. First, to ensure continuous connection, the local agent will command AP1 to switch its antenna

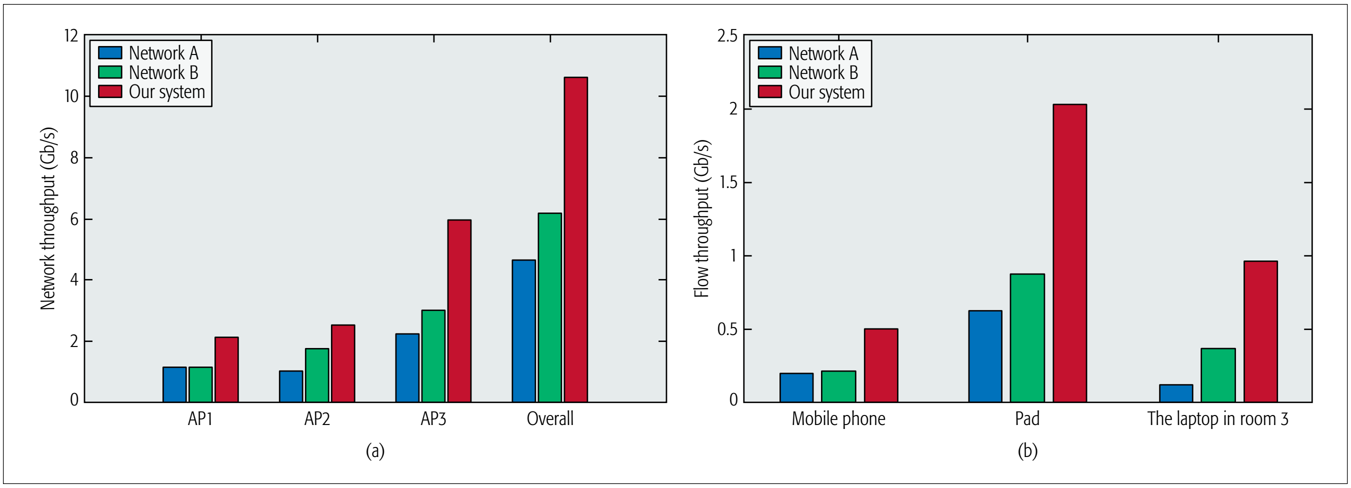


Figure 4. The network throughput and flow throughput comparison between our architecture and traditional networks: a) network throughput comparison; b) flow throughput comparison.

toward the wall to exploit the NLOS path for transmission temporarily without severe interference to current transmissions. At the same time, AP1 reports the blockage to the central controller through the measurement interface. The NLOS path has additional path losses and can only support a throughput of 1 Gb/s, while the QoS requirement (throughput) of this flow is 2 Gb/s. After the blockage has been reported to the central controller, to meet the QoS requirement of this flow, the central controller executes two actions. The first action is to check whether it is possible to schedule more time slots to this flow to meet its QoS requirement. Due to the heavy loads in BSS1, there is no additional time slot for this flow. Then the central controller executes the second action, which is to check whether there is an LOS path between the client and a neighboring AP. The central controller finds that there is one LOS path between the mobile phone and AP2, and the path can also support a throughput of 1 Gb/s if one time slot is scheduled for this flow. At the same time, the central controller finds that currently there is no client under AP2, and there are enough time slots in BSS2 to meet the QoS requirement of this flow. Thus, the central controller issues the instructions to AP1 and AP2 to perform the handover of the mobile phone through the control interface. After the handover, the packets of this flow are forwarded from the gateway to AP2, and in BSS2, time slots 1 and 2 are scheduled to this flow.

Performance Evaluation

We adopt a realistic network deployment to quantitatively evaluate the enhancement achieved by utilizing the software-defined cross-layer architecture for the mmWave mobile broadband system.

Targeted System

In the simulation, we target the scenario shown in Fig. 1. We adopt the superframe structure in IEEE 802.15.3c, where a superframe consists of the beacon period (BP), the contention access period (CAP), and the channel time allocation period (CTAP). There are at most 1000 time slots in CTAP, and time-division multiple access (TDMA) is adopted in CTAP. The duration of BP, CAP, CTAP, and each time slot is the same as that in [7]. A superframe is scheduled at each AP every 0.02 s, and the simulation time is 20 s. The network state information from each AP and the gateway is fed back to the control plane through the measurement interface during the intervals between superframes. The control flows from the control plane are also distributed to each AP and the gateway through the control interface during the intervals between superframes.

In the simulation, we compare our software-defined system with two traditional networks, denoted as *network A* and *network B*. In our system, an anti-blockage mechanism that combines beam switching and handovers between APs intelligently is applied. Besides, the selection of MCSs is achieved at both the local agent and the central controller according to the channel conditions, and the interference among BSSs is managed at the central controller. In contrast, no mechanism is adopted by *network A* to overcome blockage, while beam switching to exploit the NLOS path is applied by *network B*.

Results Analysis

We first compare the network throughput achieved by the two traditional networks to that obtained by our system. The throughputs attained by AP1, AP2, and AP3 as well as the overall network throughput are shown in Fig. 4a. We observe that our system increases the throughputs of AP1, AP2, and AP3 as well as the overall network significantly, compared to the two traditional networks. Our system improves the overall network throughput by about 128 percent compared to *network A*. Due to different traffic loads in different rooms, the increases of throughput achieved in different rooms are different. The increase achieved by AP1 is mainly due to the fact that in our network, when the blockage occurs in room 3, the local agent first exploits beam switching quickly to maintain connection, and then the central controller issues instructions to AP3 and AP1 to perform the handover of the laptop to achieve a high transmission rate. In the traditional networks, however, such beneficial handover

Motivated by the ideas of H-CRAN and SDN, we have proposed a software-defined mmWave mobile broadband system, which achieves intelligent and global control of the mobile network from the physical to network layer by a centralized controller.

will not happen due to the lack of cooperation between APs.

To evaluate the throughput enhancement for each flow achieved by our system, we also present the throughputs of three downlink flows to the mobile phone, the pad, and the laptop in room 3 in Fig. 4b. We can observe that our system improves the flow throughputs of devices significantly. Specifically, our system increases the throughput of the flow to the laptop in room 3 by about 162 percent compared to *network B*, and about 670 percent compared to *network A*, which is mainly due to the combined action of beam switching and the smooth handover between AP3 and AP1 to overcome blockage. Besides, the channel transmission rates in our system can be adapted to the varying channel conditions by the selection of MCSs more quickly compared to the traditional networks.

Conclusions

Motivated by the ideas of H-CRANs and SDN, we have proposed a software-defined mmWave mobile broadband system, which achieves intelligent and global control of the mobile network from the physical to the network layer by a centralized controller. We have also discussed its open problems and challenges, and quantitatively evaluated its performance advantages by simulations targeting a realistic system. This study thus provides a novel design for mmWave communications, and it opens up a new research direction for mmWave communications to make a significant impact on 5G mobile broadband.

Acknowledgment

This work is supported by the National Basic Research Program of China (973 Program) (No. 2013CB329001) and the National Nature Science Foundation of China (No. 61301080, No. 91338203, No. 91338102, No. 61572220, and No. 61321061).

References

[1] J. Lee *et al.*, *Physical Commun.*, Special Issue on Heterogeneous and Small Cell Networks, 2014.

[2] K. Zheng *et al.*, "Energy-Efficient Wireless In-Home: The Need for Interference-Controlled Femtocells," *IEEE Wireless Commun.*, vol. 18, no. 6, 2011, pp. 36–44.

[3] IEEE 802.15 WPAN Millimeter Wave Alternative PHY Task Group 3c (TG3c); http://www.ieee802. org/15/pub/TG3c.html.

[4] IEEE P802.11ad/D9.0, "Draft 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 4: Enhancements for Very High Throughput in the 60 Ghz Band," Oct. 2012.

[5] S. Singh *et al.*, "Blockage and Directivity in 60 GHz Wireless Personal Area Networks: From Cross-Layer Model To Multihop MAC Design," *IEEE JSAC*, vol. 27, no. 8, Oct. 2009, pp. 1400–13.

[6] J. Wang *et al.*, "Beam Codebook Based Beamforming Protocol for Multi-Gb/S Millimeter-Wave WPAN Systems," *IEEE JSAC*, vol. 27, no. 8, Oct. 2009, pp. 1390–99.

[7] J. Qiao *et al.*, "STDMA-Based Scheduling Algorithm for Concurrent Transmissions in Directional Millimeter Wave Networks," *Proc. ICC 2012*, Ottawa, Canada, June 10–15, 2012, pp. 5221–25.

[8] X. Zhang *et al.*, "Improving Network Throughput in 60GHz WLANs via Multi-AP Diversity," *Proc. ICC 2012*, Ottawa, Canada, June 10–15, 2012, pp. 4803–07.

[9] A. Gudipati *et al.*, "SoftRAN: Software Defined Radio Access Network," *Proc. ACM HotSDN 2013*, Hong Kong, China, Aug. 12–16, 2013, pp. 25–30.

[10] K. Zheng *et al.*, "An SMDP-Based Resource Allocation in Vehicular Cloud Computing Systems," *IEEE Trans. Industrial Electronics*, vol. 62, no. 12, 2015, pp. 7920–28.

[11] Z. Ding and H. V. Poor, "The Use of Spatially Random Base Stations in Cloud Radio Access Networks," *IEEE Signal Processing Lett.*, vol. 20, no. 11, Nov. 2013, pp. 1138–41.

[12] L. X. Cai et al., "REX: A Randomized Exclusive Region Based Scheduling Scheme for mmWave WPANs with Directional Antenna," *IEEE Trans. Wireless Commun.*, vol. 9, no. 1, Jan. 2010, pp. 113–21.

[13] R. Ahmed and R. Boutaba, "Design Considerations for Managing Wide Area Software Defined Networks," *IEEE Commun. Mag.*, vol. 52, no. 7, July 2014, pp. 116–23.

[14] J. Ning *et al.*, "Directional Neighbor Discovery in 60 GHz Indoor Wireless Networks," *Proc. ACM MSWiM '09*, Tenerife, Spain, Oct. 26–30, 2009, pp. 365–73.

[15] Q. Chen *et al.*, "Spatial Reuse Strategy in mmWave WPANs with Directional Antennas," *Proc. 2012 IEEE GLOBECOM*, Anaheim, CA, Dec. 3–7, 2012, pp. 5392–97.

Biographies

Yong Niu received a B.E. degree from Beijing Jiaotong University, China, in 2011. He is currently working toward his Ph.D. degree at the Department of Electronic Engineering, Tsinghua University, China. His research interests include millimeter-wave WPANs, medium access control, and software-defined networks.

Yong Li [M'09] received B.S. and Ph.D degrees from Huazhong University of Science and Technology (HUST) and Tsinghua University in 2007 and 2012, respectively. During 2012 and 2013, he was a visiting research associate at Telekom Innovation Laboratories and Hong Kong University of Science and Technology, respectively. During 2013 to 2014, he was a visiting scientist at the University of Miami. He is currently a faculty member in the Department of Electronic Engineering, Tsinghua University. His research interests are in the areas of mobile computing and social networks, urban computing and vehicular networks, and network science and future Internet. He has served as General Chair, Technical Program Committee (TPC) Chair, and TPC member for several international workshops and conferences. He is currently Associate Editor of the *Journal of Communications and Networking* and the *EURASIP Journal of Wireless Communications and Networking*.

Min Chen [SM'09] is a professor in the School of Computer Science and Technology at HUST. He was an assistant professor in the School of Computer Science and Engineering at Singapore National University (SNU) from September 2009 to February 2012. He worked as a post-doctoral fellow in the Department of Electrical and Computer Engineering at the University of British Columbia (UBC) for three years. Before joining UBC, he was a postdoctoral fellow at SNU for one and a half years. His research focuses on the Internet of Things, machine-to machine communications, body area networks, body sensor networks, e-healthcare, mobile cloud computing, cloud-assisted mobile computing, ubiquitous networks and services, mobile agents, multimedia transmission over wireless networks, and so on.

Depeng Jin received B.S. and Ph.D. degrees from Tsinghua University in 1995 and 1999, respectively, both in electronics engineering. He is an associate professor at Tsinghua University and vice chair of the Department of Electronic Engineering. He was awarded the National Scientific and Technological Innovation Prize (Second Class) in 2002. His research fields include telecommunications, high-speed networks, ASIC design, and future Internet architecture.

Sheng Chen (M'1990, SM'1997, F'2008) obtained his B.Eng. degree from the East China Petroleum Institute, Dongying, China, in January 1982, and his Ph.D. degree from City University London in September 1986, both in control engineering. In 2005, he was awarded a higher doctorate degree, his D.Sc., from the University of Southampton, United Kingdom. From 1986 to 1999, he held research and academic appointments at the Universities of Sheffield, Edinburgh, and Portsmouth, all in the United Kingdom. Since 1999, he has been with Electronics and Computer Science, University of Southampton, where he currently holds the post of professor in intelligent systems and signal processing. He is a Distinguished Adjunct Professor at King Abdulaziz University, Jeddah, Saudi Arabia. He is a Chartered Engineer and a Fellow of IET. His recent research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, and evolutionary computation methods and optimization. He has published over 470 research papers. He is an ISI highly cited researcher in the engineering category (March 2004).

Mobile Crowd Sensing and Computing: When Participatory Sensing Meets Participatory Social Media

Bin Guo, Chao Chen, Daqing Zhang, Zhiwen Yu, and Alvin Chin

ABSTRACT

With the development of mobile sensing and mobile social networking techniques, mobile crowd sensing and computing (MCSC), which leverages heterogeneous crowdsourced data for large-scale sensing, has become a leading paradigm. Built on top of the participatory sensing vision, MCSC has two characteristic features: it leverages heterogeneous crowdsourced data from two data sources: participatory sensing and participatory social media; and it presents the fusion of human and machine intelligence in both the sensing and computing processes. This article characterizes the unique features and challenges of MCSC. We further present early efforts on MCSC to demonstrate the benefits of aggregating heterogeneous crowdsourced data.

INTRODUCTION

The effective use of the incredible and continuous production of data coming from different sources (e.g., enterprises, the Internet of Things, online systems) will transform our life and work. Within this context, people are not only data consumers, but participate in different ways (e.g., smartphone sensing, online posting) in the data production process. In this article, we discuss the opportunities that heterogeneous human participation offer to systems and services that rely on large-scale sensing.

It is essential to first clarify the motivation of taking the human in the loop for large-scale sensing. In the past few years, researchers have studied the benefits of understanding urban/community dynamics [1]. However, traditional stationary wireless sensor network deployments often fail to capture such dynamics because they either do not have enough sensing capabilities or are limited in terms of scalability (e.g., high deployment and maintenance cost). Mobile crowd sensing and computing (MCSC) offers a new method of large-scale sensing and computing. On one hand, the sheer number of mobile devices (e.g., smartphones, tablets, wearable devices) and their inherent mobility provide the ability to sense and infer people's context (e.g., ambient noise) in an unprecedented manner. On the other hand, highly scalable sensing with mobile devices in combination with cloud computing support gives MCSC systems the scalability and versatility properties that are often lacking in static deployments. Although it is quite difficult to attempt a formal definition of the MCSC paradigm, we could state that MCSC is *a new sensing paradigm that empowers ordinary people to contribute data sensed or generated from their mobile devices, and aggregates and processes heterogeneous crowdsourced data in the cloud for intelligent service provision.*

From the artificial intelligence (AI) perspective, MCSC is founded on a distributed problem solving model where crowds are engaged in complex problem solving procedures through open calls. The concept of crowd-powered problem solving has been explored in several research areas. The term "crowdsourcing" was coined in 2005 by *Wired*. The definition of the term crowdsourcing is as follows:[1] *the practice of obtaining needed services or content by soliciting contributions from a large group of people, and especially from an online community.* Wikipedia,[2] where thousands of contributors from across the globe have collectively created the world's largest encyclopedia, is a typical example. MCSC extends this concept by going beyond the boundaries of online communities and reaching out to the mobile device user population for sensing participation. With participatory sensing, first proposed by Burke *et al.* [2], we see for the first time solutions that require explicit human involvement in accomplishing sensing tasks. MCSC broadens the concept of participatory sensing from two aspects. First, it takes advantage of various forms of human participation in the mobile Internet era. Generally speaking, MCSC sensing modalities can be obtained from specific hardware sensors (e.g., accelerometers, cameras) available on mobile devices and from the information trail (e.g., social media posts) directly generated by users. Second, MCSC presents the fusion of human and machine intelligence in both the sensing and computing processes. The usage of heterogeneous crowdsourced data as well as the integration of human and machine intelligence

With the development of mobile sensing and mobile social networking techniques, mobile crowd sensing and computing, which leverages heterogeneous crowdsourced data for large-scale sensing, has become a leading paradigm. The authors characterize the unique features and challenges of MCSC, and also present early efforts on MCSC to demonstrate the benefits of aggregating heterogeneous crowdsourced data.

Bin Guo and Zhiwen Yu are with Northwestern Polytechnical University; Chao Chen is with Chongqing University; Daqing Zhang is with Institut TELECOM SudParis; Alvin Chin is with Microsoft.

[1] Source: Merriam-Webster Dictionary, http://www.merriam-webster.com/dictionary/crowdsourcing

[2] http://en.wikipedia.org/

0163-6804/16/$25.00 © 2016 IEEE

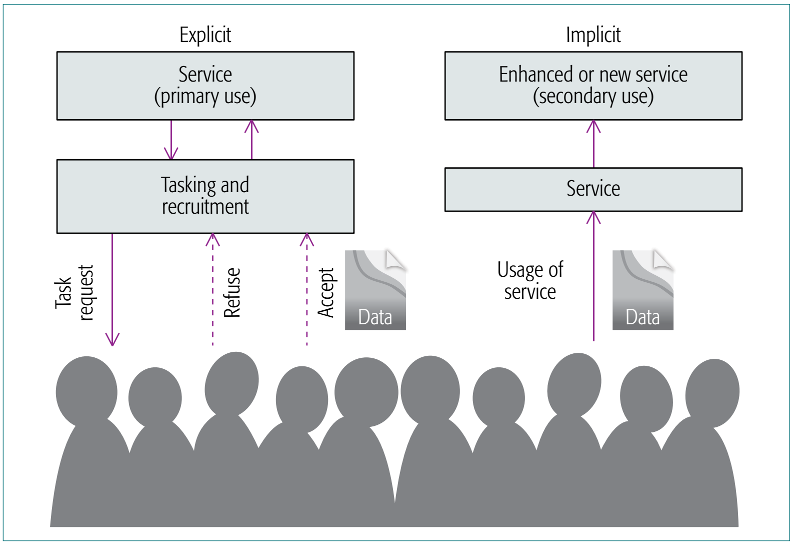


Figure 1. Data control in MCSC.

opens up new and unexpected opportunities. We use the following trip planning scenario to showcase the characteristics of MCSC.

Trip planning is a typical MCSC application. With participatory sensing, we can collect GPS trajectory data from vehicles and compute the optimal route when answering a query with departure and destination points. However, for a more complex query, that is, to generate an itinerary for a visitor to a city given the time budget (start time, end time), it is not possible to leverage a single trajectory dataset. Further information such as points of interest (POIs) in the city, the best time to visit the POIs, and user preferences for different POIs, are further needed. The information can be obtained, however, by reusing the user-generated data from location-based social networks (LBSNs).

The above scenario demonstrates the aggregated power of participatory sensing and social networks for intelligent service provision. The key contributions of this article can be summarized as follows:
- Characterizing the main features of MCSC by combining participatory sensing and participatory social media
- Exploring the fusion of human and machine intelligence in MCSC, and discussing the key research challenges such as cross-space data mining and data quality maintenance
- Presenting several representative studies to demonstrate the power and usage of MCSC, including two of our recent works and ones from other research groups

Merging Participatory Sensing and Participatory Social Media

MCSC combines two distinct data generation modes: participatory sensing advocates the involvement of citizens in the sensing loop in the real world; participatory social media refers to user-generated data in mobile social networks (MSNs). MSN services can bridge the gap between online interaction and physical elements (e.g., check-in places, activities), and the data collected from them provides another way to understand urban dynamics.

User Participation: Explicit or Implicit

Previous studies often discuss user-participating data sourcing in one dimension: the degree of user involvement in the sensing process. As presented by Ganti *et al.* [3], crowd-powered sensing can span a wide spectrum in terms of user involvement, with participatory and opportunistic sensing at the two ends. The proportion of human involvement depends on application requirements and device capabilities. With the two data generation modes in MCSC, we intend to categorize the sensing style from a new dimension: the data usage manner. For both participatory and opportunistic sensing, data collection is the primary purpose of the application. The sensing task is therefore explicit to its participants. For user-generated data in MSNs, however, the data is used for a second purpose (the primary purpose is social interaction) and is often implicit to the contributors. We thus categorize the sensing style of MCSC into two basic types regarding data usage: explicit and implicit. Note that sometimes the two basic sensing styles can be merged into a complex one. For instance, explicit crowdsensed data can also be used for a second purpose (e.g., using crowdsensed urban noise data to predict the functions of regions in a city), thus integrating explicit sensing with implicit sensing.

Data Control for Primary/Secondary Use

The sharing of personal data in MCSC applications can raise significant privacy concerns, with information (e.g., locations, preferences) being sensitive and vulnerable to privacy attacks. To motivate user participation, we should explore new techniques that protect user privacy. We identify two basic sensing styles (*explicit* and *implicit*) for MCSC, and they have distinct data control needs.

In *explicit sensing*, sensor data collection is triggered by tasks, which specify the sensing modalities (e.g., regions of interest, sampling context) based on application requests. The tasks are distributed to mobile device carriers that satisfy the tasking requirements, and people can decide to accept or refuse the task allocated (as shown in Fig. 1). We can find that data is collected under the "primary use" manner in explicit sensing. Privacy in explicit sensing should guarantee that participants maintain control over the release of their sensitive information, for example, the degree of granularity and data recipients.

In *implicit sensing*, data is contributed not for a sensing task, but for users to enjoy online services (socializing on Facebook, purchasing goods on Amazon). As shown in Fig. 1, the data is reused to enhance original services or create new services by third parties (i.e., used for a second purpose). Since most innovative secondary uses are not imagined when the data is collected, how should we protect user privacy in implicit sensing?

