THIS MONTH'S DIGITAL DELIVERY OF IEEE COMMUNICATIONS MAGAZINE

SUPPORTED BY:

June 2011, Vol. 49, No. 6

IEEE Communications MAGAZINE

www.comsoc.org

- Energy Efficiency in Communications
- Radio Communications
- Consumer Communications

Free ComSoc Tutorial
Spectrum Management
See Page 1

IEEE
IEEE COMMUNICATIONS SOCIETY
A Publication of the IEEE Communications Society

RFMD®

RFSA2013/RFSA2023

Voltage Controlled Attenuators, 50 to 4000MHz

RFMD 20 1991 • 2011

RFMD® delivers a revolutionary new circuit architecture in the RFSA2013 and RFSA2023. Both products are fully monolithic analog voltage controlled attenuators (VCAs) that feature exceptional linearity over a typical temperature compensated 30dB gain control range. Each of these VCAs is controlled by a single positive control voltage with on-chip DC conditioning circuitry. These game-changing products incorporate the complete solution in a small 3 x 3mm QFN package that reduces the footprint by 20x in board area and reduces the DC power by 10x over conventional PIN diode approaches.

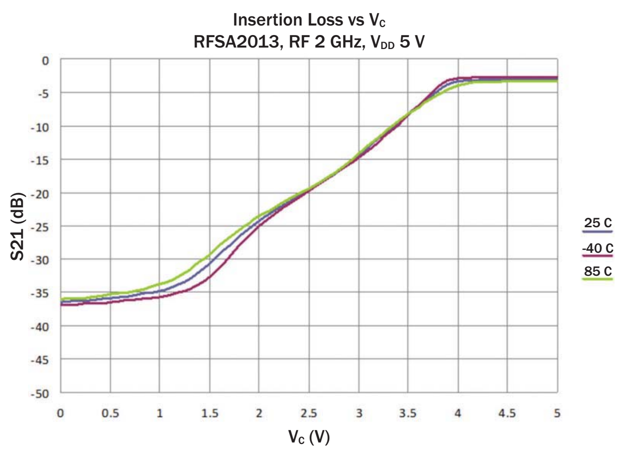
SPECIFICATIONS

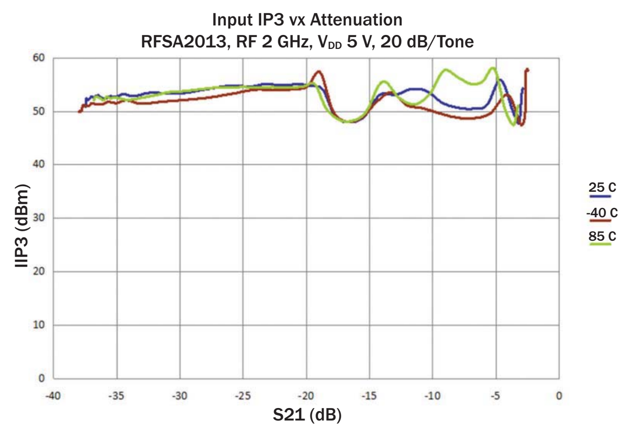
Freq Range (MHz)	*Minimum Insertion Loss (dB)*	*Gain Control Range (dB)*	*Input 1dB Compression Point (dBm)*	*IIP3 (dBm)*	*Supply Voltage (V)*	*Package (mm)*	*Part Number*
50 to 4000	2.6	33.2	30	50	5.0	3 x 3 QFN	RFSA2013
50 to 4000	2.6	33.2	30	50	3.3	3 x 3 QFN	RFSA2023

Order RFMD products online at www.rfmd.com/EDC0611.

FEATURES

- Patent pending circuit architecture
- Broadband 50 to 4000MHz frequency range
- 30dB attenuation range
- +50dBm IIP3 typical
- +80dBm IIP2 typical
- High 1dB compression point >+30dBm
- Low supply current 1mA typical
- Power supply:
  - 5V for RFSA2013
  - 3.3V for RFSA2023
- Linear in dB control characteristic
- Internal temperature compensation
- Class 1C ESD (1000 V)
- Complete solution in small 3 x 3mm QFN package
- Applications include: cellular, 3G infrastructure; WiBro, WiMAX, LTE; microwave radio; high-linearity power control





For sales or technical support, contact your authorized local sales representative (see www.rfmd.com/globalsales).
Register to receive RFMD's latest product releases with our Email Component Alerts at www.rfmd.com/emailalert.
7628 Thorndike Rd., Greensboro, NC 27409-9421 USA • Phone 336.664.1233

RFMD® Mobility. Connectivity. Energy.