"Terms and conditions" have an important role in current online services, where people are told at the time of service usage which information is being gathered and for what purpose. This law works for explicit sensing, but, as discussed above, may not work well for implicit sensing. For example, it is often difficult to pre-specify a (secondary) purpose. Improved data notice strategies should thus be studied. One basic rule to follow is that we should have users keep a sense of awareness (and control) of any use of their data. We may need to build an *evolving notice center* (ENC) for each service to make users aware of what data is collected, how long the data will be kept, and who are using their data. Enhanced legal protections for service providers on data reuse can also alleviate user concerns in implicit sensing.

Fusion of Human and Machine Intelligence

The primary feature of MCSC is having varying human participation (e.g., locating sensing objects, capturing pictures, posting in MSNs) in the large-scale problem solving process. The coexistence of human and machine power, however, needs to be orchestrated in an optimal manner to enhance them both. An important reason to combine human and machine intelligence is that they often show distinct strengths and weaknesses, as illustrated in Fig. 2. We refer to the fusion of human intelligence (HI) and machine intelligence (MI) as HMI, which *characterizes the complementary roles of HI and MI in problem solving and integrates them for MCSC service provision.*

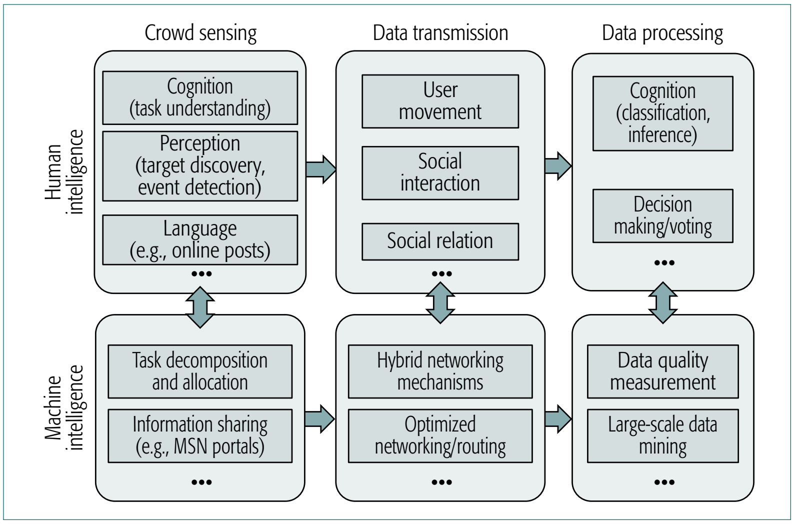
As shown in Fig. 2, there are three important layers in a generic MCSC framework, and HI and MI work collaboratively over all these layers. For example, in the *crowd sensing* layer, machines can allocate tasks to proper participants according to the task needs and human behavior patterns, and the selected workers can execute the assigned tasks using their cognition/perception abilities. In the *data transmission* layer, human mobility patterns and social interactions facilitate the development of optimized networking methods [4]. In the *data processing* layer, the integrated power of HMI can attain higher performance (e.g., accuracy of classification) than either one.

Key Research Challenges

The combination of two participatory data generation modes in MCSC also raises new research challenges and issues, some of which are discussed below.

Heterogeneous, Cross-Space Data Mining

The strength of MCSC relies on the usage of crowdsourced data from both physical and virtual societies. The same sensing object (e.g., a social gathering on a street corner) will interact with both spaces and leave fragmented data in each space, making the information obtained from different communities (online or offline) different. For instance, we can learn social relationships from online social networks, and infer group activities using smartphone sensing in the real world. Obviously, the complementary nature of heterogeneous communities will bring new opportunities to develop new human-centric services. Therefore, we should integrate and fuse the information from heterogeneous, cross-space data sources — we refer to it as *cross-space data mining* — to attain a comprehensive picture of the sensing object. Potential research issues include how to reveal the complex linkage and interplay among the data from online/offline space, how to aggregate the information from heterogeneous data sources for enhanced learning, and so on. Recent progress on this has been discussed in detail in our previous work [5].



Crowd sensing
Data transmission
Data processing
Human intelligence
Machine intelligence
Cognition (task understanding)
Perception (target discovery, event detection)
Language (e.g., online posts)
•••
User movement
Social interaction
Social relation
•••
Cognition (classification, inference)
Decision making/voting
•••
Task decomposition and allocation
Information sharing (e.g., MSN portals)
•••
Hybrid networking mechanisms
Optimized networking/routing
•••
Data quality measurement
Large-scale data mining
•••

Figure 2. The fusion of human and machine intelligence in MCSC.

Potential Fusion Patterns of HMI

The success of MCSC largely depends on HMI. However, the challenge is how HI and MI should be fused to produce aggregated effects. In other words, we should study the potential fusion patterns of HI and MI to attain HMI. Three potential patterns are identified in this article: *sequential*, *parallel*, and *hybrid*. The *sequential* pattern is often used in MCSC. For example, given a crowd sensing task, machines can decompose it into sub-tasks, and people can execute the assigned sub-tasks using their cognition abilities. HI and MI can also be combined in a *parallel* manner. Still taking the accomplishment of a complex sensing task as our example, human nodes and machines (e.g., static sensing nodes) may have complementary sensing abilities, and need to work in a parallel way to fully capture the required information. Finally, two or more parallel or sequential units can be integrated in a *hybrid* manner when more complex problems are to be solved.

Data Quality and Selection

The involvement of human participation in the sensing process also brings forth certain uncertainties to MCSC systems. For example, anonymous participants may send incorrect, low-quality, or even fake data to a data center [6]. Data contributed by different people may be redundant or inconsistent. Certain quality estimation and prediction methods are thus necessary to evaluate the quality of sensing data, and statistical processing can be used to identify outliers. Data selection is also crucial to filter low-quality or irrelevant data and generate a high-quality dataset for further data processing or information presentation. For instance, Ding *et al.* [6] proposed a data-cleansing-based robust spectrum sensing algorithm to eliminate the negative impact of abnormal or low-quality

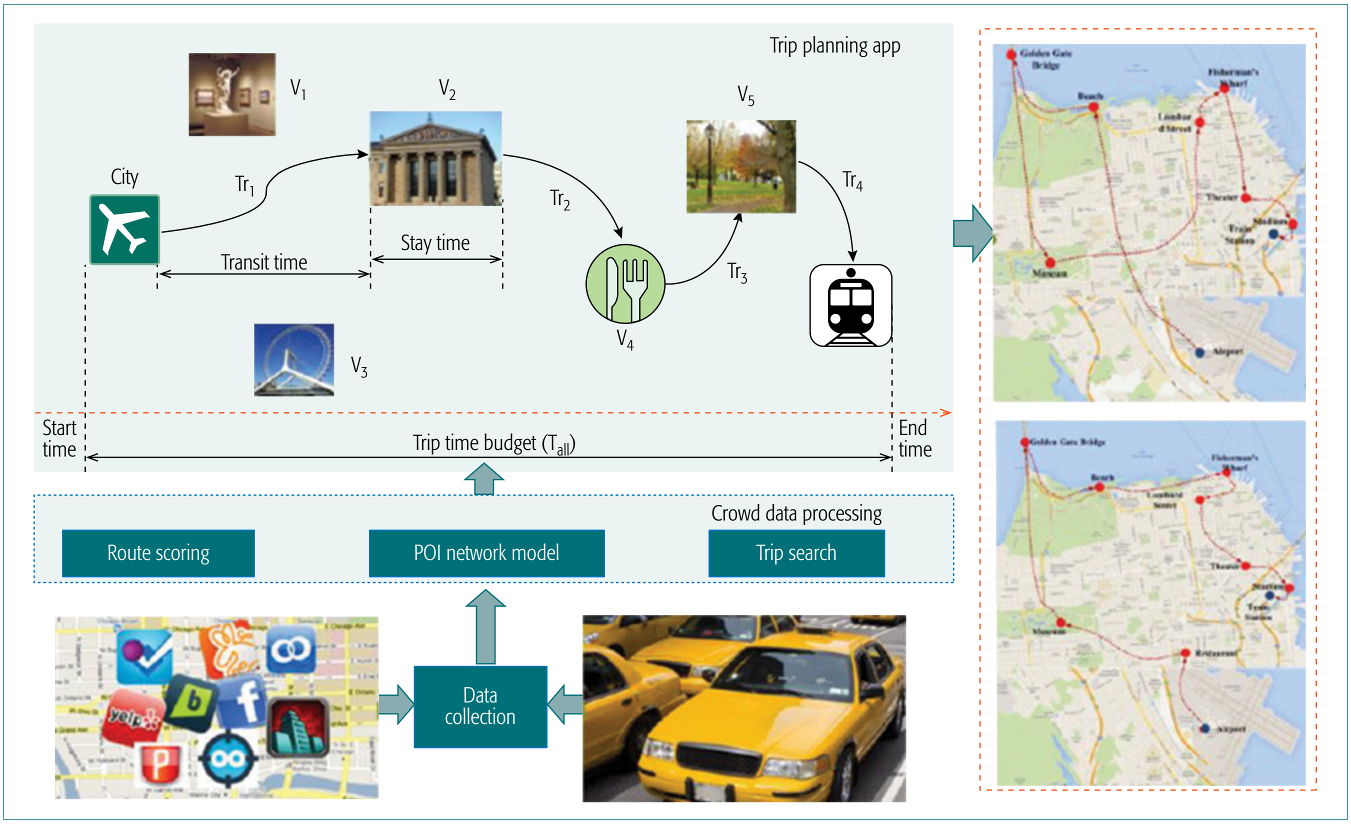


Figure 3. Trip planning over heterogeneous crowdsourced data.

data in crowd sensing. SmartPhoto [7] quantified the utility of crowdsourced photos based on the associated contextual information, such as the smartphone's orientation, position, and location.

MCSC on the Road

The study of MCSC brings new potential in many application areas. This section first makes a summary of two of our ongoing works. The first work is a trip planning application that demonstrates the power of using a combination of participatory sensing and social media data. The second work illustrates our efforts on HMI in MCSC. We also use more examples from other research groups to demonstrate the power of MCSC.

Trip Planning with Heterogeneous Crowdsourced Data

In the introduction, we described the trip planning scenario. A detailed analysis of the problem as well as our solution is presented below. As shown in Fig. 3, in order to plan a trip for visiting a popular tourist city, one needs to select a number of user-preferred POIs among hundreds of available venues (V1 to V5 in Fig. 3), figure out the order in which they are visited, ensure the total time it takes to visit them (the *stay time*), transit from one venue to the next (the *transit time*), and meet the user's *time budget*.

In order to address the trip planning problem, information about the POIs and links among POIs needs to be acquired to build a POI network model. Two types of crowdsourced data sources can be exploited:

- Location-based service networks (LBSNs) (e.g., FourSquare[3]), which can report the popularity (the average number of visitors per day), operating hours, and the best visiting time of the POIs, and an individual user's visiting history
- GPS trajectories of people and taxis, which can indicate the stay time in each place and the transit time between two places

Previous studies rely on one of the two data sources, which results in incomplete POI network models. For instance, Cao *et al.* [8] assumed that the transit time between any two POIs is static and proportionate to their distance. It does not work in real situations because the transit time between venues can be significantly distinct at different time slots due to dynamic traffic conditions.

In view of the above reasons, we propose a personalized traffic-aware trip planning framework called TripPlanner [9], which leverages heterogeneous crowdsourced digital footprints for POI network model construction. More specifically, we make use of two data sources including location data sensed by mobile vehicles (taxi GPS trace data) and user-generated data in FourSquare considering their complementary nature. The TripPlanner system consists of two major components: *route scoring* and *trip search* (Fig. 3). The route scoring module is responsible for estimating the attractiveness of a candidate route to a given user, where two factors are considered: the attractiveness of a venue to the user, and the suitability of the visiting time to each venue. The trip search module applies heuristic algorithms to add user-preferred venues iteratively to the generated routes, with the objective of maximizing the route score and satisfying the travel time constraints. More technical details can be found in [9].

[3] http://www.foursquare.com

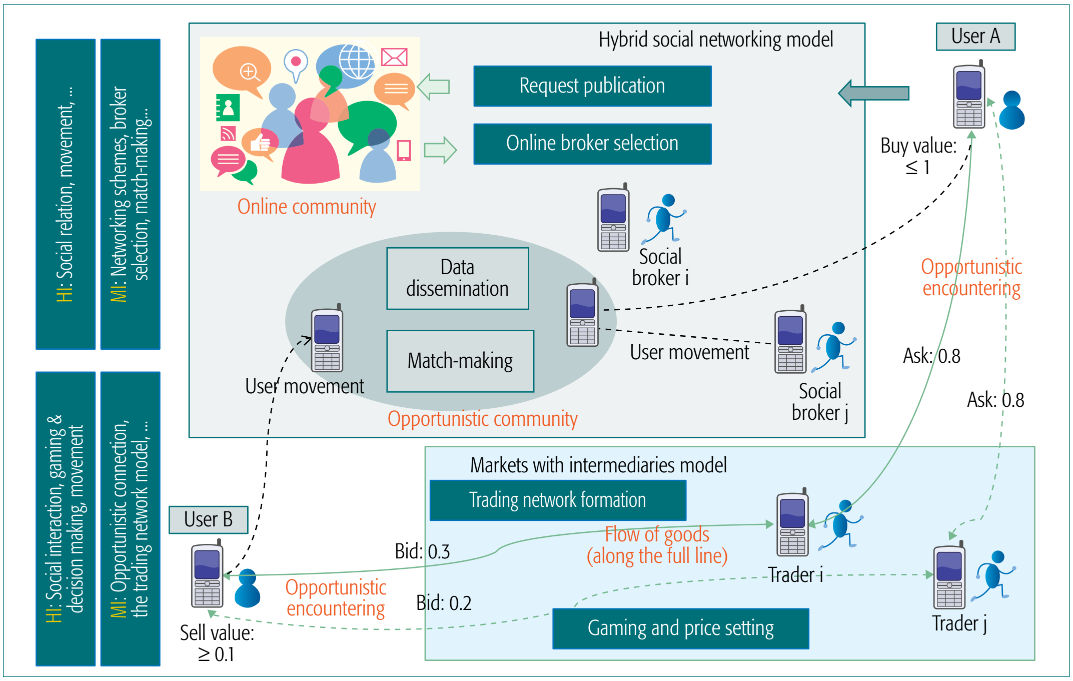

HI: Social relation, movement, ...
MI: Networking schemes, broker selection, match-making...
Hybrid social networking model
User A
Request publication
Online broker selection
Online community
Buy value: ≤ 1
Social broker i
Data dissemination
Match-making
User movement
User movement
Opportunistic encountering
Opportunistic community
Ask: 0.8
Ask: 0.8
Social broker j
HI: Social interaction, gaming & decision making, movement
MI: Opportunistic connection, the trading network model, ...
User B
Markets with intermediaries model
Trading network formation
Flow of goods (along the full line)
Bid: 0.3
Bid: 0.2
Opportunistic encountering
Trader i
Trader j
Gaming and price setting
Sell value: ≥ 0.1

Figure 4. HMI in two crowd-powered data dissemination models.

The success of MCSC relies on the effective transmission of data from individual mobile devices to the destination nodes (e.g., data requesters, back-end servers). The mobility of mobile devices and their carriers not only provides nice coverage for sensing tasks but also brings challenges to data transmission.

We have evaluated the performance of TripPlanner over the datasets collected from San Francisco. 15,759 POIs of the city were obtained from FourSquare. We ranked all POIs in descending order based on their total check-in times. The top 1000 POIs were finally used in our work, considering that tourists would seldom visit POIs with few check-ins. The taxi GPS data of San Francisco was obtained from the CRAWDAD[4] data sharing website. Two similar queries with the same time budget (8.5 h) but different trip starting times were predefined. The start time of the two queries was set to 7:00 a.m. and 10:00 a.m., respectively. As shown on the right of Fig. 3, given the time budget, the user can have seven preferred venues for the first query and eight for the second one. This is because for the first query, the planned route starts around the morning rush hour and thus needs more transit time. We also find that the route for the two queries are different. These results indicate that TripPlanner is traffic-aware with the usage of taxi GPS data. Moreover, with LBSN data in use, our method is more venue-aware. For example, it is suggested that the user go to an Italian restaurant for lunch just after leaving the airport, since it is almost lunch time, and many people visit that venue during that time period.

HMI in Crowd-Powered Data Transmission

The success of MCSC relies on the effective transmission of data from individual mobile devices to the destination nodes (e.g., data requesters, backend servers). The mobility of mobile devices and their carriers not only provides nice coverage for sensing tasks but also brings challenges to data transmission. For instance, network topology, device connectivity, and communication interfaces evolve over time, which makes it hard to find stable routes for crowdsensed data transmission. We refer to it as the "transient" networking issue in MCSC.

To address this issue, people often form opportunistic social networks (OPSNs) [4]. Since the source node and destination nodes may never meet in OPSNs, forwarding data packets from their sender to the nodes of interest is often based on a broker-based solution. In this solution, a selected broker node first stores the data from its sender, carries it while in motion, and then forwards it to intermediate or destination nodes. The assumption is that all users are willing to act as brokers. However, this assumption does not always hold: according to sociological theories, socially selfishness is a basic attribute of human beings, and thus the selected brokers may deny requests from other nodes to save their own resources (e.g., storage, power). Therefore, how to motivate people to participate in opportunistic data transmission becomes a crucial challenge of MCSC. We have developed two HMI-enhanced approaches to promote user participation in broker-based data dissemination. In the first approach people are inspired by social/ethical reasons, while in the second one a solid economic model is leveraged. We illustrate them in Fig. 4 and describe them in detail below.

The Hybrid Social Networking (HSN) Model: The HSN model [10] is inspired by the multi-community involvement and cross-community traversing nature of modern people. For example, at one moment, *Bob* is staying at a place with Internet connection and can communicate with his online friends; later, he may travel by train with merely opportunistic connection to nearby passengers. We use HSN to indicate the smooth switching and collaboration between online and opportunistic communities, as shown in Fig. 4 (top).

One of the key issues addressed by HSN is social selfishness. According to [11], people are willing to help their friends. Following this finding, HSN allows the data sender to choose brokers online from their social connections to avoid

[4] http://crawdad.org/index.html

We have characterized the key features of MCSC, such as explicit/implicit sensing and heterogeneous, cross-space data mining. To fully leverage the power of crowd participation, MCSC needs the deep fusion of human and machine intelligence.

the selfishness problem. The online approach also reduces the cost of popular node selection (to shorten transmission latency, the ones with high probability to meet more people are selected as brokers), which does not require direct contact in the real world. The selected brokers will disseminate the information and do matchmaking with potential interested nodes in opportunistic communities. We compared the performance of HSN with single-community-dependent methods (e.g., the pure opportunistic networking method). Experiment results indicate that great performance improvement is obtained when using HSN [10]. This is because that the integration of online communities shortens the broker selection process, and increases the opportunity to select popular brokers.

HSN is a typical HMI-powered approach, where HI and MI are fused in a hybrid manner via two units. The prior unit is sequential, where social relation and user popularity (HI) are derived first and then used for broker selection (MI); the latter is parallel, where user movement (HI) and matchmaking (MI) work simultaneously to fulfill the data transmission task.

The Market Model with Intermediaries: Inspired by how buyers and sellers interact in traditional markets, we introduce the model of markets with intermediaries as an incentive mechanism to stimulate node cooperation in MCSC. In many markets (e.g., stock markets, agricultural markets in developing countries), individual buyers and sellers do not interact directly with each other, but trade via intermediaries instead [11]. These intermediaries, also called traders, often set the prices of transactions. The similarities between markets with intermediaries and data dissemination in opportunistic social networks drive us to use the former as an incentive mechanism in the latter. A data sender is like a seller in a market, and a data receiver is like a buyer: she "buys" a unit of goods if she receives the data from traders. As shown in Fig. 4 (bottom), the connections (built based on direct contacts) among traders, sellers, and buyers will form a trading network.

For simplicity, each seller i initially holds one unit of the good, which she values at v_i; she is willing to sell it at any price that is at least v_i (*sell value*). Each buyer j values one copy of the good at v_j (*buy value*) and will try to obtain a copy of the good if she can do it by paying no more than v_j. Each trader t offers a *bid price* b_{ti} to each seller i with which she connects, and an *ask price* a_{tj} to each connected buyer j. After receiving offered prices by traders, each seller and buyer can only choose at most one trader with which to deal. A *flow of goods* from sellers through traders to buyers is finally generated. Figure 4 gives an example of such a trading model, including the bid price, ask price, and flow of goods (indicated by the solid lines).

In this approach, transactions are made based on a game process. In the game, a trader's strategy is a choice of bid and ask prices to propose to each neighboring seller and buyer; a seller or buyer's strategy is a choice of a neighboring trader to deal with, or the decision not to take part in a transaction. The participants (sellers, traders, and buyers) are motivated to get their payoffs. Note that in traditional markets, currency is generally used as a medium for buying and selling. In our model, we further expand this concept by allowing virtual currency to be used. That is to say, services can pay "virtual coins" to the participants, and participants need to spend some coins in service usage. In our case, both data senders and brokers can receive their payoffs in "virtual coins." Experiments indicate that our approach can enhance user participation in MCSC data transmission [12]. HMI is also embedded in a hybrid way in this case. In association with social interactions and price settings (HI), the trading network model (MI) is formed (a parallel unit); afterward, the stakeholders bargain and make decisions (HI) to fulfill the data transmission task (a sequential unit).

Other Efforts on MCSC

Beyond our recent works, MCSC has also been found useful in several other studies. Zheng *et al.* [13] proposed a model to infer fine-grained air quality information throughout a city. The learning model leveraged the air quality data reported by existing monitor stations and a variety of crowd-contributed data in the city, such as traffic flow (offline space) and POIs in LBSNs (online space). Du *et al*. [14] leveraged a combination of social media and historical physical activity data to predict activity attendance and facilitate social interaction in the real world.

Conclusion

MCSC shows its difference in the literature by leveraging varying levels of user participation in data contribution and aggregating heterogeneous crowdsourced data for novel service provision. We have characterized the key features of MCSC, such as explicit/implicit sensing, and heterogeneous cross-space data mining. To fully leverage the power of crowd participation, MCSC needs deep fusion of human and machine intelligence. Three HMI patterns are thus identified. We further present the early efforts on MCSC.

As an emerging paradigm for large-scale sensing, numerous challenges and research opportunities remain to be investigated. First, MCSC is an instance that bridges the gap between cyber space and physical space. The problems to solve in such a hyperspace are much more complex, and need to integrate various HI and MI units as interdependent parameters in a unique solution. Second, in hyperspace, we should exploit cross-space features for aggregated sensing and data understanding. Third, as community-enabled sensing, the generic features of a community, such as sensing scale (ranging from a group to an urban scale), community structure, and user collaboration should be further studied [5, 15], which are paid little attention in existing studies.

Acknowledgment

This work was partially supported by the National Basic Research Program of China (no. 2015CB352400), the National Natural Science Foundation of China (no. 61332005, 61373119, 61222209), and the Scientific and Technology New Star of Shaanxi Province (2014KJXX-39).

References

[1] D. Cuff, M. Hansen, and J. Kang, "Urban Sensing: Out of the Woods," *Commun. ACM*, vol. 51, no. 3, 2008, pp. 24–33.

[2] J. Burke *et al.*, "Participatory Sensing," *Proc. ACM Sensys Wksps.*, 2006.

[3] R. K. Ganti, F. Ye, and H. Lei, "Mobile Crowdsensing: Current State and Future Challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, Nov. 2011, pp. 32–39.

[4] B. Guo *et al.*, "Opportunistic IoT: Exploring the Harmonious Interaction between Human and the Internet of Things," *J. Network and Computer Applications*, vol. 34, no. 6, 2013, pp. 1531–39.

[5] B. Guo *et al.*, "Cross-Community Sensing and Mining," *IEEE Commun. Mag.*, vol. 52, no. 8, 2014, pp. 144–52.

[6] G. Ding et al., "Robust Spectrum Sensing with Crowd Sensors," *IEEE Trans. Commun.*, vol. 62, no. 9, 2014, pp. 3129–43.

[7] Y. Wang *et al.*, "Smartphoto: A Resource-Aware Crowdsourcing Approach for Image Sensing with Smartphones," *Proc. ACM MobiHoc '14*, 2014, pp. 113–22.

[8] X. Cao *et al.*, "Keyword-Aware Optimal Route Search," *Proc. VLDB Conf.*, 2012, pp.1136–47.

[9] C. Chen *et al.*, "TripPlanner: Personalized Trip Planning Leveraging Heterogeneous Crowdsourced Digital Footprints," *IEEE Trans. Intell. Transportation Sys.*, 2014.

[10] B. Guo *et al.*, "Hybrid SN: Interlinking Opportunistic and Online Communities to Augment Information Dissemination," *Proc. IEEE UIC '12*, 2012, pp. 188–95.

[11] D. Easley and J. Kleinberg, *Networks, Crowds, and Markets*, Cambridge Univ. Press, 2010.

[12] S. Huangfu *et al.,* "Using the Model of Markets with Intermediaries as an Incentive Scheme for Opportunistic Social Networks," *Proc. IEEE UIC '13*, 2013.

[13] Y. Zheng, F. Liu, and H. P. Hsieh, "U-Air: When Urban Air Quality Inference Meets Big Data," *Proc. KDD '13*, 2013, pp. 1436–44.

[14] R. Du *et al.*, "Predicting Activity Attendance in Event-Based Social Networks: Content, Context and Social Influence," *Proc. ACM UbiComp'14*, 2014, pp. 425–34.

[15] D. Zhang, B. Guo, and Z. Yu, "The Emergence of Social and Community Intelligence," *Computer*, vol. 44, no. 7, 2011, pp. 21–28.

Biographies

Bin Guo (guob@nwpu.edu.cn) is a professor at the School of Computer Science, Northwestern Polytechnical University, China. He received his Ph.D. degree from Keio University, Tokyo, Japan, in 2009. During 2009–2011, he was a post-doctoral researcher at Institut TELECOM SudParis, France. His research interests include pervasive computing, social computing, and mobile crowd sensing. He has served as an Editor or Guest Editor for a number of international journals, such as *IEEE Communications Magazine*, *IEEE Transactions on Human-Machine Systems*, *IEEE IT Professional, Personal and Ubiquitous Computing*, and *ACM Transactions on Intelligent Systems and Technology*. He has served as General/Program/Workshop Chair for several conferences, including IEEE UIC, IEEE SCI@PerCom, IEEE iThings, and so on.

Chao Chen (chunchaaonwpu@gmail.com) is an associate professor at Chongqing University, China. He received his Ph.D. degree from Pierre and Marie Curie University, France, in 2014. His research interests include pervasive computing, social network analysis, and big data analytics for smart cities.

Daqing Zhang (daqing.zhang@it-sudparis.eu) is a professor at Institut TELECOM SudParis. He obtained his Ph.D. from University of Rome "La Sapienza," Italy in 1996. He has organized a dozen international conferences as General Chair or Program Chair. He is an Associate Editor for four journals including *ACM TIST*, *Springer Journal of Ambient Intelligence and Humanized Computing*, and two others. He has also served on the Technical Committees for conferences such as UbiComp, Pervasive, and PerCom. His research interests include ubiquitous computing, context-aware computing, big data analytics, and social computing.

Zhiwen Yu (zhiwenyu@nwpu.edu.cn) is a professor and director of the Department of Discipline Construction, Northwestern Polytechnical University, China. He worked as an Alexander Von Humboldt Fellow at Mannheim University, Germany, from November 2009 to October 2010, and a research fellow at Kyoto University, Japan, from February 2007 to January 2009. His research interests cover pervasive computing, context-aware systems, and personalization. He has served as an Editor or Guest Editor for a number of journals, such as *IEEE Communications Magazine*, *PUC*, and *ACM TIST*. He was General Chair of UIC 2014 and Workshop Chair of UbiComp 2011.

Alvin Chin (alvin.chin@utoronto.ca) is a senior researcher at Microsoft (formerly Nokia) and previously was in the Mobile Social Experiences group at Nokia Research Center, Beijing, from 2008 to 2012. He has Bachelor's and Master's degrees in computer engineering from the University of Waterloo and a Ph.D. in computer science from the University of Toronto, Canada. His research interests include social networking and ubiquitous computing. He has served as General/Program/Workshop Chair for several conferences including IEEE CPSCom, IEEE CIT, IEEE UIC, ACM Hypertext, and so on. He is an Associate Editor for the *New Review of Hypermedia and Multimedia* and Co-Editor of the Special Issue on Smartphone-Based Technologies, Applications and Systems of *ACM TOMM*.

Accepted from Open Call

Long Term Evolution for Wireless Railway Communications:Testbed Deployment and Performance Evaluation

Yong-Soo Song, Juyeop Kim, Sang Won Choi, and Yong-Kyu Kim

The authors show the feasibility of the LTE-R testbed with essentially IP-based network architecture. Specifically, they discuss procedures of deploying LTE-R by describing their construction of a testbed in a commercial railway through cell planning and optimization. Then they demonstrate the performance enabled by the implementation of a testbed for LTE-R.

Abstract

In this article, we show the feasibility of the LTE-R testbed with essentially IP-based network architecture. Specifically, we discuss procedures of deploying LTE-R by describing our construction of a testbed in a commercial railway through cell planning and optimization. Then we demonstrate the performance enabled by the implementation of a testbed for LTE-R. We confirm that not only reliable communications but also multimedia services requiring high data rates are feasible, which gives us some guarantee of the prosperity of various advanced train services. We also discuss a number of valuable technical communication issues related to inherent characteristics of railway communications that are unlike those of commercial wireless communications.

Introduction

Over the past few decades, mobile communications have been subject to incredible change and innovation, mainly due to a massive increase in user data requirements [1–2]. One major milestone was the creation of all-IP-based services. The revolutionary change was realized through the conversion from the convergence of circuit-switched and packet-switched networks into packet-switched networks alone. Other milestones weree the introduction of orthogonal frequency-division multiple access (OFDMA) and new core network (CN) architecture, referred to as Evolved Packet Core (EPC), from the radio access network (RAN) and CN perspectives, respectively, which led to fourth generation (4G) cellular communications [1]. In recent years, beyond 4G cellular communications, the imminent arrival of the 5G communications era is expected. In 5G communications, we expect the peak data rate to be measured in tens of gigabits per second, providing end users with 1 Gb/s [2].

In contrast to the rapid and innovative changes in commercial mobile communications, the development of railway communications has been rather more measured. One of the main reasons for this deliberation is that railway communications are mission-critical. The other reason is that it can take many years to verify the applicability of any railway communication standard in terms of safety in the field. As representative railway communications, the Global System for Mobile Railway (GSM-R) is the most widely used standard, particularly in Europe. GSM-R is a unique standard for an integrated wireless railway communications system, and its stability has been verified for more than 10 years [3, 4]. However, due to limited transmission capacity, GSM-R has been applied mainly to data communications for train control, that is, the European Train Control System (ETCS).

In recent years, the center of gravity of research and development in railway communications systems has been moving toward packet-switched wireless communications [5–9]. The International Union of Railways (UIC) recently agreed that the current GSM-R is not sufficient to provide new railway services such as video supervision and real-time monitoring. Also, the operators of commercial mobile communications have begun to deploy and operate Long Term Evolution (LTE), and the demand for GSM devices has decreased. Consequently, the procurement of GSM devices has become more difficult, which has caused problems in terms of maintenance. These issues caused various alternatives, including general packet radio service (GPRS), to be considered in order to improve the performance of conventional systems [5, 6], but it was eventually concluded that the most sensible approach was to adopt a new system altogether. Research efforts are now oriented toward LTE as the next generation technology [7–9]. In particular, the capabilities of LTE in terms of future railway operational needs are currently being reviewed under the Future Railways Mobile Communications System (FRMCS) project, which was initiated by UIC in 2013.

In this situation, it is reasonable to expect railway communications to keep pace with mobile communications. Nevertheless, the question "*Will the mobile communication technologies satisfy railway communications' own requirements?*" must be answered. Motivated by this fundamental question, we have proposed LTE Railway (LTE-R) as an integrated wireless railway communications system by validating a deployed testbed of LTE-R. In this article, our main focus is to assess the feasibility of LTE-R by dealing with

The authors are with the Korea Railroad Research Institute. Sang Won Choi is the corresponding author.

0163-6804/16/$25.00 © 2016 IEEE

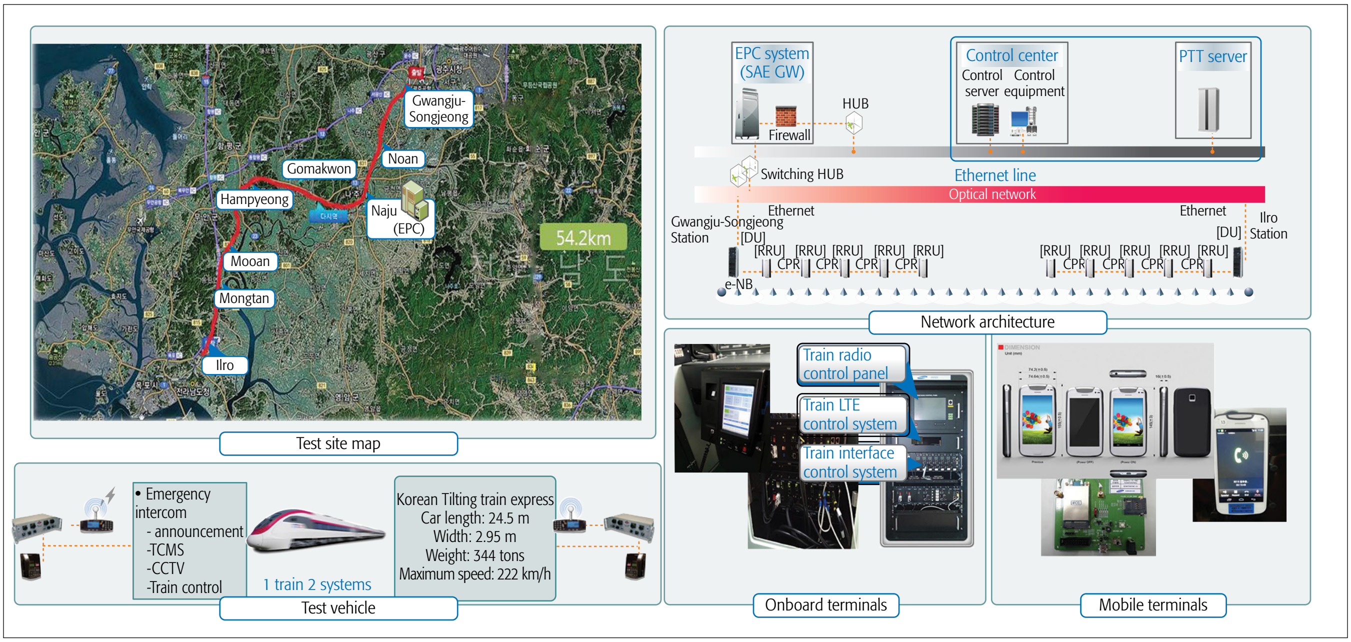


Figure 1. LTE-R Test-Bed.

the performance of LTE-R on the constructed LTE-R testbed. Lastly, further technical research issues are addressed from the inherent peculiarity of railway communication networks.

LTE-R Deployment

The LTE-R consists of base station equipments, core equipments, and LTE-R service servers. The base station consists of multiple radio units (RUs) and a digital unit (DU). The RU is for radio wave transmission and reception, and the DU is responsible for various types of processing, including the use of radio resources for packet scheduling and effective processing of the physical layer. The core equipment comprises a serving and packet gateway (S/P-GW), mobility management entity (MME), home subscriber server (HSS), and policy and charging rules function (PCRF). These entities are mainly for effective bearer management, mobility management including handover, subscriber management, and quality of service (QoS) management, respectively, based on priority. The LTE-R service servers consist of a call server, a push to talk (PTT) server, a conference server, a video server, a control server, and a recoding server. These servers support diverse group communications including emergency calls, private calls, and multimedia broadcasting via real-time protocol, group management, call management, interlocking with PCRF, and so on.

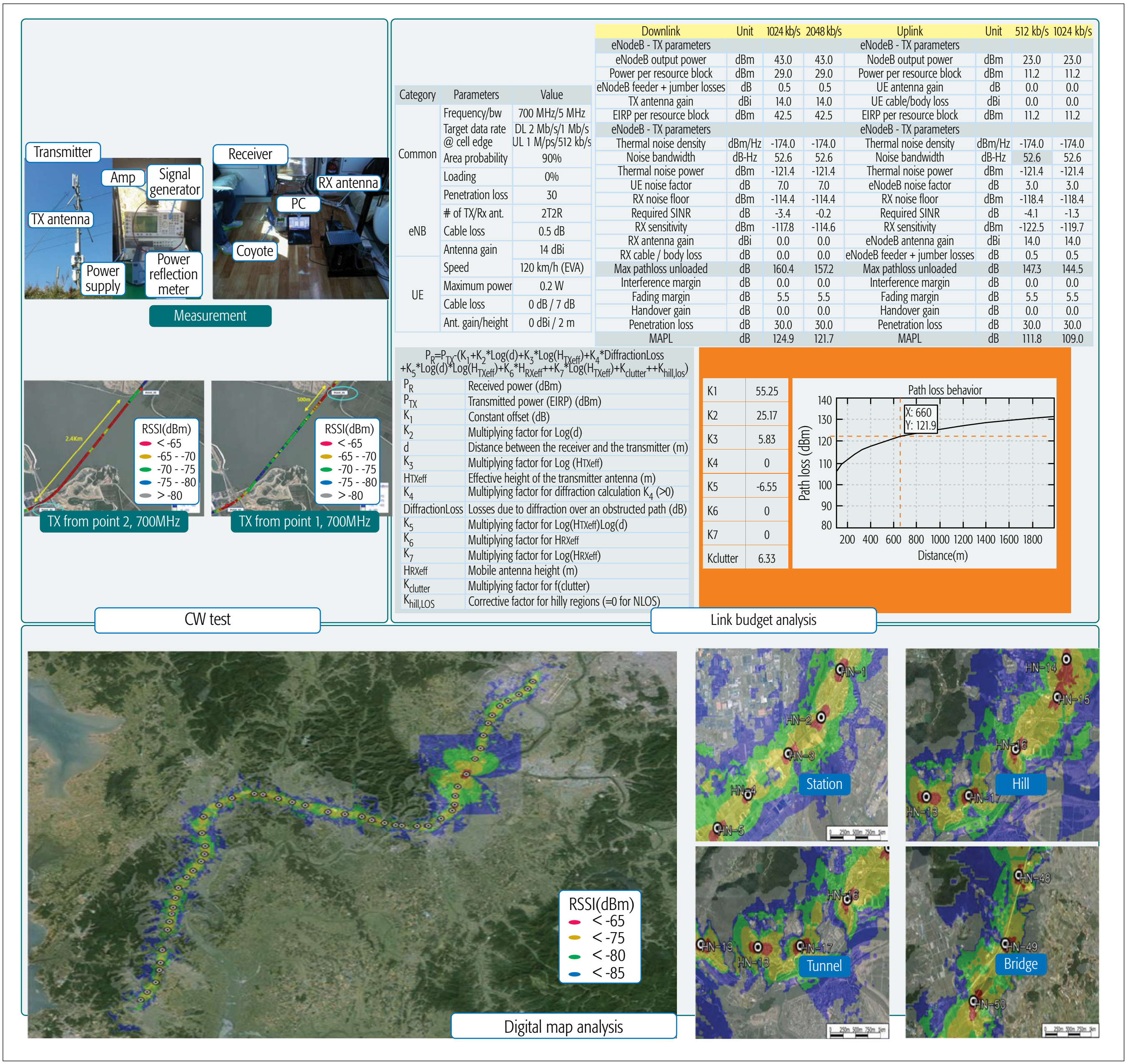
As for terminals, there are two kinds of terminals: an onboard terminal and a mobile terminal. The onboard terminal is located at each driver's cabin, and performs signal transmission and reception through the antenna on the roof of the train. It interfaces with locomotive engineers and other onboard devices such as closed circuit TV (CCTV) through the train radio control panel (TRCP) and train interface control (TIC), respectively. Note that the architecture of the mobile terminal is similar to that of the commercial mobile terminal except for having a physical button for floor control for a PTT service.

Figure 1 shows the overview of our LTE-R testbed built on a commercial railway. The test site ran from Gwangju-Songjeong station to Ilro station on the Honam line. On the 54.4 km test section, we deployed Samsung Rel. 9 LTE network devices including 51 RUs, 7 DUs, and 1 EPC. Each RU, having directional antennas, was located near the trackside, and DUs were located to cover the RUs. An EPC was located in Naju station and was connected to DUs by wired lines.

Cell Planning

The first stage in the cell deployment was to conduct cell planning, in which the positions of the RUs and DUs were decided based on a certain channel model. The basic rule is that the service region should be sufficiently covered while minimizing the number of RUs to deploy. In addition, geographical features, antenna height, and mobility pattern must be considered in cell planning. Various channel models must be utilized in a railway environment because the geographical features near a railway include mountain valleys, hills, curve sections, bridges, tunnels, and so on.

To identify the channel model suited to the railway environment, we used a continuous wave (CW) test as the step of field quality measurement. A reference signal was emitted for the given antenna height and transmission power, and the path loss was estimated by observing the received power strength of a mobile terminal in a cabin. The upper left part of Fig. 2 shows an example of a CW test in a bridge for both straight and curved sections. Based on the worst case, we concluded that the cell radius required to maintain the received signal level at –70 dBm was about 500 m. In addition, the result shows that the signal attenuation of the mobile terminal was more significant in the straight section than in the curved section. This reveals that the characteristics of the wireless channel can be changed according to the direction of motion of the train.