These products comply with RFMD's green packaging standards.

IEEE Communications MAGAZINE

June 2011, Vol. 49, No. 6

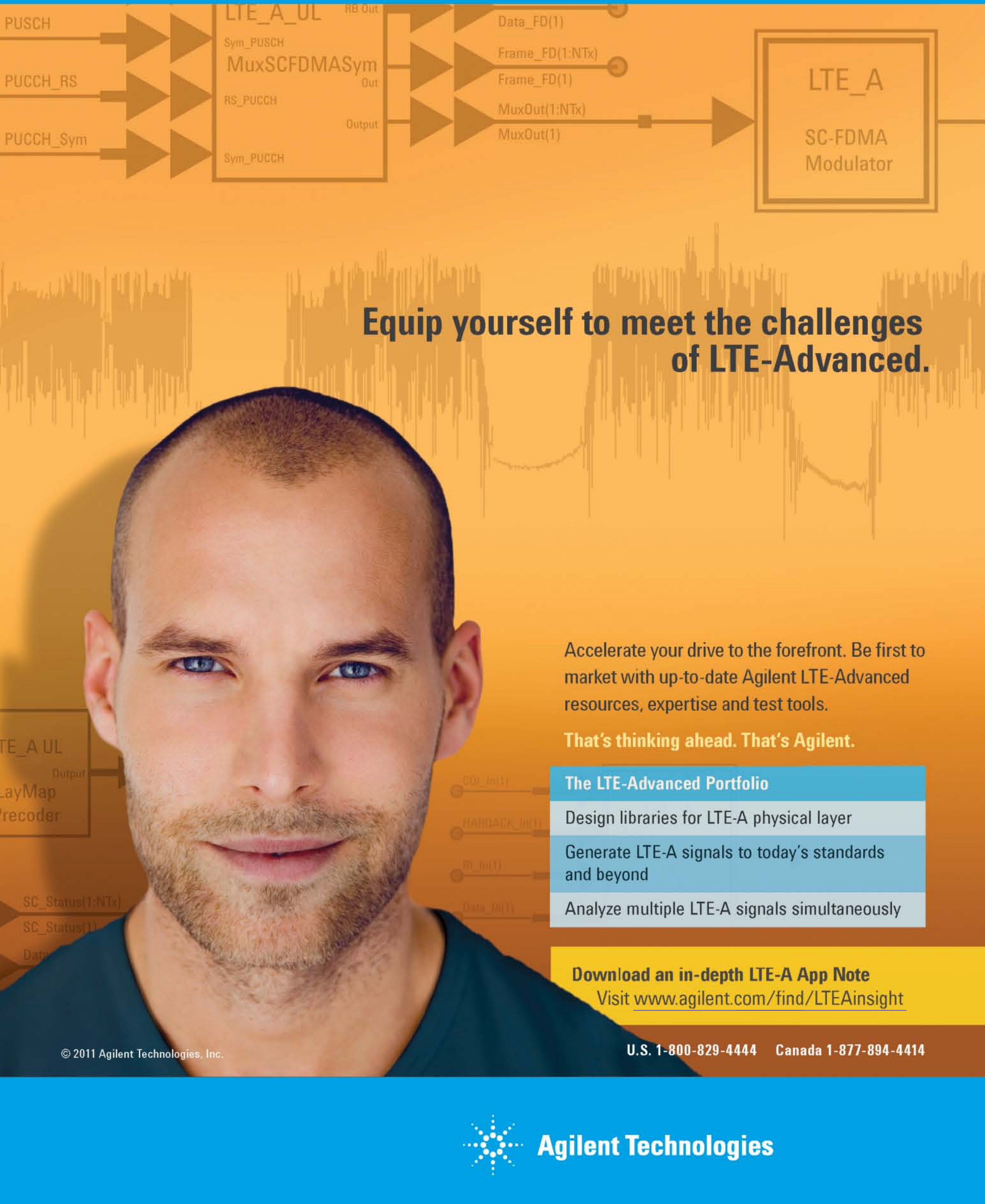
www.comsoc.org/~ci

Energy Efficiency in Communications: Part II

Guest Editors: Honggang Zhang, Andreas Gladisch, Mario Pickavet, Zhifeng Tao, and Werner Mohr