Category	Parameters	Value
Common	Frequency/bw	700 MHz/5 MHz
	Target data rate @ cell edge	DL 2 Mb/s/1 Mb/s UL 1 M/ps/512 kb/s
	Area probability	90%
	Loading	0%
	Penetration loss	30
eNB	# of TX/Rx ant.	2T2R
	Cable loss	0.5 dB
	Antenna gain	14 dBi
UE	Speed	120 km/h (EVA)
	Maximum power	0.2 W
	Cable loss	0 dB / 7 dB
	Ant. gain/height	0 dBi / 2 m

Downlink	Unit	1024 kb/s	2048 kb/s	Uplink	Unit	512 kb/s	1024 kb/s
eNodeB - TX parameters				eNodeB - TX parameters			
eNodeB output power	dBm	43.0	43.0	NodeB output power	dBm	23.0	23.0
Power per resource block	dBm	29.0	29.0	Power per resource block	dBm	11.2	11.2
eNodeB feeder + jumber losses	dB	0.5	0.5	UE antenna gain	dB	0.0	0.0
TX antenna gain	dBi	14.0	14.0	UE cable/body loss	dBi	0.0	0.0
EIRP per resource block	dBm	42.5	42.5	EIRP per resource block	dBm	11.2	11.2
eNodeB - TX parameters				eNodeB - TX parameters			
Thermal noise density	dBm/Hz	-174.0	-174.0	Thermal noise density	dBm/Hz	-174.0	-174.0
Noise bandwidth	dB-Hz	52.6	52.6	Noise bandwidth	dB-Hz	52.6	52.6
Thermal noise power	dBm	-121.4	-121.4	Thermal noise power	dBm	-121.4	-121.4
UE noise factor	dB	7.0	7.0	eNodeB noise factor	dB	3.0	3.0
RX noise floor	dBm	-114.4	-114.4	RX noise floor	dBm	-118.4	-118.4
Required SINR	dB	-3.4	-0.2	Required SINR	dB	-4.1	-1.3
RX sensitivity	dBm	-117.8	-114.6	RX sensitivity	dBm	-122.5	-119.7
RX antenna gain	dBi	0.0	0.0	eNodeB antenna gain	dBi	14.0	14.0
RX cable / body loss	dB	0.0	0.0	eNodeB feeder + jumber losses	dB	0.5	0.5
Max pathloss unloaded	dB	160.4	157.2	Max pathloss unloaded	dB	147.3	144.5
Interference margin	dB	0.0	0.0	Interference margin	dB	0.0	0.0
Fading margin	dB	5.5	5.5	Fading margin	dB	5.5	5.5
Handover gain	dB	0.0	0.0	Handover gain	dB	0.0	0.0
Penetration loss	dB	30.0	30.0	Penetration loss	dB	30.0	30.0
MAPL	dB	124.9	121.7	MAPL	dB	111.8	109.0

$$P_R=P_{TX}-(K_1+K_2*Log(d)+K_3*Log(H_{TXeff})+K_4*DiffractionLoss +K_5*Log(d)*Log(H_{TXeff})+K_6*H_{RXeff}++K_7*Log(H_{TXeff})+K_{clutter}++K_{hill,los})$$

P_R	Received power (dBm)
P_{TX}	Transmitted power (EIRP) (dBm)
K_1	Constant offset (dB)
K_2	Multiplying factor for Log(d)
d	Distance between the receiver and the transmitter (m)
K_3	Multiplying factor for Log (HTXeff)
HTXeff	Effective height of the transmitter antenna (m)
K_4	Multiplying factor for diffraction calculation K_4 (>0)
DiffractionLoss	Losses due to diffraction over an obstructed path (dB)
K_5	Multiplying factor for Log(HTXeff)Log(d)
K_6	Multiplying factor for HRXeff
K_7	Multiplying factor for Log(HRXeff)
HRXeff	Mobile antenna height (m)
$K_{clutter}$	Multiplying factor for f(clutter)
$K_{hill,LOS}$	Corrective factor for hilly regions (=0 for NLOS)

K1	55.25
K2	25.17
K3	5.83
K4	0
K5	-6.55
K6	0
K7	0
Kclutter	6.33

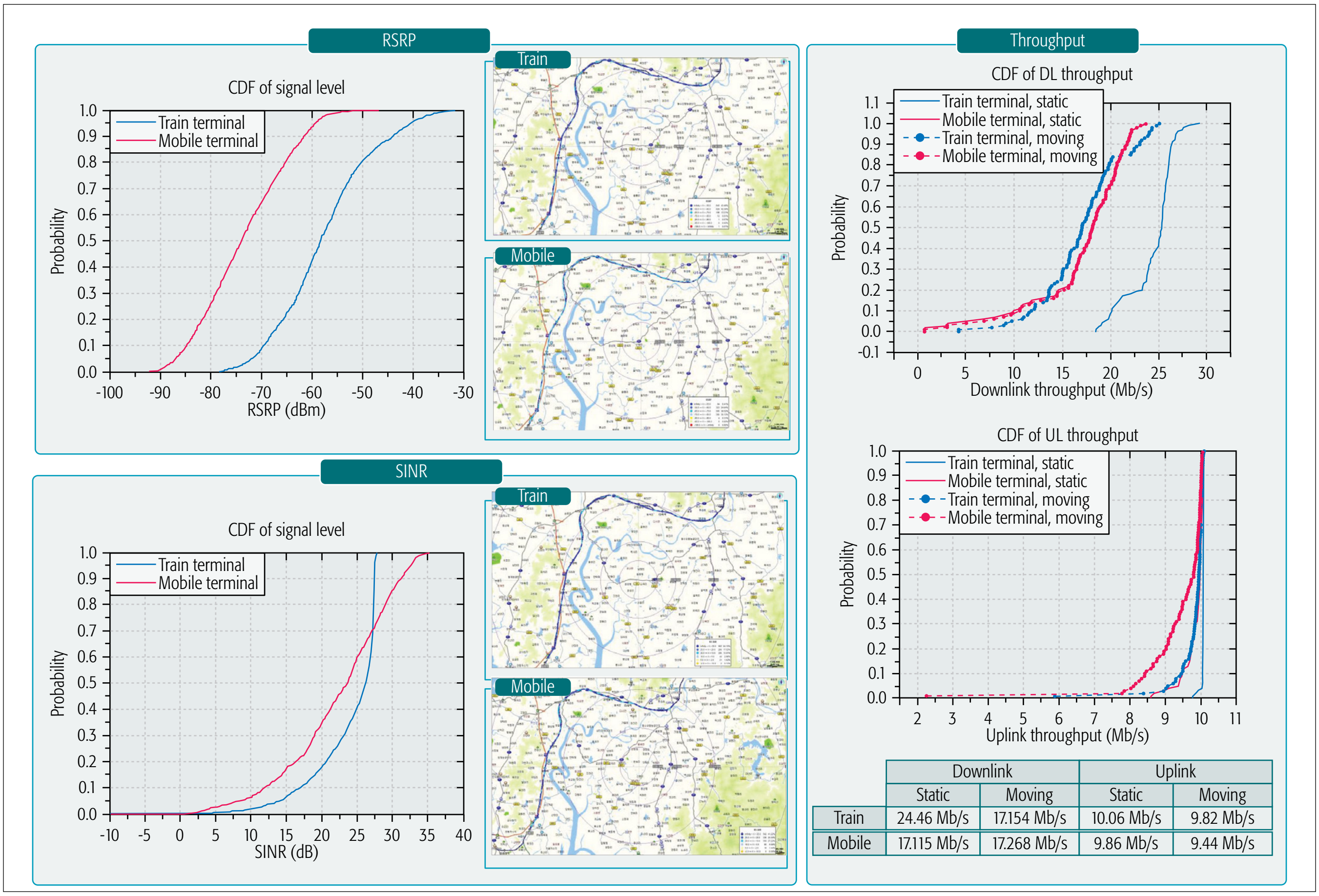
Figure 2. Cell deployment procedure.

Taking those channel characteristics into account, we defined the design criteria for cell planning. The service range and the area of a cell can be determined using a link budget analysis, in which the maximum allowable path loss (MAPL) is calculated from various loss and gain factors between an RU and a terminal. The upper right part of Fig. 2 shows the detailed derivation of a link budget analysis. Using the Atoll Simulation Tool, we derived a standard propagation model (SPM) based on the Hata model. The parameter $K_{clutter}$ used in the SPM was defined as the measured value from the CW test. After the cell plan was completed, it was verified using a coverage simulation by digital map analysis. The simulation, which is shown in the lower part of Fig. 2, indicates that the received signal level in most of the railway service region was greater than –85 dBm.

Cell Optimization

The objective of the optimization procedure is to pass through the target region after cell installation in order to ensure cell coverage. Issues found during cell optimization are generally resolved by adjusting the tilt or swing angle of antennas, varying the transmission power and cell parameters such as time to trigger (TTT) or handover hysteresis. After installing RUs and DUs in the field according to the cell plan, we passed through the test section and measured the received signal level on the train in order to check whether the cell coverage was formed as we intended.

Several issues were found and resolved during cell optimization. For example, the signal-to-interference-plus-noise ratio (SINR) was very low in a region near a hill, because the hill was higher than the antenna height of RU#16, and the



	Downlink		Uplink	
	Static	Moving	Static	Moving
Train	24.46 Mb/s	17.154 Mb/s	10.06 Mb/s	9.82 Mb/s
Mobile	17.115 Mb/s	17.268 Mb/s	9.86 Mb/s	9.44 Mb/s

Figure 3. Performance of the cell optimization.

signal from it flowed weakly in the region. By adjusting the tilt angle of RU#16 and the transmission power of RU#17, we could increase the SINR by 5–7 dB. In addition, handover ping-pong frequently occurred at regions near train stations, because the train passed slowly through the regions in which the received signal levels of the adjacent cells were similar. By adjusting the tilt angles of the antennas and the transmission power, we succeeded in suppressing handover ping-pong in the region.

To check the validity of the cell optimization, we applied a set of basic service tests to identify any other issues related to the service aspect. We measured performance of the testbed to assess whether it could meet the requirements. The measurements were conducted via both onboard and mobile terminals and in the train, which operated under a general scenario at speeds of up to 160 km/h. We measured the reference signal received power (RSRP) and SINR for each terminal, and also tested the handover behavior and FTP throughput. During the FTP throughput test, we accessed an FTP server and tried downloading a 50 MB file for the downlink and a 10 MB file for the uplink.

As shown in Fig. 3, our testbed meets the requirements for LTE-R. The RSRP level in the 95 percent coverage region was estimated at about –72 dBm for the onboard terminal and –83.5 dBm for the mobile terminal. This reveals that our testbed was deployed appropriately, covering the railway region, and that it is ready to provide various railway services in the region. In addition, no handover failure events occurred at either of the terminals. This shows the ability of the testbed to provide railway services seamlessly in a moving environment.

We note from the results of the SINR distribution in Fig. 3 that the quality of the received signal is proper in overall regions. The SINR for the 95 percent coverage region was 14 dB for the onboard terminal and 8 dB for the mobile terminal, which implies that the signal quality was quite good in the cell edge region. This is because the distance between RU and the track was rather small, and the line of sight (LOS) between an RU and a terminal was guaranteed in overall regions. These characteristics led to a high score on the FTP throughput test in which the downlink average throughput was greater than 17 Mb/s and the uplink average throughput was around 10 Mb/s. This result provides some assurance that the testbed is capable of providing various railway services simultaneously from the perspective of data rate.

Feasibility Verification of LTE-R

In this section, we validate our LTE-R testbed constructed in a railway through basic performance tests as a feasibility proof. As for the LTE-R testbed, we use the system parameters depicted in Table 1. In the last part, we address the potential of LTE-R for public safety based on the performance evaluation tests.

System parameters	Values
3GPP LTE Spec.	Release 9
Bandwidth (BW)	Scalable BW 5 MHz (Band 13, 10 MHz)
FFT size	512
Subcarrier spacing	15 kHz
Duplex mode	FDD mode
Maximum transmit power	23 dBm (onboard and mobile terminal), 43 dBm (eNodeB)
Maximum speed of train	222 km/h
Test-bed scale	About 50 km
eNodeB output power	43 dBm
UE output power	23 dBm
Antenna type	Directional antenna
MIMO antenna configuration	2 TX, 2 RX
Half power beam width (HPBW)	(Horizontal) 65° ± 6° (Vertical) 15° ± 3°
Front to back ratio (FBR)	(Horizontal) ≥ 23dB (at 180° ± 30° (Vertical) ≥ 23dB (at 180° ± 60°)
Array gain	≥ 14 dBi
Effective isotropic radiated power (EIRP) per resource block	42.5

Table 1. System parameters of the LTE-R testbed. 3GPP: Third Generation Partnership Project; FDD: frequency-division duplex.

Performance Requirements

For guaranteeing LTE-R performance from the end-to-end service perspective, the performance index needs to be considered first. Table 2 depicts performance requirements for GSM-R and LTE-R. Note that in Table 2, we illustrate the upper bound of packet data rate that can be achieved using the available coding scheme (8-phase shift keying, PSK) based on Enhanced Data for General Packet Radio Service Evolution (EDGE) [10]. Contrary to GSM-R, LTE-R can afford to cover data transmission regarding vital and non-vital data. Here, we establish criteria for data rate by considering several data rates to support typically used train service [11]. In terms of data integrity, probabilities of data loss, data duplication, out-of-sequence data, and data corruption can be considered in general packet radio service (GPRS). In LTE-R, data plane management is accomplished in radio link control (RLC) along with cyclic redundancy check (CRC). Hence, it is reasonable to consider only the data loss rate of a data packet.

Voice communications on LTE is accomplished via voice over IP (VoIP), which is different from circuit-switched GSM for original GSM-R. Inherently, GSM-R has a performance limitation regarding packet delay and call connection establishment time mainly due to the relatively long frame structure and naive protocols. Considering the evolution of LTE-R in call connectivity, we apply stricter and more differentiated criteria for LTE-R than GSM-R according to each type of group communications.

On the other hand, it must be noted that a functional split between the radio access and CN has been realized in the LTE network. Specifically, all radio functionalities such as RLC and medium access control (MAC) are placed in eNodeB, which is a logical combination of a conventional NodeB and a radio network controller (RNC). Consequently, a flat network structure is formed. By considering the improvement in terms of handover latency coming from the LTE's flat structure, stricter criteria of LTE-R were applied than those of GSM-R.

Table 2 depicts the performance requirements of LTE-R required for serving railway services in comparison to those of GSM-R. It should be noted that Table 2 includes the assumption that an LTE-R system should guarantee the required performance as stably as GSM-R, which is a circuit-switched network. It implies that LTE-R, which is a packet-switched network, should satisfy the performance requirements in Table 2 in any network condition including network congestion. Basically, the LTE-R system can achieve stability by its own QoS management scheme [12]. By managing QoS per data bearer, the LTE-R system can satisfy the performance requirements for each corresponding railway service.

Performance Test Cases

Performance test cases aim to check whether the performance requirements are met. In performance test cases, we deduct a performance index as a quantified value by repetitive measurement, and determine whether each criterion is met. Here, the criteria are described in Table 2. The following are the representative test cases:

- Network registration: Make the terminal power on, check whether it finishes the attach procedure properly, and measure the network registration time.
- Call connectivity: Try a point-to-point voice call, a video call, an emergency call, or a conference call, check if it succeeds, and measure the call setup time.
- PTT connectivity: Try floor control during PTT service, check if it succeeds, and measure the setup time.
- Long duration call: Keep a call, check whether a call drop event happens, check whether handover failure happens, and measure the handover delay time.
- FTP throughput: Try downloading or uploading a certain file through FTP and measure the data rate.
- Ping data transfer: Send pings consecutively, measure round-trip time (RTT), and check whether data pause happens for a certain time.

Performance Analysis

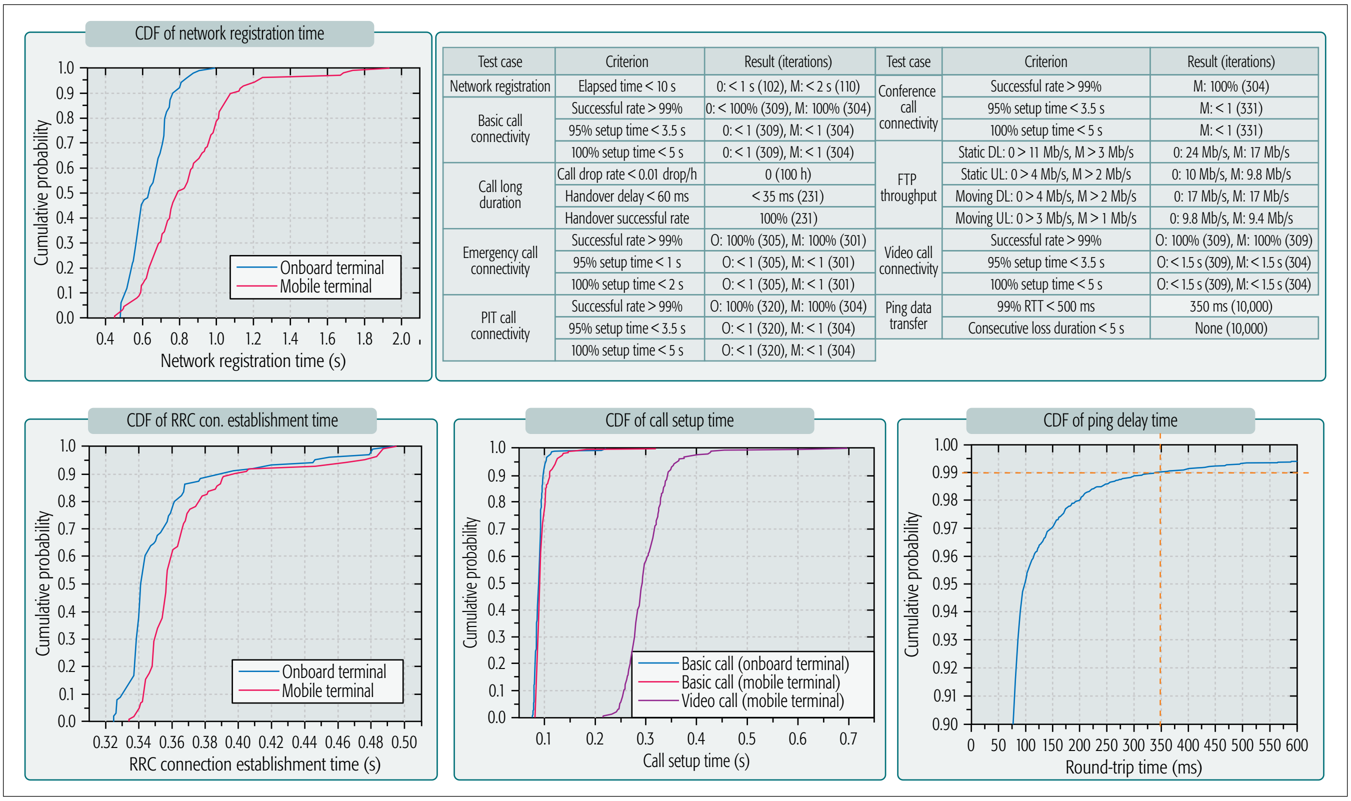
Figure 4 summarizes the results of the performance test. The results satisfy all the criteria, revealing that LTE-R fully satisfies the corresponding performance requirements and is an appropriate replacement system beyond GSM-R. From the control plane latency perspective, it is observed that LTE-R managed remarkable advancement compared to GSM-R. For example, it took at most 2 s to finish network registration, where network registration can take up to 20 s in GSM-R. Even considering the latency due to radio resource control (RRC) connection establishment, the call setup time was less than 0.7 s for a basic call and 0.9 s for a video call. The main factors enabling LTE-R to reduce the control plane latency are the shorter frame length and the evolved CN architecture, which is optimized to transfer packets in a short time.

Through the long call duration test, it is confirmed that service outage rarely happens during train service. Note that it is hard to meet this performance criterion in commercial mobile communications, in which some weak signal regions are inevitable in spite of cell optimization. Although it is better to perform cell optimization in a railway environment, it still requires sophisticated skills to secure cell coverage throughout railway service regions and to keep connectivity in weak signal environments in order to achieve the goal. In addition, there was no handover failure event, and the handover latency was less than 40 ms, which mainly contributes to suppress occurrence of service outage events.

In terms of packet transmission, LTE-R outperforms GSM-R for all performance indices. Due to LTE-R's use of relatively larger bandwidth and relatively higher spectral efficiency with the aid of sophisticated LTE technologies such as multiple-input multiple-output (MIMO) and orthogonal frequency-division multiple access (OFDMA), both the downlink and uplink throughputs for LTE-R are much higher than those for GSM-R. This indicates that LTE-R is qualified to provide various kinds of advanced railway services which require much greater data transfer, which is difficult to see when using GSM-R. Also, the end-to-end packet delay was less than 175 ms (with 99 percent), which is much shorter than the typical latency time in GSM-R, and data pause was kept to no more than 5 s. Furthermore, it was observed that this characteristic was confirmed functionally even in the full loading case. This reveals that LTE-R has potential to carry train control traffic and provide train control service smoothly.

Performance index	Value for GSM-R [3]	Value for LTE [10]
Connection establishment delay of mobile originated calls	< 8.5 s (95%), 10 s (99 %)	Emergency call < 1 s (90%), 2 s (100%) Broadcasting < 1 s (90%), 2.5 s (100%) Group communications < 1 s (90%), 2.5 s (100%) Others < 3.5 s (90%), 5 s (100%)
Connection establishment failure probability (per attempt)	10^{-2}	10^{-2}
Maximum end-to-end delay	500 ms (99%)	250 ms (99%)
Data rates (voice)	2.4, 4.8, and 9.6 kb/s	6.4~23.85 kb/s
Data rate (packet)	≤ 59.2 kb/s (EDGE)	Static: DL ≥ 11 Mb/s, UL ≥ 4 Mb/s Moving: DL ≥ 4 Mb/s, UL ≥ 3 Mb/s
Connection loss rate	< 10^{-2}/h	< 10^{-2}/h
Maximum break during handover	500 ms	60 ms
Network registration time	≤ 30 s (95%), ≤ 35 s (99%), ≤ 40 s (100%)	10 s (100%)

Table 2. Performance requirements for GSM-R and LTE-R.



Test case	Criterion	Result (iterations)	Test case	Criterion	Result (iterations)
Network registration	Elapsed time < 10 s	0: < 1 s (102), M: < 2 s (110)	Conference call connectivity	Successful rate > 99%	M: 100% (304)
Basic call connectivity	Successful rate > 99%	0: < 100% (309), M: 100% (304)		95% setup time < 3.5 s	M: < 1 (331)
	95% setup time < 3.5 s	0: < 1 (309), M: < 1 (304)		100% setup time < 5 s	M: < 1 (331)
	100% setup time < 5 s	0: < 1 (309), M: < 1 (304)	FTP throughput	Static DL: 0 > 11 Mb/s, M > 3 Mb/s	0: 24 Mb/s, M: 17 Mb/s
Call long duration	Call drop rate < 0.01 drop/h	0 (100 h)		Static UL: 0 > 4 Mb/s, M > 2 Mb/s	0: 10 Mb/s, M: 9.8 Mb/s
	Handover delay < 60 ms	< 35 ms (231)		Moving DL: 0 > 4 Mb/s, M > 2 Mb/s	0: 17 Mb/s, M: 17 Mb/s
	Handover successful rate	100% (231)		Moving UL: 0 > 3 Mb/s, M > 1 Mb/s	0: 9.8 Mb/s, M: 9.4 Mb/s
Emergency call connectivity	Successful rate > 99%	O: 100% (305), M: 100% (301)	Video call connectivity	Successful rate > 99%	O: 100% (309), M: 100% (309)
	95% setup time < 1 s	O: < 1 (305), M: < 1 (301)		95% setup time < 3.5 s	O: < 1.5 s (309), M: < 1.5 s (304)
	100% setup time < 2 s	O: < 1 (305), M: < 1 (301)		100% setup time < 5 s	O: < 1.5 s (309), M: < 1.5 s (304)
PIT call connectivity	Successful rate > 99%	O: 100% (320), M: 100% (304)	Ping data transfer	99% RTT < 500 ms	350 ms (10,000)
	95% setup time < 3.5 s	O: < 1 (320), M: < 1 (304)		Consecutive loss duration < 5 s	None (10,000)
	100% setup time < 5 s	O: < 1 (320), M: < 1 (304)			

Figure 4. Performance results of LTE-R.

In addition, it is observed that onboard terminals perform better than mobile terminals. This is because the received signal in a mobile terminal is more degraded than that in an onboard terminal due to less straightness of the wireless channel, which results in lower SINR than in the onboard terminal. Note that the onboard terminal got a far higher score than the mobile terminal in terms of FTP throughput and network registration since the onboard terminal requires exchange of more packets on the control and data planes than the mobile terminal.

Potential of LTE-R for Public Safety

To gain insight into the extensions from LTE-R services to public safety LTE (PS-LTE) services, we need to take into account a comparison between PS-LTE and LTE-R. LTE-R supports most of the PS-LTE functionalities regarding group communications except for device-to-device-communication-based functions, which are in the midst of lively discussion in 3GPP standard activities [13].

In group communications for PS, priority and preemption are indispensable functionalities, which are feasible with LTE-R and PS-LTE. Specifically, each set of different users and service is provided with differentiated priority via PCRF. In the PCRF, allocation retention priority (ARP) and the QoS class identifier (QCI) are the main parameters that steer the priority. Here, ARP is associated with deciding which bearer request should be accepted on the congested network, and the QCI parameter is used to prioritize among packets in the limited radio resource situation. Hence, LTE-R together with PS-LTE attains functionalities for priority and preemption. We observed that from performance test results through the implemented LTE-R testbed, various types of group communications such as basic calls, emergency calls, PTT calls, and conference calls, together with data transmission, can be supported with an acceptable connection setup time and access rate.

We observed that from performance test results through the implemented LTE-R test-bed, various types of group communications such as basic calls, emergency calls, PTT calls, and conference calls, together with data transmission, can be supported with an acceptable connection setup time and access rate.

Hence, LTE-R can be further developed not only for integrated wireless railway services but also for PS by keeping the principles that *core technologies of railway communications should be aligned with those of mobile communications.*

Discussion

In this section, we introduce several critical issues found via cell deployment and performance evaluation that need further and deeper technical consideration.

Cell Planning and Optimization: The most critical issue during cell optimization is the achievement of a target level of received signal strength for both onboard terminals and mobile terminals in the cabin. The antenna of an onboard terminal is usually placed on the top of the train, which is much higher than mobile terminals, which are at human height. Furthermore, a mobile terminal in a cabin receives a wireless signal that comes through the windows of the train, and this matters for cell optimization if the RU uses a directional antenna, as is common in railway environments. The antenna parameters should therefore be carefully determined so that both onboard and mobile terminals ensure a certain level of signal quality simultaneously.

Network Design: In essence, railway communications rely on deterministic train movement with regular speed patterns. For example, trains tend to have low speed near stations. In addition, railway communication networks use a cell structure of a sequential chain type, while commercial networks typically use a cell structure of a hexagonal type. Because of the uniqueness of a railway network, the communication scheme can be optimized further with low complexity and without loss of optimality [14].

Guaranteeing QoS for Safety Services: Some railway services, such as train control and railway emergency calls, are strictly required to guarantee their own QoS. This is because any service outage can threaten human safety by causing accidents, or lead to delays and inherent financial losses for operators. Train control services usually generate small amounts of traffic, but each traffic packet must be transferred with low latency at any time. Therefore, in order to support these services, it is important to keep data connectivity at the highest priority level. To do so, LTE-R must make use of sophisticated network management schemes for handling exceptional situations such as traffic congestion.

High Speed: One of the most challenging issues is connected with the high speeds of trains. Communication theory suggests that system performance degradation may be expected due to the decrease in coherence time caused by the increase in Doppler shift. For reliable communications, carrier frequency and train speed should be considered jointly so that the estimated channel profile applies during the time between two different reference symbols in an LTE subframe. An alternative approach would be to use the predictable mobility pattern of terminals. Terminals can perform measurement or handover more effectively and precisely when using information about their direction and speed of motion.

Position Assisted System: In a railway communications system, the position of a train is detected using a sensor so that the train position information can be used jointly with communication-related information. Based on the acquired train position, serving cell management is done more correctly, which effectively leads to seamless connections of voice, video, and data communications.

Accommodation of Numerous Antennas: In general, the number of antennas is limited due to the space required by a mobile terminal. However, it is possible to accommodate a number of antennas in order to make use of the vast areas on top of trains so that additional diversity and multiplexing gains of wireless communications can be obtained simultaneously. Note that there is still a trade-off between complexity and performance. Specifically, reliable and efficient communications with a high data rate usually requires a great deal of channel state information, which is challenging for high-speed trains.

Conclusion

As a feasibility proof for integrated wireless railway communication networks for reliable train services, we have evaluated the performance of an implemented LTE-R system. For performance validation, we constructed an LTE-R testbed by achieving cell deployment in a commercial railway. We have demonstrated that our LTE-R testbed was validated via basic performance tests together with a variety of group call tests. In light of LTE-R's performance and also the standardization movement toward mission-critical group communications, LTE-R shows great potential as a communication technology for both integrated railway communications and public safety. For further evolution of LTE-R, we have discussed unique features of railway communication networks with technical issues of particular interest.

In light of LTE-R's performance and also the standardization movement toward mission critical group communications, LTE-R shows great potential as a communication technology both for integrated railway communications and for PS.

Acknowledgment

This work was partly supported by ICT R&D program of MSIP/IITP [B0101-15-1361, Development of PS-LTE System and Terminal for National Public Safety Service], and partly supported by Future Metropolitan Rail Technique R&D program of MOLIT [10PURT-B056851-01, Study on Wireless Network for Railway Application].

References

[1] E. Dahlman, S. Parkvall, and J. Sköld, *4G LTE/LTE-Advanced for Mobile Broadband*, 2nd ed., Academic Press, 2014.

[2] C.-X. Wang *et al*., "Cellular Architecture and Key Technologies for 5G Wireless Communication Networks," *IEEE Commun. Mag*., vol. 52, no. 2, Feb. 2014, pp. 186–95.

[3] ERTMS SUBSET-093 v2.3.0, GSM-R Interfaces: Class 1 Requirements, 2005.

[4] UIC CODE 950 v7.3.0, EIRENE Functional Requirements Specification, 2012.

[5] Banedanmark, "Boundaries between ETCS and the GSM-R Network," 2008.

[6] Banedanmark, "Requirements on the GSM-R Network for ETCS support," 2009.

[7] W. Luo, R. Zhang, and X. Fang, "A CoMP Soft Handover Scheme for LTE Systems in High Speed Railway," *EURASIP J. Wireless Commun. and Net.*, vol. 2012, no. 1, June 2012, pp. 1–9.

[8] A. Sniady and J. Soler, "LTE for Railways: Impact on Performance of ETCS Railway Signaling," *IEEE Vehic. Tech. Mag.*, vol. 9, no. 2, Apr. 2014, pp. 69–77.

[9] A. Sniady *et al*., "LTE Micro-Cell Deployment for High-Density Railway Areas," *Commun. Tech. Vehicles Proc*., Springer, 2014, pp. 143–55.

[10] UIC, GSM-R Procurement & Implementation Guide, 2014.

[11] TTA TTAK.KO-06.0369, Functional Requirements for LTE Based Railway Communication Systems, 2014.

[12] M. Alasti *et al*., "Quality of Service in WiMAX and LTE Networks," *IEEE Commun. Mag*., vol. 48, no. 5, May 2010, pp. 104–11.

[13] 3GPP TS 23.303 V12.4.0, Tech. Spec. Group Services and System Aspects; Proximity-Based Services (ProSe); Stage 2, 2015.

[14] H. Maleki and S. A. Jafar, "Optimality of Orthogonal Access for One-Dimensional Convex Celluar Networks," *IEEE Commun. Lett*., vol. 17, no. 9, Sept. 2013, pp. 1770–73.

Additional Reading

[1] 3GPP TS 36.104 V12.7.0, Tech. Spec. Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception, 2015.

Biographies

Yong-Soo Song (adair@krri.re.kr) received his Master's degree in electrical engineering from Yonsei University in 2004. He has been with KRRI since 2004. He is working toward his Ph.D in electrical engineering from Yonsei University.

Juyeop Kim (jykim00@krri.re.kr) is a senior researcher in the ICT Convergence Team at Korea Railroad Research Institute (KRRI). He received his M.S. and Ph.D. in electrical engineering and computer science from Korea Advanced Institute of Science and Technology (KAIST) in 2010. His current research interests are railway communications systems, group communications, and mission-critical communications.

Sang Won Choi (swchoi@krri.re.kr) received his M.S. and Ph.D. in electrical engineering and computer science from KAIST in 2004 and 2010, respectively. He is currently a senior researcher in the ICT Convergence Research Team of KRRI. His research interests include mission-critical communications, mobile communication, communication signal processing, and multi-user information theory. He was the recipient of a Silver Prize at the Samsung Humantech Paper Contest in 2010.

Yong-Kyu Kim (ygkim1@krri.re.kr) received his M.S. in electronic engineering from Dankook University, Korea, in 1987, and his D.E.A. and Ph.D. in automatic and digital signal processing from Institute National Polytechnique de Lorraine, France, in 1993 and 1997, respectively. He is currently an executive researcher inthe ICT Convergence Team at KRRI. His research interests are in automatic train control, communication-based train control, and driverless train operation.

Landing on the Mobile Web: From Browsing to Long–Term Modeling

Troy Johnson and Patrick Seeling

As caching frequently used data locally is a common initial approach employed to limit network traffic and energy expenditures while "on the go," the authors evaluate the long-term suitability of approximating a basic set of parameters for a cache and request behavior model using a popular large dataset. They present a convenient approach that can employ a general approximation of parameters over time.

ABSTRACT

Browsing the web has become a common task performed using personal mobile devices, resulting in significant access network and battery limitation challenges. Efforts to alleviate these challenges are commonly based around approaches incorporating elements of on-device and network optimizations. Energy-efficient mobile web content delivery has, in turn, attracted a significant body of research and practical developments. However, the efforts put forth today might not result in long-term applicable results should the underlying characteristics of the mobile content change drastically over time. As caching frequently used data locally is a common initial approach employed to limit network traffic and energy expenditures while "on the go," we evaluate the long-term suitability of approximating a basic set of parameters for a cache and request behavior model using a popular large data set. We present a convenient approach that can employ a general approximation of parameters over time. Our long-term modeling of the underlying factors results in an acceptable level of peak inaccuracies in simulations for more than a year's time horizon. In turn, practitioners and researchers are enabled to readily employ modeling and simulation approaches over a significant period of time with only slight impacts on their approaches and results.

INTRODUCTION

Mobile users oftentimes need to quickly access information that is provided using the web, rather than in an application-wrapped context. With web access predictions indicating that browsing will be predominantly performed by users employing their personal mobile devices, a significant portion of future network traffic will be delivered over the wireless air interfaces of non-stationary personal handheld devices [1]. Other mobile user habits and a (mobile) web-first strategy for new service developments might additionally contribute to the frequency and amounts of data that mobile users request from service landing pages. Additionally, the increased pressure to create applications that can be spread across mobile platforms through the use of web-based technologies, such as HTML5 and JavaScript, increases the likelihood of future applications employing a hybrid strategy of pre-packaged content that is supplemented by requesting specific landing pages from the web. Similarly, if new web-centric services emerge (or new information about an existing service is sought by users, e.g., by sharing information about them in social contexts), mobile users might be reluctant to engage in the process of installing an application to their device and prefer to try out the web-based version of the service.

Based on this emerging trajectory, we initially presented a comparative overview of the landing web page characteristics differences between desktop and mobile versions in [2], which complements observations made in [3]. As web pages are constituted by a typically large range of individual objects that need to be retrieved, the authors of [4] demonstrate the possibility of energy-optimized mobile web browsing. Circumventing the need to download individual web objects, caching performed by (mobile) browsers commonly can be considered a first optimization step for improvement (e.g., as outlined in [5]). While some limits exist for individual devices, multi-level caching approaches could yield additional benefits, as recently described in [6]. An interesting approach discussed therein is the notion of decentralized caching, an intuitive example of which is to exchange content between different device types belonging to the same user to save energy (from the user and content provider perspectives) by essentially generating a locally shared cache. The differences as well as similarities between the different versions of web pages inherently offer possibilities for optimization efforts. Information about cache lifetime expirations of web objects in general could be exploited to locally forward content in an opportune fashion; the potential of this approach was evaluated in [7] using a popular user browsing behavior model. The integration of social networking concepts when considering today's mobile shared economy trajectories [8] increases the importance of providing web-based content to mobile users and was outlined in [9]. Here, the authors exploit the significant improvements in mobile resources as well as social and other contexts to proactively deliver content off-peak, thus reducing the "network crunch."

Research efforts, for example, striving to optimize energy-efficient mobile web content provisioning rely on modeling of the underlying

The authors are with Central Michigan University. Patrick Seeling is the corresponding author.

0163-6804/16/$25.00 © 2016 IEEE

content characteristics. As the vast parameter space is typically prohibitive for experimentation, the models subsequently drive simulations for approximations of the mobile web characteristics. To date, a continuous stream of efforts continues to capture the state of the web as delivered to mobile devices. Typically, resulting models include a combination of states and/or multiple distributions to take user behavior and content characteristics into account. Comparing exemplary models developed over time, such as [10] in 2011, [11] in 2012, or [12] in 2014, emerging differences of individual model parameters can be observed. This trend is indicative of the majority of models representing snapshots in time. However, long-term trends could have a significant impact on the suitability of efforts made today that employ characteristics of actual content found on the web. A major drawback that is, in turn, inherent to the presentation of developed models is the limited evaluation of suitability over time. Given the overarching trends presented in [2], here we provide an evaluation of an accessible long-term approach to modeling the mobile landing web page characteristics with respect to cache implications and resulting network requests. Employing the large httparchive.org data set [13], we employ a model that is able to capture the main facets of the cache behavior a mobile user's browser would encounter over the course of an unconnected time as presented in [14], and evaluate its suitability for longer time horizons by predictions of its parameters.

The remainder of this article is structured as follows. In the subsequent section, we provide an overview of the typical mobile device web cache's content characteristics based on prior works. We continue by describing a convenient methodology for simulating a mobile browser's means of accessing the web content represented by web landing pages. Then we evaluate the extrapolation of past observations for the simulation of mobile device web access characteristics with caching, specifically considering long-term trends and implications. We conclude in the final section.

A Mobile Browser's Cache

Initially, we consider the contents and the overall simulation of mobile browsing and subsequent cache utilization in this section.

Cache Contents

The contents that constitute a mobile browser's assumed cache can be evaluated by considering its composition first. To this extent, the common web landing pages of popular web sites were found to provide an adequate upper limit for complexity considerations [3]. We provide the cache contents as identified in [2] for these landing pages in Fig. 1. We initially note that the number of objects per mobile web page increases at several time instances, commonly coinciding with an update of the underlying httparchive data set. Within each updated sequence, however, the number is overall at a stable level, which reached around 60 objects in October 2013.

For the distribution of the object sizes by type that make up a web page, we note in Fig. 1b that a general upward trend exists for most object types; they are more independent of the time period than the number of objects in comparison. Overall, HTML content exhibits a steady trend, while all other types of objects commonly embedded feature a steady increase in relative size. Combined with the number of objects per page, this results in a continuous increase of mobile web landing page sizes.

A mobile browser requesting these pages would in turn be required to download a significant amount of this data were general cache expiration times low. Indeed, when considering the categorized object expirations illustrated in Fig. 1c for October 15, 2013, we note that a significant portion of objects immediately expire or are very short lived. Further discussed in [2], these results are fairly stable over time, which makes the distribution of expiration ages and related objects another rather stable component to consider. Thus, a mobile browser's cache would have to expire a significant portion of data, resulting in required re-downloads (with subsequent impacts on air interface use and resulting battery utilization).

Overall, these trends motivate us to consider a more stable distribution of the page compositions of objects and their expirations, with a focus on size trends over time, as discussed now in greater detail.

Simulating a Mobile Browser's Cache

We developed a model that captures the main facets of the web object behavior that a mobile user's browser would encounter by separating the sizes and expiration age limitations for web page objects, as little correlation is exhibited among them [14]. We note that we consider only objects with cache expirations below one week, as those above can be considered static and pre-provisioned in the context of mobile web browsing. Employing this approach, synthetic web pages are generated in a multi-step process. Initially, a page is generated to be of either a current, short-lived, medium-lived, or regular type. Based on this selection, the number of objects for the page is drawn randomly, followed by the sizes of the objects. Subsequently, the object expirations are randomly distributed over the objects based on the synthetic page type under which they fall. Finally, for a set of generated web pages, a unique popularity index is randomly assigned to each page (as little correlation was found for page compositions and their popularity ranking in the underlying data set). 500 sets of pages are generated to create data sets to be used in a browsing simulation that takes user behavior into account.

We consider the model presented in [12] as a baseline and employ a modified version as described in [15], maintaining the main facets of user behavior modeling. We illustrate the overall model used for simulating a mobile browsing experience in Fig. 2.

While the user is in a browsing session, assumed to last a specific amount of time that is Weibull distributed with $d_B \sim WEB(\alpha = 0.4, \beta = 80)$, requests to (different) synthetically generated web pages follow the popular Zipf distribution for content popularity $z_i \sim Zipf(\alpha = 0.85)$.

The contents that constitute a mobile browser's assumed cache can be evaluated by considering its composition first. To this extent, the common web landing pages of popular web sites were found to provide an adequate upper limit for complexity considerations.

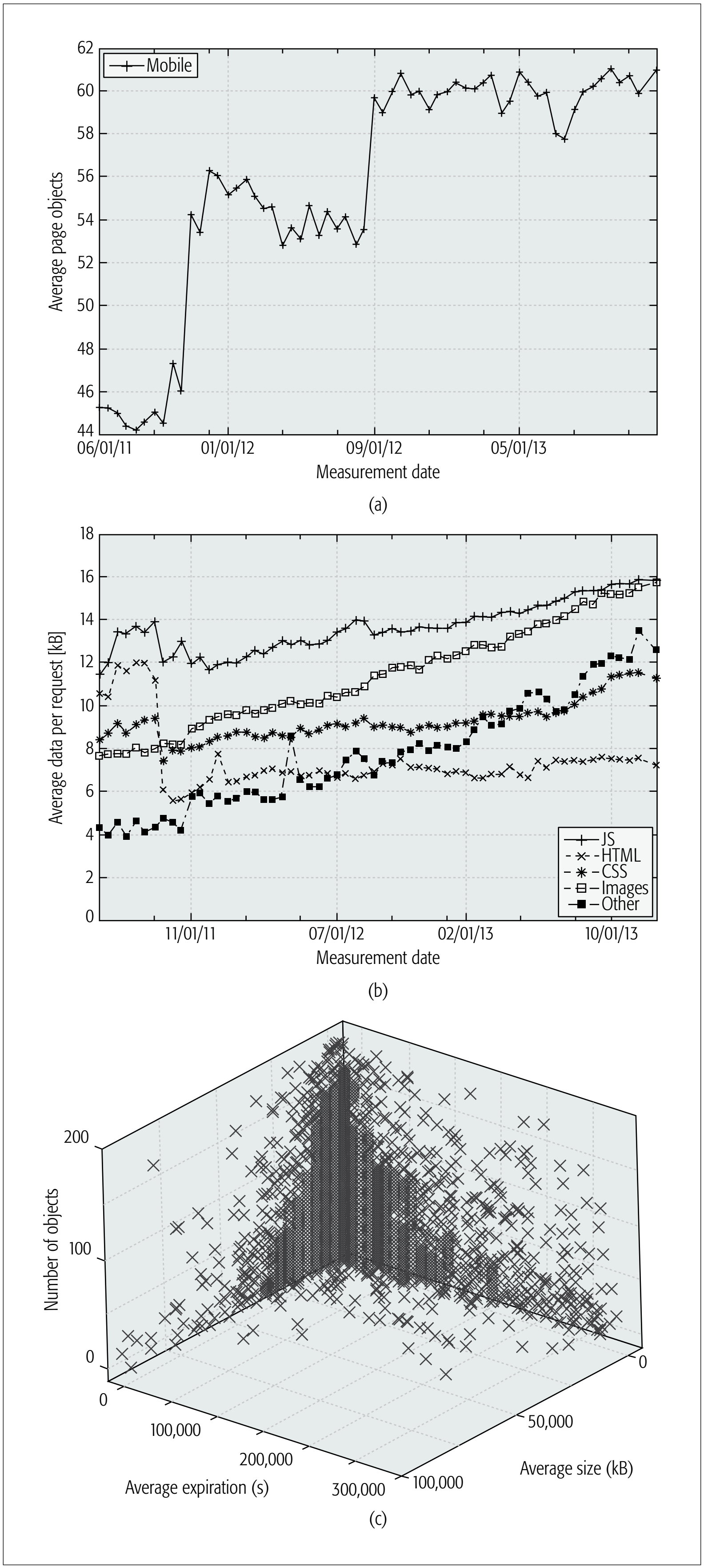


Figure 1. Web object numbers, sizes by types, and expirations over time encountered for the httparchive.org data set; see [2, 14] for more details: a) average number of objects per page; b) object types and sizes; c) object expirations for October 15, 2013.

The mobile user is assumed to continue browsing after a Pareto-distributed waiting time in seconds [16] for viewing the page (also commonly referred to as user think time, UTT) has passed, $d_w \sim Pareto(\beta = 1.5, k = 30)$. Alternatively, the user could end the currently active browsing session with P_{BW} of 40 percent, in which case the overall system enters a waiting state for a period of time. The duration of the waiting state is exponentially distributed with $d_w \sim Exp(\lambda = 0.05)$. We refer the interested reader to [15] for a more thorough description of the simulation of the employed mobile user's web browsing behavior.

Initial Simulations of Cache Benefits

We initially employ our simulation model for the approximation of the October 15, 2013 data set's characteristics, starting with an assumed *hot* cache (i.e., the cache is assumed to be fully populated at the beginning of the simulation). Here, we evaluate the overall suitability to determine the relative amounts of objects and associated amounts of data that would require a download instead of being serviceable from the mobile device browser's local cache. We perform 500 synthetic web page data set generations and simulate the mobile browsing over each of these sets 500 times for a duration of three days.

We present the resulting graphic comparison in Fig. 3, whereby we omit confidence intervals for readability purposes (noting that they are typically within 5 percent of the presented averages).

With a significant portion of the objects expiring instantly or almost instantly, we initially note that even for just a very short time offset, a significant amount of data would require downloading. Next, we observe that there is a "kink" in the relative amounts of data requests around a day's border, coinciding with a common web object expiration value on the shorter timescale observable from Fig. 1. We additionally note that our model follows this trend, but slightly overestimates the download requirements before the time boundary of a day, while for the remaining time, the model slightly underestimates the number of objects to require a download. While the number of bytes resulting from the actual data also exhibits a "kink," employing the simulated web pages does result in a slightly increased margin of estimated downloads below a day's threshold, which narrows as the simulated time increases. Overall, however, we note that our model is within approximately five percent of the actual data simulation.

Approximating the Cache's Content over Time

In this section, we consider an approximation of the long-term trends exhibited in prior works and combine those with the simulations of page-level popularities, number of objects, and their characteristics (sizes and expirations).

Parameter Estimations for Long Horizons

Following the simulation approach outlined above, the entire set of object sizes (i.e., individual web objects not aggregated based on the page to which they belong) is approximated using a Weibull distribution before the objects are aggregated to determine the synthetically generated page level characteristics. The number of responses per page, in turn, is approximated by employing a gamma distribution with fixed parameters for each month, effectively plateauing the current level of page complexities for the purpose of our approach; this reflects the general trend observed within the underlying data sets.

We anchor our evaluation around the October 15, 2013 data set as a base to extrapolate the long-term estimation of the object sizes only, as their overarching characteristics have little correlation. Continuing the trend of a bi-weekly basis as provided from the underlying httparchive.com data sets, we illustrate the resulting shape and scale parameters for the Weibull distribution employed to generate synthetic object sizes in Fig. 4.

We observe a somewhat jittery behavior for shape and scale parameters over time. However, the deviations are all within fairly narrow confines, allowing us to perform a basic approximation of their trends over time. Implementing a basic linear fitting to this set of parameter variability, we note an increase of approximately 1.4 percent each half month for the scale parameter. For the shape parameter, we observe a slower increase of 0.05 percent each half month. In the next steps, we now consider employing these parameter estimates within our model as a long-term trend for the synthetically generated object sizes reaching beyond the time horizon of the parameter estimations.

Impact on Long Horizons

Initially, we commence the simulation with a hot cache, that is, we assume all objects are locally available. We subsequently continue with an evaluation period of up to three days, for which we generate 500 sets of artificially composed web pages. The number of individual web pages in each of the generated sets was chosen to match the numbers in the original source data for the individual months (ranging from 4803 pages for June 15, 2013 to 4713 pages for January 15, 2015).

For each individual web page, the characteristics of page compositions out of objects and object attributes are generated according to our previous modeling, described in greater detail in [14], with parameters of individual distributions replaced by their long-time approximations. Since no strong correlation was found to exist between page rank (popularity) and other characteristics, we randomize the popularity of synthetically generated pages employing the uniform distribution. To account for the impact that the popularity variability can have on the simulated user browsing behavior, each set of artificially generated web pages (500 sets) were employed in 500 simulations (for a combined total of 250,000 simulations at each artificially generated monthly data point). The results were captured in terms of bins with sizes of 30 s, 300 s, and 3600 s for processing. To determine the impact that the approximation of values has on the quality of the simulation, we determine the difference between the simulations employing the approximated parameters and the actual httparchive data set, illustrated in Fig. 5a. We note that for readability purposes, the results for the 30 s bin sizes were smoothed after generation of the data. We additionally note that for readability purposes, we do not illustrate the upper and lower confidence intervals, as they are very narrow (typically within 5 percent of the averages presented).

For the impact the approximation has on object requests, we initially evaluate the fine-grained 30 s bin results. We observe an initial increase in the difference, which is followed by a plateau or more pronounced peak, and succeeded by a slower decrease over simulated time observing the cache behavior. Comparing the different snapshots in time for which the comparison was performed, we note that the further the simulation using approximations moves away from the anchor point, the higher the impact of approximating the parameters for the web page composition.

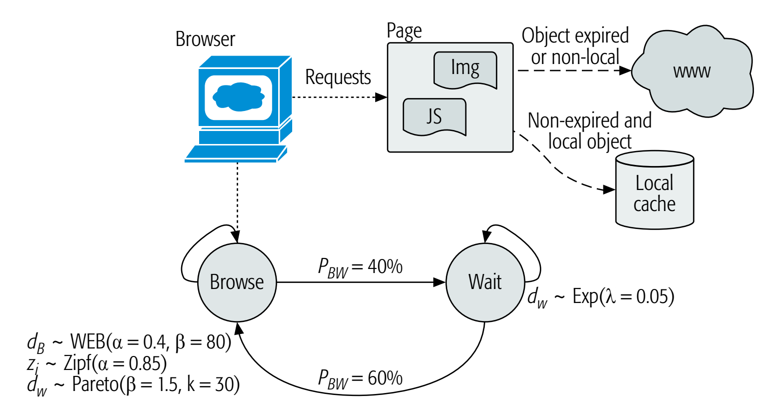


Figure 2. Overview of the assumed user browsing behavior and content retrieval from local cache or the Internet.

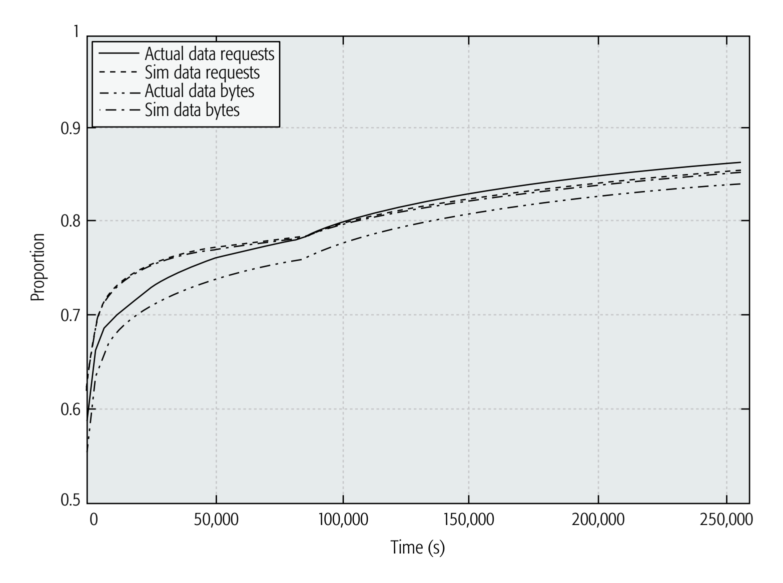


Figure 3. Overview of the simulated mobile browser required external requests based on expired cache contents employing the underlying original httparchive and synthetically generated mobile web landing pages.

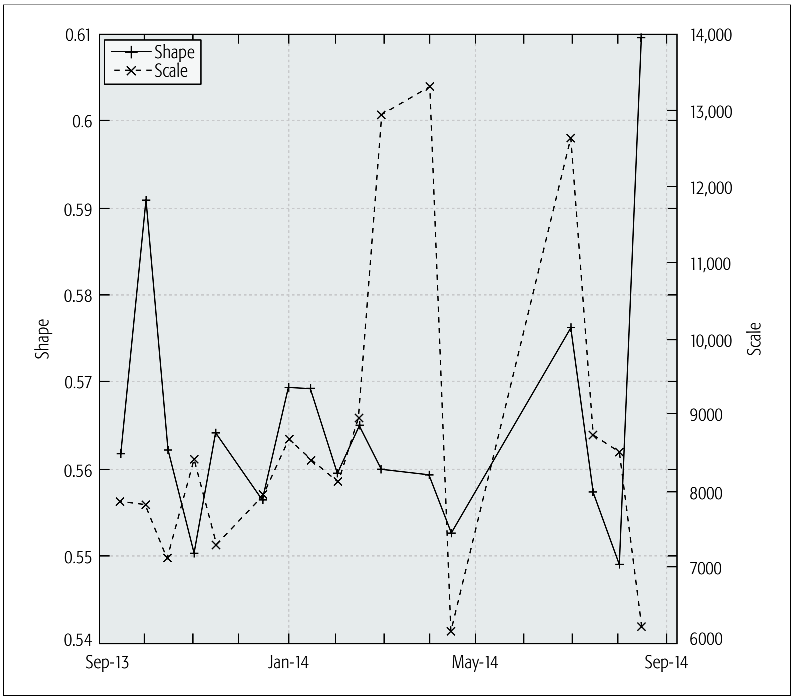


Figure 4. Estimation of shape and scale parameters for the Weibull distribution used to approximate the overall web object sizes of the data set.

Overall, however, the level of difference is very narrow (below 5 percent) for the highest difference.

Next, we shift our view to the approximation that includes the sizes at fine granularity in Fig. 5b. We observe an immediate rise in the differences between the two approaches, followed by a distinct peak, which is trailed by an asymptotic decay. Upon closer inspection of the different time distances from the October 15, 2013 data set used as the base, we note that here, a more distinct order can be found than for the number of requests alone. Specifically, the further away the approximation becomes, the higher the differences, as can be expected for such an approach. It is noteworthy, however, that even for a time distance of more than one year from the last point used for estimation of the parameter development over time, our model is still only 13 percent off in its peak difference.

We can attribute the additional increase in the difference between approximated and actual parameters used in the two simulation approaches to the additional variability that stems from the simulations of object sizes that constitute individual simulated web pages in the cache. This effect has less impact on the simulation times beyond a day, for which even the furthest evaluated approximation yields less than 8 percent difference of the data in the simulated cache.

As different simulation scenarios might require different time horizons, we aggregate the results into larger different bin sizes of 300 s and 3600 s, illustrated in Fig. 5 as well. We observe from comparing Figs. 5a, c, and e for the number of cached objects and Figs. 5b, d, and f for the amount of data in the cache, respectively, that few characteristic behavior changes can be observed. We note, however, that the increase in the bin size to 1 h results in the next close approximation for March 15, 2014 exhibits significantly lower differences. Thus, when considering smoothed or capacity-level aggregations for simulations, the averaging effects have a significant effect on the result accuracy.

Conclusion

In summary, our approach shows great promise to facilitate the simulation of mobile web landing pages, which can be regarded as reflecting the upper-bound characteristics of actual web pages visited by mobile users. In particular, it enables the evaluation of how a mobile web browser could perform realistic web requests over time in the presence of its local cache. The continuous evolution of the mobile web's characteristics from human interface and interaction design principles are embodied into the web objects to be retrieved by a browser.

Over time, it might not always be possible to predict or derive up-to-date representations of the latest developments to include those in performance evaluations of new mechanisms in the realm of mobile communications. We present an accessible approach that can employ a general approximation of its parameters over time with only manageable impact on accuracy. Specifically, for more than one year's changes in the underlying data set, our model incurs only a maximum 13 percent penalty in quality. While this indicates that a long-term prediction can be facilitated by our model with reasonable accuracy, it also indicates a sensible result quality of performance evaluations for long time horizons can be achieved by other researchers when approaches operate on the granularity of web page objects.

References

[1] Cisco, Inc., "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2014–2019," tech. rep., Feb. 2015.

[2] T. A. Johnson and P. Seeling, "Desktop and Mobile Web Page Comparison: Characteristics, Trends, and Implications," *IEEE Commun. Mag.*, vol. 52, no. 9, Sept. 2014, pp. 144–51.

[3] M. Butkiewicz, H. Madhyastha, and V. Sekar, "Characterizing Web Page Complexity and its Impact," *IEEE/ACM Trans. Netw.*, vol. 22, no. 3, June 2013, pp. 943–56.

[4] B. Zhao *et al.*, "Energy-Aware Web Browsing on Smartphones," *IEEE Trans. Parallel Distrib. Sys.*, vol. 26, no. 3, Mar. 2015, pp. 761–74.

[5] H. Shen *et al.*, "Energy-Efficient Data Caching and Prefetching for Mobile Devices Based on Utility," *Mobile Networks and Applications*, vol. 10, no. 4, Aug. 2005, pp. 475–86.

[6] M. A. Maddah-Ali and U. Niesen, "Fundamental Limits of Caching," *IEEE Trans. Info. Theory*, vol. 60, no. 5, May 2014, pp. 2856–67.

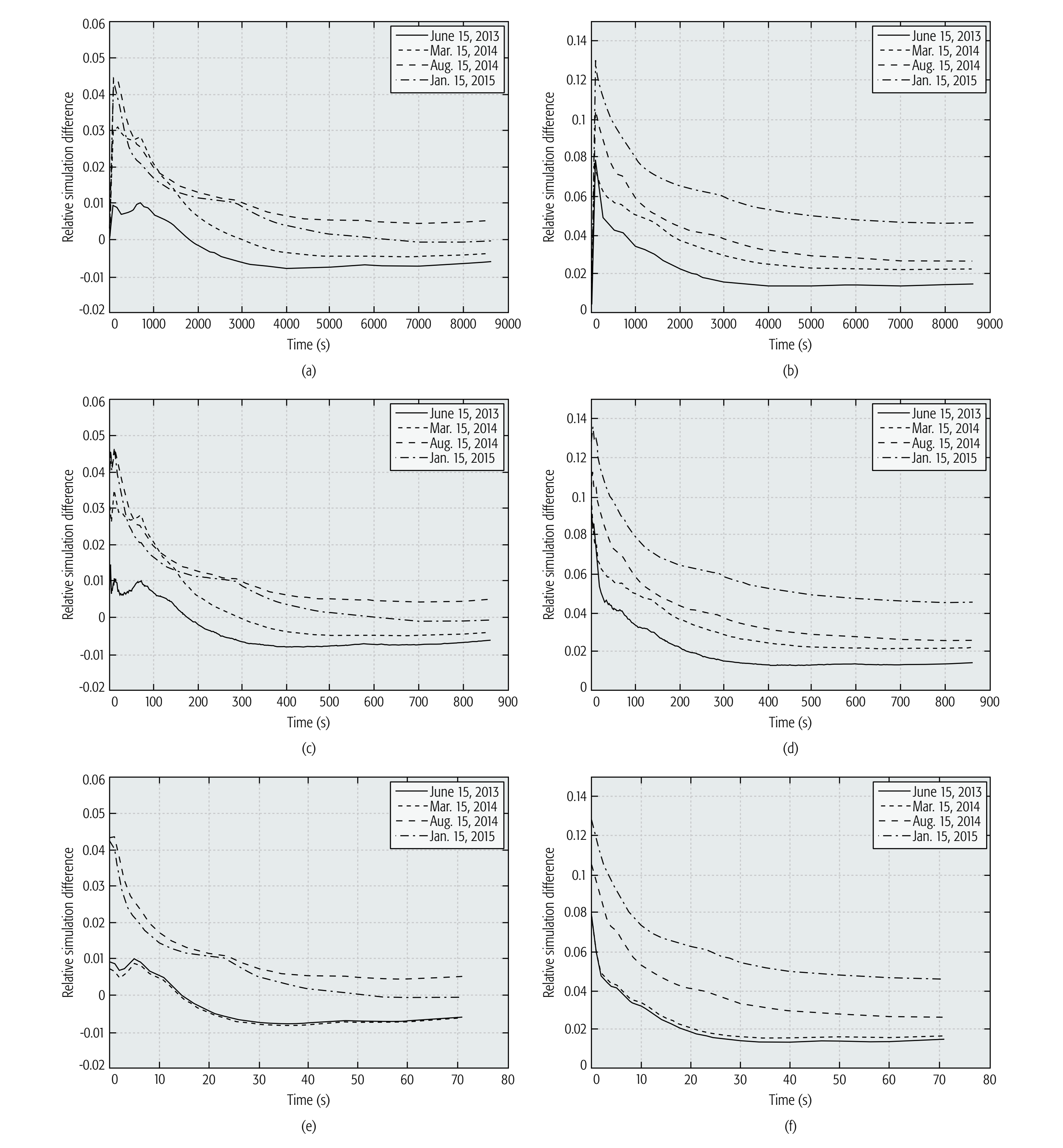
[7] T. A. Johnson and P. Seeling, "Web Cache Object Forwarding from Desktop to Mobile for Energy Consumption Optimizations," *Proc. IEEE OnlineGreencomm*, Tucson, AZ, USA, Nov. 2014, pp. 1–7.

[8] M. Katz, *et al.*, "Sharing Resources Locally and Widely: Mobile Clouds as the Building Blocks of the Shareconomy," *IEEE Vehic. Tech. Mag.*, vol. 9, no. 3, Sept. 2014, pp. 63–71.

[9] E. Bastug, M. Bennis, and M. Debbah, "Living on the Edge: The Role of Proactive Caching in 5G Wireless Networks," *IEEE Commun. Mag.*, vol. 52, no. 8, 2014, pp. 82–89.

[10] G. Feng Zhao *et al.*, "Modeling Web Browsing on Mobile Internet," *IEEE Commun. Lett.*, vol. 15, no. 10, Oct. 2011, pp. 1081–83.

[11] R. Pries, Z. Magyari, and P. Tran-Gia, "An HTTP Web Traffic Model Based on the Top One Million Visited Web Pages," *Proc. EURO-NGI Conf. Next Generation Internet*, June 2012, Karlskrona, Sweden, pp. 133–39.



(a)

(b)

(c)

(d)

(e)

(f)

Figure 5. Differences between the simulations employing the synthetically generated data sets with linearly approximated parameters for long-term evaluations and the underlying httparchive data set: a) 30 s bins, requests; b) 30 s bins, data; c) 300 s bins, requests; d) 300 s bins, data; e) 3600 s bins, requests; f) 3600 s bins, data.

[12] I. Tsompanidis, A. H. Zahran, and C. J. Sreenan, "Mobile Network Traffic: A User Behavior Model," *Proc. 7th IFIP Wireless and Mobile Networking Conf.*, Vilamoura, Algarve, Portugal, May 2014, pp. 1–8.

[13] S. Souders, httparchive.org, Apr. 2015; http://www.httparchive.org

[14] T. Johnson and P. Seeling, "Landing Page Characteristics Model for Mobile Web Performance Evaluations on Object and Page Levels," *Proc. IEEE ICC*, London, U.K., June 2015.

[15] T. Johnson and P. Seeling, "Browsing the Mobile Web: Device, Small Cell, and Distributed Mobile Caches," *Proc. IEEE ICC Wksp. Cooperative and Cognitive Networks*, London, U.K., June 2015.

[16] G. Anastasi *et al.*, "Performance Comparison of Power–Saving Strategies for Mobile Web Access," *Performance Evaluation*, vol. 53, no. 3–4, Aug. 2003, pp. 273–94.

Biographies

Troy Johnson (johns4ta@cmich.edu) is currently with Ford Motor Company, Dearborn, Michigan. He received his B.S. in 2013 and his M.S. in 2015 from Central Michigan University, both in computer science. His main research interests revolve around mobile and green computing.

Patrick Seeling [S'03, M'06, SM'11] (pseeling@ieee.org) is an associate professor in the Department of Computer Science at Central Michigan University. He received his Ph.D. in electrical engineering from Arizona State University, Tempe, in 2005. His research interests include mobile and multimedia networking mechanisms and their optimization, quality of service and resulting impacts on quality of experience, and computer-mediated education. He is a Senior Member of the ACM.

An Overview of the CPRI Specification and Its Application to C-RAN-Based LTE Scenarios

Antonio de la Oliva, José Alberto Hernández, David Larrabeiti, and Arturo Azcorra

The authors present the CPRI specification, its concept, design, and interfaces, provide a use case for fronthaul dimensioning in a realistic LTE scenario, and propose interesting open research challenges in the next-generation 5G mobile network.

Abstract

The CPRI specification has been introduced to enable the communication between radio equipment and radio equipment controllers, and is of particular interest for mobile operators willing to deploy their networks following the novel cloud radio access network approach. In such a case, CPRI provides an interface for the interconnection of remote radio heads with a baseband unit by means of the so-called fronthaul network. This article presents the CPRI specification, its concept, design, and interfaces, provides a use case for fronthaul dimensioning in a realistic LTE scenario, and proposes some interesting open research challenges in the next-generation 5G mobile network.

Introduction and Motivation

Mobile network operators (MNOs) have realized that the cloud radio access network (C-RAN) approach can provide a significant advantage with respect to their competitors in a market scenario where the trend in revenue per user is almost flat or decreasing. C-RAN has been recently introduced and further shown that significant operational expenditure (OPEX) and capital expenditure (CAPEX) reductions can be achieved with respect to traditional equipment deployments. A recent trial from China Mobile has shown 53 and 30 percent savings in OPEX and CAPEX, respectively [1].

The C-RAN approach advocates for the separation of the radio elements of the base station (called remote radio heads, RRHs) from the elements processing the baseband signal (called baseband units, BBUs), which are centralized in a single location or even virtualized into the cloud. This approach benefits from simpler radio equipment at the network edge, easier operation, and cheaper maintenance, while the main RAN intelligence (BBUs) is centralized in the operator-controlled premises. The challenge of C-RAN deployments is that such a functional split requires these two elements to be connected through a high-speed, low-latency, and accurately synchronized network, the so-called fronthaul. Such critical requirements are currently met with fiber optics [2, 3].

The C-RAN approach has some some clear benefits with respect to traditional integrated base stations (BSs). First, the cost of deploying RRHs decreases considerably since the installation footprint is much smaller. RRHs do not need any refrigeration or costly on-site construction, thus shortening the time for deployment compared to traditional integrated BSs. On the other hand, BBUs can be aggregated and further virtualized in BBU pools. In this way, BBUs can be shared and turned off when necessary, reducing the cost of maintaining a network with low loads. Finally, another benefit of C-RAN is that it enables the use of cooperative radio techniques, cooperative multipoint (CoMP), allowing reduction of the interference between different radio transmissions and improving its performance. This further enables denser RRH deployments than traditional ones since interference among BSs can be better mitigated [4].

A number of radio equipment manufacturers have defined two main specifications for the transport of fronthaul traffic: the Common Public Radio Interface (CPRI) [5] and the Open Base Station Architecture Initiative (OBSAI). Both solutions are based on the implementation of the digital radio over fiber (D-RoF) concept, whereby the radio signal is sampled and quantized, and, after encoding, transmitted toward the BBU pool. These two specifications differ in the way that information is transmitted. CPRI is a serial line interface transmitting constant bit rate (CBR) data over a dedicated channel, while OBSAI uses a packet-based interface. The mapping methods of CPRI are more efficient than OBSAI [6], and most global vendors have chosen CPRI for their products.

The aim of this article is to present the CPRI specification, its concept, design, and interfaces, and further provide a guideline for fronthaul dimensioning in realistic Long Term Evolution (LTE) scenarios. We also provide some interesting open research challenges and current initiatives to bring the C-RAN concept to the fifth-generation (5G) mobile network. Accordingly, the following section briefly reviews the LTE physical layer (PHY) specifications required to understand the design of CPRI. We then introduce the top-level fronthaul network requirements demanded by CPRI and its main features, including the user plane data, control and management, and synchronization information multiplexing. After that we provide an application example of

The authors are with Universidad Carlos III de Madrid.

0163-6804/16/$25.00 © 2016 IEEE

CPRI in a realistic LTE scenario. Finally, we conclude this work providing a number of open research issues and challenges regarding CPRI and the fronthaul.

LTE Physical Media

This section presents the main features of the LTE PHY; in particular, LTE frequency-division duplex (LTE-FDD) is considered for brevity.

Concerning the downlink (DL), LTE uses orthogonal frequency-division multiple access (OFDMA), while in the uplink (UL) LTE uses single-carrier frequency-division multiple access (SC-FDMA). In both techniques data is encoded on multiple narrowband subcarriers, minimizing the negative effects of multi-path fading, distributing the interference effect across different users.

Concerning downlink, LTE uses OFDMA, while in the uplink LTE uses SC-FDMA. In both techniques data is encoded on multiple narrowband subcarriers, minimizing the negative effects of multi-path fading, distributing the interference effect across different users.

LTE allows spectrum flexibility where the channel bandwidth can be configured from 1.25 to 20 MHz. As an example, the DL with a 20 MHz channel and a 4 × 4 multiple-input multiple-output (MIMO) configuration can provide up to 300 Mb/s of user plane data. The UL peak data rate is 75 Mb/s.

LTE defines a generic frame structure that applies to both DL and UL for FDD operation. Each LTE frame has a duration of 10 ms, and is subdivided into 10 equal-size subframes of 1 ms; each subframe comprises two slot periods of 0.5 ms duration. Depending on the cyclic prefix (CP) duration, each slot carries a number of orthogonal frequency-division multiplexing (OFDM) symbols (7 for the short CP or 6 for the long CP) with T_{symbol} = 66.67 μs.

In the frequency domain, groups of N_{sc} = 12 adjacent subcarriers (15 kHz/subcarrier) are grouped together on a slot-by-slot basis to form so-called physical resource blocks (PRBs), which are the smallest bandwidth unit (180 kHz) assigned by the BS scheduler (Fig. 1). Thus, different transmission bandwidths use various PRBs per time slot, ranging from N_{PRB} = 6 to 100, as shown in Table 1.

Thus, each time slot carries a number of bits depending on the number of symbols per time slot (either 6 or 7), the modulation chosen, and the transmission bandwidth B_{tx}. For example, for B_{tx} = 2.5 MHz (144 subcarriers) with 64-quadrature amplitude modulation (QAM) (6 b/symbol) and short CP (N_{CP} = 7 OFDM symbols per time slot), the number of bits carried in a time slot of 0.5 ms duration is 6048 bits (144 subcarriers × 7 OFDM symbols × 6 b/symbol), and the resulting data rate is approximately 12 Mb/s. The effective data rate is actually less than this value since some resource elements of the PRB are reserved for control and signaling. It is also worth noting that there is one resource grid for each transmitting antenna; in other words, in a 2 × 2 MIMO configuration the value above doubles (24 Mb/s).

In order to recover all of the data transmitted, the receiver must take N_{FFT} samples per OFDM symbol (T_{symbol}) as specified in Table 1. In the example above, the receiver must take N_{FFT} = 256 samples per OFDM symbol (66.67 s) in order to recover the data transmitted in B_{tx} = 2.5 MHz. In this case, the sampling frequency is f_s = 3.84 MHz (1.536 · B_{tx}, as shown in the table), and the sampling period $T_s = 1/f_s$ = $260.4141\overline{6}$ ns.

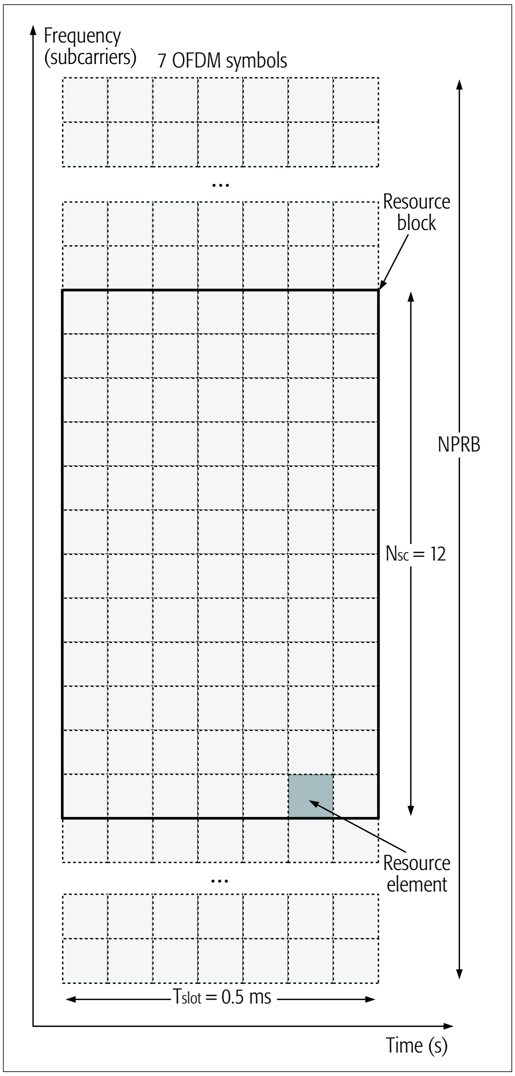


Figure 1. Downlink resource grid defined in LTE.

It is worth highlighting the importance of the f_s = 3.84 MHz sampling reference value of LTE FDD, since the timing and synchronization design of CPRI revolves around this number. Essentially, f_c = 3.84 MHz defines the main clock for CPRI framing, which is then oversampled to obtain the timing references for the other LTE channel bandwidths.[1] In addition, one CPRI basic frame is generated every $1/f_c$ = $260.41\overline{6}$ ns to carry the sampled digitized OFDM symbol, thus completely aligned with the LTE time reference.

Overview of CPRI

Concept and Requirements

According to the CPRI specification v6.1 [5], "*the Common Public Radio Interface (CPRI) is an industry cooperation aimed at defining a publicly available specification for the key internal interface of radio base stations between the Radio Equipment Control (REC) and the Radio Equipment (RE).*" In other words, the CPRI specification provides the physical (L1) and data link layer (L2) details for the transport of digitized radio information between REC and RE.

Figure 2 shows the functional split between REC and RE as defined in the CPRI specification

[1] The value of this clock is inherited from the single clock used in multimode WCDMA user equipments.

Tx BW (Btx)	1.25 MHz	2.5 MHz	5 MHz	10 MHz	15 MHz	20 MHz
Number of PRB (N_{PRB})	6	12	25	50	75	100
FFT size (N_{FFT})	128	256	512	1024	1536	2048
Sampling frequency	1.92 MHz	3.84 MHz	7.68 MHz	15.36 MHz	23.04 MHz	30.72 MHz
(f_s = 15KHz × N_{FFT})	(1/2 × 3.84 MHz)		(2 × 3.84 MHz)	(4 × 3.84 MHz)	(6 × 3.84 MHz)	(8 × 3.84 MHz)
Subcarriers/PRB (N_{sc})	12					
OFDM symbols (N_{CP})	7/6 (Short/Long CP)					
Modulation	QPSK, 16-QAM, 64-QAM					
MIMO configurations	4 × 2, 2 × 2, 1 × 2, 1 × 1					
I/Q data rate (Gb/s) per AxC	0.0576	0.1152	0.2304	0.4608	0.6912	0.9216

Table 1. Downlink OFDM modulation parameters and CPRI bandwidth required for the case of $M = 15$ b/sample.

(DL). As shown in the figure, all the operations above the PHY and most of those of the PHY are performed by the REC, which generates the radio signal, samples it, and sends the resulting data to the RE. The RE basically reconstructs the waveform and transmits it over the air. The uplink case is similar, although the sampling of the radio signal must be performed in the RE. The main benefit of this split is that almost no digital processing functions are required at the RRHs, making them very small and cheap. In addition, the centralization of all the signal processing functions in the BBU simplifies the adoption of cooperative techniques such as CoMP, which require advanced processing of the radio signal of several RRHs simultaneously. Further discussion on alternative functional splits can be found in [7].

Some of the main design features and requirements of CPRI are listed below:

- CPRI supports a wide variety of radio standards: Third Generation Partnership Project (3GPP) Universal Terrestrial Radio Access (UTRA) FDD, WiMAX, 3GPP Evolved UTRA (E-UTRA, LTE), and 3GPP GSM/EDGE. This article only focuses on the use of CPRI for the transport of the E-UTRA interface.
- Although in most practical configurations CPRI will be configured in a point-to-point fashion, the specification also allows different topology configurations: star, chain, tree, ring, and multihop options to carry CPRI data over multiple hops. For example, CPRI natively supports the multiplexing of two CPRI-1 (614.4 Mb/s) into a single CPRI-2 (1228.8 Mb/s) frame through daisy chaining of the REs.
- CPRI requires strict synchronization and timing accuracy between REC and RE: the clock received at the RE must be traceable to the main REC clock with an accuracy of 8.138 ns. This number is exactly a fraction of $T_c = 260.41\overline{6}$, in particular $T_c/32$.
- CPRI equipment must support an operating range of at least 10 km.
- The main requirements for CPRI transmission apart from the required bandwidth are delay and bit error rate (BER). CPRI links should operate with at most 5 μs delay contribution excluding propagation delay, and a maximum allowed BER of 10^{-12}. In addition, the frequency deviation from the CPRI link to the radio BS must be not larger than than 0.002 ppm.

Design and Implementation

CPRI defines three different logical connections between the REC and the RE: user plane data, control and management plane, and synchronization and timing. These three flows are multiplexed onto a digital serial communication line.

User Plane Data: Transported in the form of one or many in-phase and quadrature (IQ) data flows. Each IQ data flow reflects the radio signal, sampled and digitized, of one carrier at one independent antenna element, the so-called antenna carrier (AxC). In the particular case of LTE, an AxC contains one or more IQ samples for the duration of one UMTS chip ($T_c = 1/f_c = 260.41\overline{6}$ ns since $f_c = 3.84$ MHz).

Synchronization Data Used for Time and Frame Alignment: The interface shall enable the RE to achieve the frequency accuracy specified in 3GPP TS 45.10 [8]. The central clock frequency generation in the RE shall be synchronized to the bit clock of one of the ports connecting RE and REC. With 8B/10B or 64B/66B line coding, the bit clock rate of the interface shall be a multiple of 38.4 MHz in order to allow for a simple synchronization mechanism and frequency regeneration.

Control and Management: C&M data can be transmitted by either an in-band protocol (for time-critical signaling data) or higher-layer protocols not defined by CPRI. The inband protocol is used for synchronization and timing, and also for error detection/correction. This makes use of the line codings specified in IEEE 802.3 (line codes 8B/10B and 64B/66B). The physical layer is capable of detecting link failures and synchronization issues as a result of line code violations.

Vendor-Specific: CPRI reserves some time slots for the transmission of any vendor-specific data, allowing manufacturers to customize their solutions.

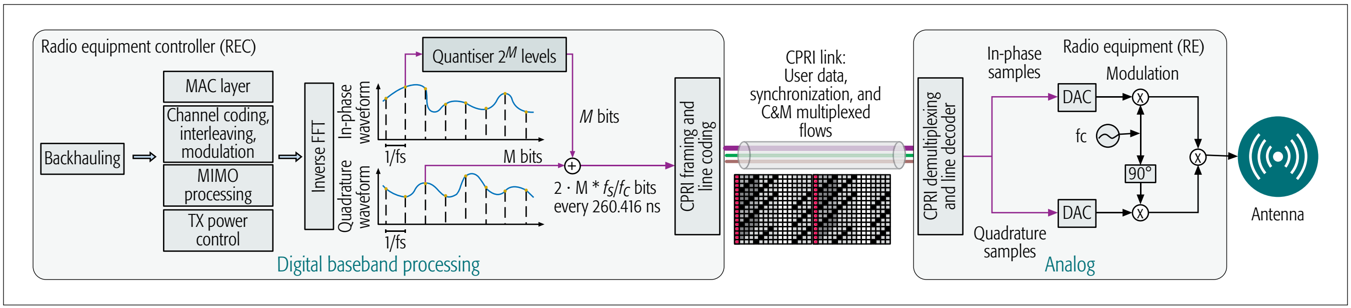

Radio equipment controller (REC)
MAC layer
Channel coding, interleaving, modulation
MIMO processing
TX power control
Backhauling
Inverse FFT
In-phase waveform
Quadrature waveform
Quantiser 2^M levels
M bits
1/fs
M bits
1/fs
2 · M * fs/fc bits every 260.416 ns
CPRI framing and line coding
Digital baseband processing
CPRI link: User data, synchronization, and C&M multiplexed flows
CPRI demultiplexing and line decoder
In-phase samples
Radio equipment (RE)
Modulation
DAC
fc
90°
DAC
Quadrature samples
Analog
Antenna

Figure 2. Conceptual explanation of REC/RE functional split.

Transmission of User Plane Data

The transmission of user plane data is based on the concept of an antenna carrier (AxC). Given that the LTE radio signal is first sampled and then quantized (Fig. 2), the amount of information carried by an AxC depends on two parameters:

- The sampling frequency f_s, which is a multiple of the nominal chip rate f_c = 3.84 MHz (Table 1).
- The number of bits M used in the quantization process of the I and Q radio signals. In E-UTRA, $M = 8, \ldots, 20$ either DL or UL. Previous work [9] and actual field programmable gate array (FPGA) implementation of CPRI consider $M = 15$ for capacity efficiency.

For example, in a configuration with $M = 15$ b/sample, one AxC comprises 15 + 15 = 30 b/IQ sample, which are transmitted in the following interleaved sequence:

$$I_0Q_0I_1Q_1 \ldots I_{M-1}Q_{M-1},$$

that is, from the least significant bit (LSB) to the most significant bit (MSB).

In CPRI, one basic frame is created and transmitted every $T_c = 260.41\overline{6}$ ns, which is based on the Universal Mobile Telecommunications System (UMTS) clock rate, that is, 3.84 MHz. This duration remains constant for all CPRI line bit rate options. As already indicated, this value of T_c is designed to transport one fast Fourier transform (FFT) sample for an LTE channel bandwidth of 2.5 MHz, two samples for the 5 MHz bandwidth, four samples for the 10 MHz channel, and so on.

A basic frame comprises $W = 16$ words ($w = 0, \ldots, 15$) whereby the length T of each word depends on the CPRI line bit rate option (Table 2). The exact line bit rate values for each option are computed in the second column of Table 2. In all cases, the first word $w = 0$ is reserved for control, while the other 15 words are used to carry IQ data samples. For example, in CPRI option 1, there is room for 120 (= 15 words × 8 b/word) bits for transporting the IQ samples of several AxCs. Thus, in a configuration of $2M = 30$ b/AxC, one basic frame can carry up to 4 AxCs consisting of one sample each. This is a basic configuration for an antenna serving four sectors with 2.5 MHz LTE channel bandwidth. It is worth remarking that four 2.5 MHz AxCs carry about 4 · 12 = 48 Mb/s of actual LTE data, and are spread over 614.4 Mb/s after CPRI encapsulation; this is about 13 times higher bit rate.

CPRI defines a hierarchical framing with three layers (Fig. 3), chosen this way to match the framing numbers of the LTE FDD frame structure:

- Basic frame, of variable size, created and transmitted every $T_c = 260.41\overline{6}$ ns.
- Hyperframe, which is a collection of 256 basic frames. One hyperframe is created every 256 × T_c = 66.67 μs, which is the OFDM symbol time in LTE. Thus, a hyperframe carries all the FFT samples required to decode the whole OFDM symbol.
- CPRI frame, which is a collection of 150 hyperframes. A CPRI frame is created every 10 ms and carries the digital samples of a whole LTE frame.

Control and Management, and Synchronization

As noted before, the first word ($w = 0$) in every basic frame (control word) carries C&M information; thus, 256 control words are available per hyperframe. These 256 control words are organized into 64 subchannels of 4 control words each (Fig. 4). As shown, every control word can be addressed by a subchannel ID (0, …, 63).

Each subchannel belongs to one category out of seven:

Synchronization: The control word on the first basic frame (CW 0 in Fig. 4) is reserved to indicate the starting of a new hyperframe. This control word uses a special 8B/10B (K28.5) or 64B/66B (50h) code. The three remaining words in the synchronization subchannel (words 64, 128, and 192) are used to signal the hyperframe number and the node B frame number (BFN) for synchronization purposes with the LTE framing.

L1 In-Band Protocol: Subchannel 2 carries the necessary signaling required to set up the different C&M links, including starting up, resetting, and tearing down the CPRI link, and also to handle alarms at the PHY for different events such as loss of synchronization.

Slow C&M Link: The subchannels assigned to this category enable the transmission of high-level data link control (HDLC) frames. HDLC is a well-known layer 2 protocol providing basic functionalities such as flow control and error correction based on retransmission.

Ctrl_AxC: A Ctrl_AxC designates one AxC-specific control data stream. The mapping of Ctrl_AxCs to AxCs as well as the actual content of the control data bytes are not defined in CPRI but are vendor-specific.

Fast C&M Link: In addition to the slow C&M link, the operator of the CPRI link is provided

			Number of AxCs of channel bandwidth and bit rate required per AxC						
Option #	CPRI data rate (Mb/s)	Coding	T	1.25 MHz (76.8 Mb/s)	2.5 MHz (153.6 Mb/s)	5 MHz (307.2 Mb/s)	10 MHz (614.4 Mb/s)	15 MHz (921.6 Mb/s)	20 MHz (1228.8 Mb/s)
1	614.4	8B/10B	8	8	4	2	1	–	–
2	1228.8	8B/10B	16	16	8	4	2	1	1
3	2457.6	8B/10B	32	32	16	8	4	2	1
4	3072	8B/10B	40	40	20	10	5	3	2
5	4915.2	8B/10B	64	64	32	16	8	5	4
6	6144	8B/10B	80	80	40	20	10	6	5
7	9830.4	8B/10B	128	128	64	32	16	10	8
				(63.36 Mb/s)	(126.72 Mb/s)	(253.44 Mb/s)	(506.88 Mb/s)	(760.32 Mb/s)	(1013.76 Mb/s)
7A	8110.08	64B/66B	128	128	64	32	16	10	8
8	10137.6	64B/66B	160	160	80	40	20	13	10
9	12165.12	64B/66B	192	192	96	48	24	16	12

Table 2. Maximum number of AxC transported in a CPRI link, $M = 15$ bits.

with a fast C&M subchannel to transmit other control information. Such control frames are first encapsulated over Ethernet and then transmitted over this subchannel. Fragmentation and reassembly are needed. For this purpose, CW 194 carries a pointer to the CW in the hyperframe containing the first byte of the Ethernet frame (shown in Fig. 4 as pointer P).

Reserved for Future Use and Vendor-Specific.

CPRI Fronthaul Dimensioning in C-RAN Scenarios

General Dimensioning Guidelines

Following the discussion earlier, the D-RoF transmission (i.e., sampling and quantization) of an AxC requires a data bit rate of $B_{AxC} = (2M) f_s$ b/s, expanded by factors 16/15 (15 words data, 1 word C&M) and either 10/8 or 66/64 (8B/10B or 66B/64B line coding, respectively). According to this, a 2.5 MHz LTE channel requires 153.6 Mb/s per AxC.

In this light, Table 2 shows the bit rate required per AxC for different LTE bandwidths and the maximum number of AxCs transported for standard CPRI bit rates. This table provides a good starting point for dimensioning fronthaul networks in C-RAN scenarios, and should be read as follows: CPRI option 6 (6144 Mb/s) can carry 80 AxCs @ 1.25 MHz LTE bandwidth, 40 AxC @ 2.5 MHz, or 5 AxC @ 20 MHz. On the other hand, if the LTE setup is fixed to a number of 3 sectors and 2 × 2 MIMO @ 10 MHz LTE bandwidth (i.e., 2 × 3 AxCs), a lookup in Table 2, column "10 MHz LTE bandwidth" reveals that at least CPRI option 5 is required to carry such a number of AxC.

Use Case: CPRI Downlink Requirements for a Four-Antenna Site, 2×2 MIMO, 20 MHz Channel Scenario

Consider the four-antenna/four-sector scenario operating an LTE 2×2 MIMO channel of 20 MHz bandwidth depicted in Fig. 3a. This scenario requires the multiplexing and transmission of four AxC groups (one per sector), while each AxC group comprises two AxCs, as shown in the figure.

Figure 3b shows the amount of information carried in each AxC. As shown, one IQ sample (2M = 30 bits) is generated every $1/f_s$, where f_s = 30.72 MHz for 20 MHz LTE channels (Table 1). Thus, a total of 8 × 30 = 240 bits are generated every $1/f_s$. It is also worth remarking that f_s = 30.72 MHz is exactly $8f_c$; hence, 8 IQ samples are generated every $1/f_c = T_c$ = 260.416 ns (i.e., 1920 IQ b/T_c total). This amount of information requires 8 × 1228.8 Mb/s = 9830.4 Mb/s (8B/10B assumed), which is CPRI option 7 in Table 2. Alternatively, CPRI option 7A is also suitable for carrying the same 8 AxCs @ 20 MHz LTE channel and even requires slightly less bandwidth since 64B/66B is used. In both cases, a 10 Gb/s Ethernet transceiver is suitable as a physical medium for this scenario.[2]

Figure 3c shows how the different AxC are grouped together and multiplexed over the line. The CPRI specification defines three mapping methods to multiplex different AxCs; we have chosen mapping method 3, which is backward compatible with previous CPRI specifications. Essentially, the IQ samples are arranged in order per AxC group (group 1 first, group 4 last) and interleaved within the group (30 bits AxC0, then 30 bits AxC1, then 30 bits AxC0 again, etc. for group 1).

Such ordering is then used to construct a CPRI basic frame (Fig. 3d) noting that one word for C&M is added ahead of the 1920 data bits. One basic frame is constructed this way every 260.416 ns; 256 basic frames form a hyperframe (66.67 μs), which includes the information of one LTE OFDM symbol; and 150 hyperframes form a super frame, which is synchronized with the 10 ms LTE frame (Fig. 3e).

Other scenarios would follow the same guidelines as before. For instance, the same configu-

[2] It is worth noting that this configuration requires daisy chaining of the different REs. If this is not possible, a potential configuration may use 4 CPRI-3 (2457.3 Mb/s) links.

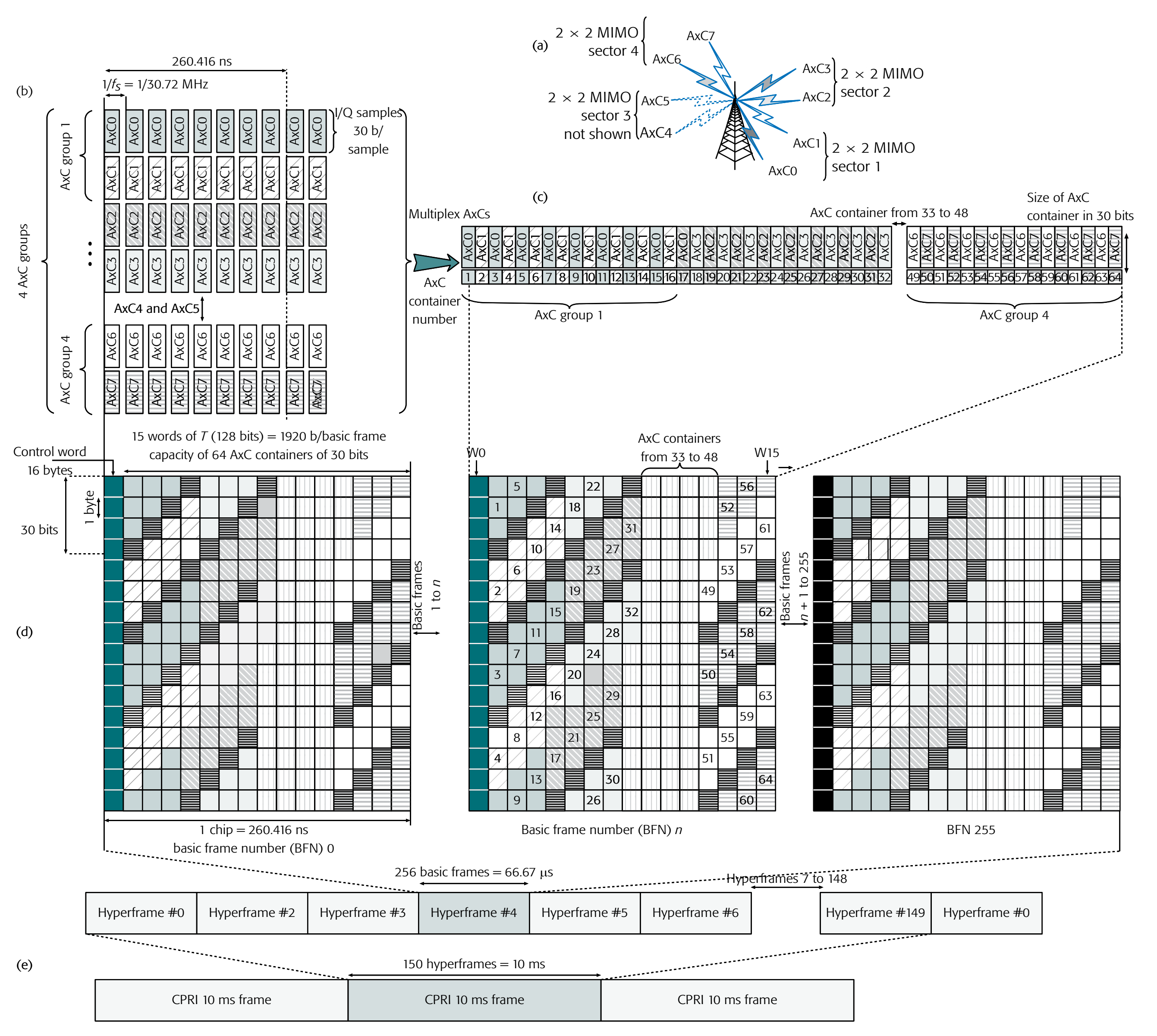


Figure 3. CPRI multiplexing of AxC data in a 2 × 2 MIMO 20 MHz channel use case: a) scenario; b) AxC generation; c) AxC arrangement and serialization; d) basic frame construction; e) hyperframes and 10 ms CPRI frame.

ration in a 4 × 4 MIMO scenario would require the same sampling frequency f_s, but the data rate would double since we now have 4 AxC groups with 4 AxCs per group, that is, a total of 16 AxCs. The arrangement of Fig. 3c would be the same for the AxC group (group 1 first, group 4 last), but AxCs within the group would alternate (AxC0, AxC1, AxC2, AxC3, AxC0 again, and so on for group 1).

Summary, Challenges and Future Research

This work has provided a short overview of CPRI, including concept, design, specification, and use case in an LTE C-RAN-based environment. The concept of C-RAN has recently appeared in the market, and the idea of separating RECs (BBUs) from REs (RRHs) is gaining traction in the mobile network industry.

On the research side, there is a common consensus on the key challenges of CPRI technology [10]. First, the amount of bandwidth required to transmit the radio signal is simply overwhelming for LTE. Moreover, the upcoming 5G RANs, where 100 MHz channels with massive MIMO are envisioned, may require several tens or even hundreds of gigabits per second capacity in the fronthaul [11]. As an example, an 8 × 8 MIMO antenna covering four sectors produces 32 AxCs, which translate into around 32 Gb/s for 20 MHz bandwidth channels. In the case of 100 MHz LTE channels, this same scenario requires five times (i.e., 160 Gb/s) the previous CPRI bandwidth.

Second, CPRI is a serial CBR interface with new frames transmitted every T_c = 260.416 ns. This, together with the low-latency and strict synchronization requirements demanded, makes it very challenging to have CPRI and other traffic sources over the same link. Recent studies have approached this problem focusing on band-

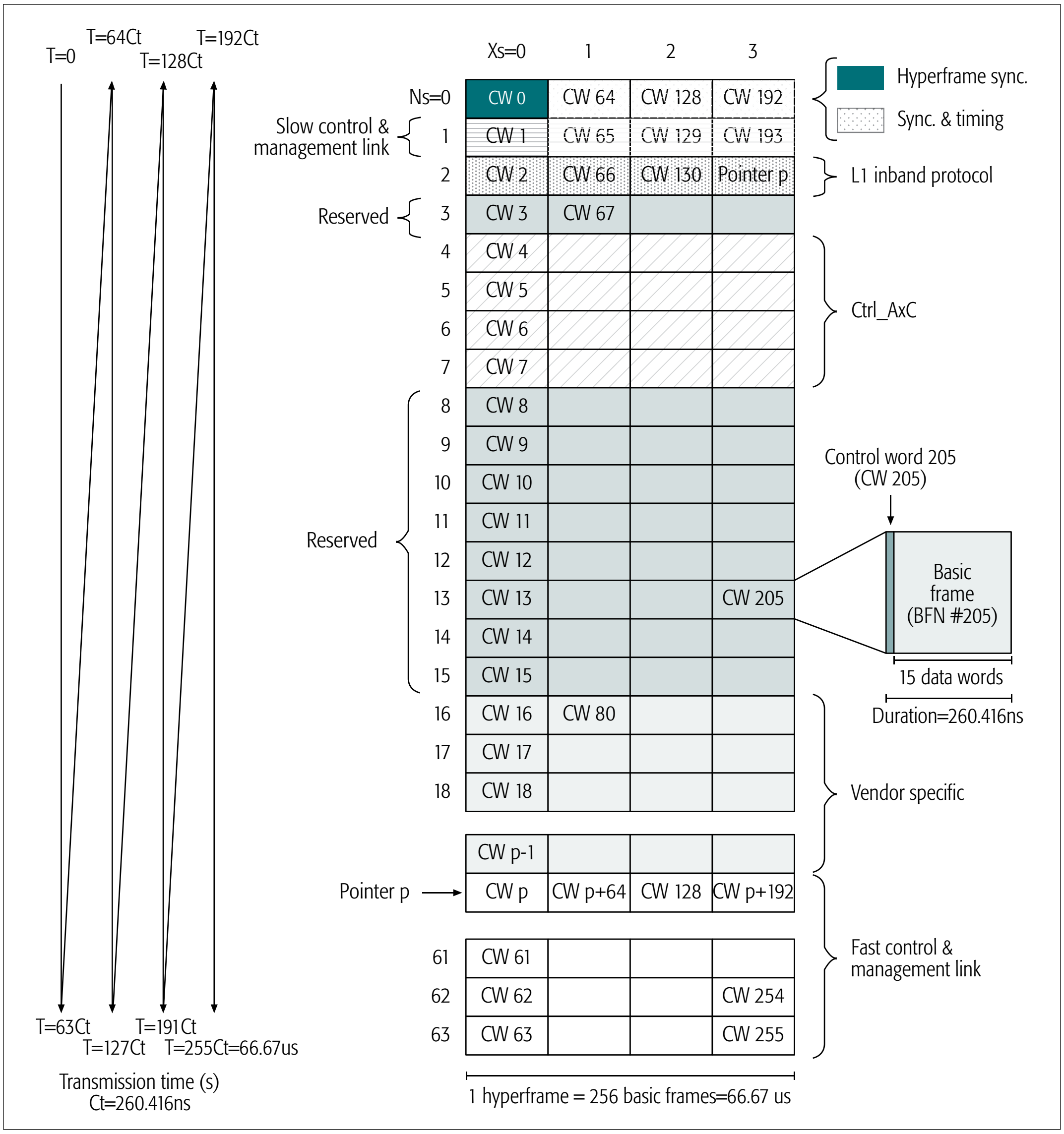


Figure 4. CPRI multiplexing of C&M channels in the hyperframe. C&M information is carried in the control word (CW) of each CPRI frame.

The recently created Time Sensitive Networking Task Group of IEEE 802.1 is working on developing new extensions to support the forwarding of Ethernet traffic with delay and jitter guarantees, including mechanisms such as frame preemption, expedited traffic forwarding, and jitter reduction techniques, mainly buffering.

width compression techniques. For example, the authors in [12] claim to provide about 1/5 compression ratios within the 5 μs delay budget allowed by CPRI, thus significantly reducing the link load.

Bandwidth compression is indeed a starting point toward the packetization of CPRI data, via Ethernet framing, for instance. However plain Ethernet is asynchronous and best effort, and therefore not suitable as such for the transport of CPRI traffic. In this light, the recently created Time Sensitive Networking (TSN) Task Group of IEEE 802.1[3] is working on developing new extensions to support the forwarding of Ethernet traffic with delay and jitter guarantees, including mechanisms such as frame preemption, expedited traffic forwarding, and jitter reduction techniques, mainly buffering [13].

In addition, the use of synchronous Ethernet seems mandatory in multihop scenarios [14]. Nevertheless, although high-precision timing protocols over Ethernet exist (see IEEE 1588v2), their accuracy is in the range of a few hundred nanoseconds, while CPRI requires at most tens of nanoseconds between REC and RE. New approaches using frequency adjustable oscillators or GPS signals are under study to solve this issue.

Finally, both research projects and standardization bodies (e.g., IEEE 1904.3 Standard for Radio over Ethernet Encapsulation and Mappings[4]) are exploring the possible gains of redefining the RE/REC functional split of C-RAN in the next-generation networks [15]. Examples include the decoupling of fronthaul bandwidth and antenna number by moving antenna related operations to the RE (DL antenna mapping, FFT, etc.), or enabling traffic-dependent bandwidth adaptation by effectively coupling fronthaul bandwidth with the actual traffic served in the cell. The latter relies on the fact that many cell processing functions do not depend on the number of users, including FFT, cyclic prefix addition/removal, synchronization signals, and so on. More information about this novel approach can be found in [7].

[3] http://www.802tsn.org/

[4] http://www.ieee1904.org/3/

Acknowledgments

The authors would like to acknowledge the support of projects CRAMnet (grant no. TEC2012-38362-C03-01) and EU H2020 5G-Crosshaul Project (grant no. 671598) to the development of this work.

References

[1] "C-RAN: The Road towards Green RAN," China Mobile White Paper, v2, 2011.

[2] A. Lometti *et al.*, "Backhauling Solutions for LTE Networks," *Proc. 16th Int'l. Conf. Transparent Optical Networks*, 2014 , July 2014, pp. 1–6.

[3] A. Pizzinat *et al.*, "Things You Should Know About Fronthaul," *IEEE/OSA J. Lightwave Tech.*, 2015.

[4] R. Irmer *et al.*, "Coordinated Multipoint: Concepts, Performance, and Field Trial Results," *IEEE Commun. Mag.*, vol. 49, no. 2, Feb. 2011, pp. 102–11.

[5] Specification, CPRI, "V6.1 Common Public Radio Interface (CPRI); Interface Specification," NEC Corp., Nortel Networks SA, Siemens Networks GmbH & Co. KG, Ericsson AB, and Huawei Technologies Co Ltd., July, 2014, 129 pages.

[6] M. Nahas *et al.*, "Base Stations Evolution: Toward 4G Technology," *Proc. IEEE Int'l. Conf. Telecommun.*, 2012, pp. 1–6.

[7] D. Wubben *et al*., "Benefits and Impact of Cloud Computing on 5G Signal Processing: Flexible Centralization through Cloud-RAN," *IEEE Signal Processing Mag.*, vol. 31, no. 6, Oct. 2014, pp. 35–44.

[8] 3GPP, "3GPP TS 45.010: Radio Subsystem Synchronization," Rel. 10, V10.1.0, 2011.

[9] C. F. A. Lanzani, L. Dittmann, and M. S. Berger, *4G Mobile Networks: An Analysis of Spectrum Allocation, Software Radio Architectures and Interfacing Technology*, Ph.D. dissertation, Tech. Univ. Denmark, Dept. of Photonics Engineering.

[10] A. Saadani *et al*., "Digital Radio over Fiber for LTE-Advanced: Opportunities and Challenges," *Proc. 17th IEEE Int'l. Conf. Optical Network Design and Modeling*, 2013, 2013, pp. 194–99.

[11] J. E. Mitchell, "Integrated Wireless Backhaul over Optical Access Networks," *IEEE/OSA J. Lightwave Technology*, vol. 32, no. 20, Oct. 2014, pp. 3373–82.

[12] B. Guo *et al.*, "CPRI Compression Transport for LTE and LTE-A Signal in C-RAN," *Proc. CHINACOM*, vol. 0, 2012, pp. 843–49.

[13] T. Wan and P. Ashwood, "A Performance Study of CPRI over Ethernet;" http://www.ieee1904.org/3/meetingarchive/2015/02/tf31502ashwood1a.pdf

[14] J. Aweya, "Implementing Synchronous Ethernet in Telecommunication Systems," *IEEE Commun. Surveys & Tutorials*, vol. 16, no. 2, 2nd qtr. 2014, pp. 1080–1113.

[15] P. Rost *et al*., "Cloud Technologies for Flexible 5G Radio Access Networks," *IEEE Commun. Mag*., vol. 52, no. 5, May 2014, pp. 68–76.

Biographies

Antonio de la Oliva (aoliva@it.uc3m.es) received his telecommunications engineering degree in 2004 and his PhD. in 2008 from Universidad Carlos III Madrid (UC3M), Spain, where he has been working as an associate professor since. His current line of research is related to networking in extremely dense networks. He is an active contributor to IEEE 802, where he has served as Vice-Chair of IEEE 802.21b and Technical Editor of IEEE 802.21d. He has also served as a Guest Editor of *IEEE Communications Magazine*.

José Alberto Hernández (jahgutie@it.uc3m.es) completed his five-year degree in telecommunications engineering at UC3M in 2002, and his Ph.D. degree in computer science at Loughborough University,Leicester, United Kingdom, in 2005. He has been a senior lecturer in the Department of Telematics Engineering since 2010, where he combines teaching and research in the areas of optical WDM networks, next-generation access networks, metro Ethernet, energy efficiency, and Hybrid optical-wireless technologies. He has published more than 75 articles in both journals and conference proceedings on these topics. He is a co-author of the book *Probabilistic Modes for Computer Networks: Tools and Solved Problems*.

David Larrabeiti is a professor of switching and networking architectures at UC3M since 1998. He has participated in EU-funded research projects related to next-generation networks and protocols since FP6. He was UC3M Principal Investigator of the FP7 BONE European network of excellence on optical networking. He is currently involved in the FP7 Fed4FIRE and H2020 5G-Crosshaul research projects, participating in the development of technology and testbeds for backhaul network design.

Arturo Azcorra received his M.Sc. degree in telecommunications engineering from Universidad Politécnica de Madrid (UPM) in 1986 and his Ph.D. from the same university in 1989. In 1993, he obtained an M.B.A. with honors from Instituto de Empresa He has participated in and directed 49 research and technological development projects, including European ESPRIT, RACE, ACTS, and IST programs. He coordinated the CONTENT and E-NEXT European Networks of Excellence, and the CARMEN EU project. In addition to his scientific achievements, he has a relevant track record of research management. He was deputy vice rector of Academic Infrastructures at UC3M from 2000 to 2007. He served as director general for Technology Transfer and Corporate Development at the Spanish Ministry of Science and Innovation from 2009 to 2010, and then as director general of CDTI (the Spanish agency for industrial research) from 2010 to 2012. He is the founder of the international research center IMDEA Networks, and currently is its director, with a double affiliation as a full arofessor at UC3M.

The use of Synchronous Ethernet seems mandatory in multi-hop scenarios. Nevertheless, although high-precision timing protocols over Ethernet exist, their accuracy are in the range of few hundred nanoseconds, while CPRI requires at most tens of nanoseconds between REC and RE.

Advertisers' Index

ADVERTISING SALES OFFICES

Closing date for space reservation: 15th of the month prior to date of issue

NATIONAL SALES OFFICE
James A. Vick
Sr. Director Advertising Business, IEEE Media
EMAIL: jv.ieeemedia@ieee.org

Marion Delaney
Sales Director, IEEE Media
EMAIL: md.ieeemedia@ieee.org

Mark David
Sr. Manager Advertising & Business Development
EMAIL: m.david@ieee.org

Mindy Belfer
Advertising Sales Coordinator
EMAIL: m.belfer@ieee.org

NORTHERN CALIFORNIA
George Roman
TEL: (702) 515-7247
FAX: (702) 515-7248
EMAIL: George@George.RomanMedia.com

SOUTHERN CALIFORNIA
Marshall Rubin
TEL: (818) 888 2407
FAX: (818) 888-4907
EMAIL: mr.ieeemedia@ieee.org

MID-ATLANTIC
Dawn Becker
TEL: (732) 772-0160
FAX: (732) 772-0164
EMAIL: db.ieeemedia@ieee.org

NORTHEAST
Merrie Lynch
TEL: (617) 357-8190
FAX: (617) 357-8194
EMAIL: Merrie.Lynch@celassociates2.com

Jody Estabrook
TEL: (77) 283-4528
FAX:: (774) 283-4527
EMAIL: je.ieeemedia@ieee.org

SOUTHEAST
Scott Rickles
TEL: (770) 664-4567
FAX: (770) 740-1399
EMAIL: srickles@aol.com

MIDWEST/CENTRAL CANADA
Dave Jones
TEL: (708) 442-5633
FAX: (708) 442-7620
EMAIL: dj.ieeemedia@ieee.org

MIDWEST/ONTARIO, CANADA
Will Hamilton
TEL: (269) 381-2156
FAX: (269) 381-2556
EMAIL: wh.ieeemedia@ieee.org

TEXAS
Ben Skidmore
TEL: (972) 587-9064
FAX: (972) 692-8138
EMAIL: ben@partnerspr.com

EUROPE
Christian Hoelscher
TEL: +49 (0) 89 95002778
FAX: +49 (0) 89 95002779
EMAIL: Christian.Hoelscher@husonmedia.com

COMPANY	PAGE
Fraunhofer Heinrich Hertz Institute	Cover 4
IEEE Media	8
IEEE Xplore	29
Keysight	Cover 2
National Instruments	3
REMCOM	7
Rohde & Schwarz GmbH	13
Wiley	Cover 3

Currently Scheduled Topics

Topic	Issue Date	Manuscript Due Date
Enabling Mobile and Wireless Technologies for Smart Cities	December 2016	February 29, 2016
Next Generation 911	November 2016	March 15, 2016
Integrated Communications, Control, and Computing Technologies for Enabling Autonomous Smart Grid	December 2016	April 1, 2016
New Waveforms and Multiple Access Methods for 5G Networks	November 2016	April 2, 2016
Impact of Next-Generation Mobile Technologies on IoT-Cloud Convergence	January 2017	April 15, 2016
Practical Perspectives on IoT in 5G Networks: From Theory to Industrial Challenges and Business Opportunities	February 2017	May 1, 2016

www.comsoc.org/commag/call-for-papers

Topics Planned for the March Issue

Critical Communications and Public Safety Networks

Radio Communications

Network Testing

Communications Standards Supplement:
Semantics for Anything-as-a-Service

FROM THE OPEN CALL QUEUE

Device-to-Device Broadcasting Communications in Beyond 4G Cellular Networks

Standards for Media Synchronization

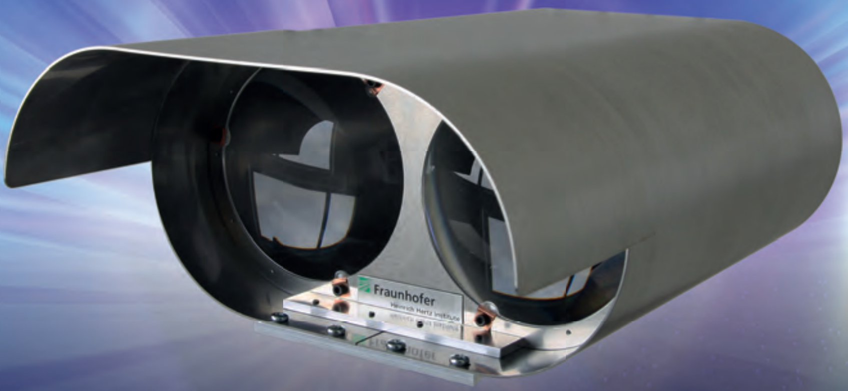
Network Functions Virtualization in 5G

Why to Decouple the Uplink and Downlink in Cellular Networks and How to Do It

Stuff You Don't Learn in Engineering School
Skills for Success in the Real World
CARL SELINGER
Fast-Tracking Your Career
Soft Skills for Engineering and IT Professionals
Wushow "Bill" Chou
IEEE PRESS
WILEY
Ten Essential Skills for Electrical Engineers
BARRY L. DORR
Negotiating Cultural Encounters
Narrating Intercultural Engineering and Technical Communication
EDITED BY Han Yu · Gerald Savage
Information Overload
An International Challenge for Professional Engineers and Technical Communicators
EDITED BY Judith B. Strother · Jan M. Ulijn · Zohra Fazal
Slide Rules
Design, Build, and Archive Presentations in the Engineering and Technical Fields
Traci Nathans-Kelly · Christine G. Nicometo
ENGINEERING Your RETIREMENT
Retirement Planning for Technology Professionals
MIKE GOLIO

Fraunhofer
Heinrich Hertz Institute

LED BASED OPTICAL WIRELESS BACKHAUL LINK



Robust, low latency infrared LED link for mobile backhaul

Specifications

- Infrared LED based
- Easy alignment:
  500 Mbps over 100 m
  250 Mbps over 200 m
- Bidirectional data exchange
- Dynamic rate adaptation
- Latency: < 2 ms
- 1 GbE chipset and interface
- Footprint and weight:
  240 mm x 230 mm x 130 mm, 3 kg

Benefits

- Low cost optical wireless link based on infrared LEDs
- Improved link robustness due to rate adaptation
- No active tracking needed

Applications

- Wireless point-to-point communication in industrial environments
- Backhauling for WiFi and LTE
- Building to building connectivity
- Redundancy for fixed line connection

E-Mail: products-pn@hhi.fraunhofer.de
Web: www.hhi.fraunhofer.de/LED-Backhaul