Topics in Radio Communications

Series Editors: Joseph Evans and Zoran Zvonar



PUSCH
PUCCH_RS
PUCCH_Sym
LTE_A_UL
MuxSCFDMASym
LTE_A
SC-FDMA Modulator
Equip yourself to meet the challenges of LTE-Advanced.
Accelerate your drive to the forefront. Be first to market with up-to-date Agilent LTE-Advanced resources, expertise and test tools.
That's thinking ahead. That's Agilent.
The LTE-Advanced Portfolio
Design libraries for LTE-A physical layer
Generate LTE-A signals to today's standards and beyond
Analyze multiple LTE-A signals simultaneously
Download an in-depth LTE-A App Note
Visit www.agilent.com/find/LTEAinsight
© 2011 Agilent Technologies, Inc.
U.S. 1-800-829-4444 Canada 1-877-894-4414
Agilent Technologies

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/ci. 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, Steve Gorshe, PMC-Sierra, Inc., 10565 S.W. Nimbus Avenue, Portland, OR 97223; tel: +(503) 431-7440, e-mail: steve_gorshe@pmc-sierra.com.

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.

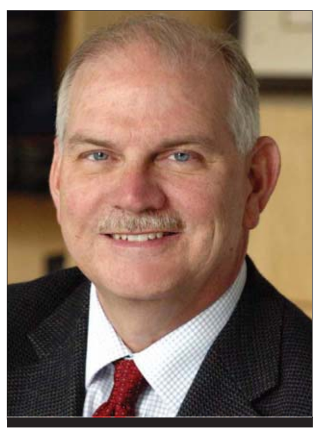
SUBMISSIONS: The magazine welcomes tutorial or survey articles that span the breadth of communications. Submissions will normally be approximately 4500 words, with few mathematical formulas, accompanied by up to six figures and/or tables, with up to 10 carefully selected references. Electronic submissions are preferred, and should be sumitted through Manuscript Central http://mc.manuscriptcentral.com/commag-ieee. Instructions can be found at the following: http://dl.comsoc.org/livepubs/ci1/info/sub_guidelines.html. For further information contact Sean Moore, Associate Editor-in-Chief (smoore-phd@ieee.org). All submissions will be peer reviewed.

BPA WORLDWIDE

COMSOC AWARDS: RECOGNIZING OUR COLLEAGUES

One of the most important things we in the IEEE Communication Society (ComSoc) do is recognize the contributions of our colleagues to the field and to humanity. Such recognition comes from many quarters, from the Institute, from our Technical Committees, from our Chapters, etc. A major part of this activity involves our Society-wide awards, which are awarded based on nominations received from the membership and administered by the ComSoc Awards Committee.

BYEONG GI LEE



H. VINCENT POOR

I share this issue of the President's Page with Vince Poor, the ComSoc Awards Committee Chair. Vince is Dean of Engineering and Applied Science at Princeton University, where he is also the Michael Henry Strater University Professor. He received the Ph.D. degree from Princeton in 1977, and prior to joining the Princeton faculty in 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. He has also held visiting appointments at several institutions, including Harvard, Stanford, and Imperial College. His current research interests are primarily devoted to statistical signal processing, stochastic analysis and information theory, and their applications to various types of networks, including wireless networks, social networks, and electricity grids. He is the author or coauthor of more than 1,100 publications in these areas, including thirteen books and eleven issued patents. An IEEE Fellow, Vince is also a member of the U. S. National Academy of Engineering and the U. S. National Academy of Sciences, a Fellow of the American Academy of Arts & Sciences, and an International Fellow of the Royal Academy of Engineering of the U. K. He received a Guggenheim Fellowship in 2002 and the IEEE Education Medal in 2005. Recent recognition of his work includes ComSoc's 2009 Armstrong Award and the 2010 IET Fleming Medal. This month, he will receive the 2011 IEEE Eric E. Sumner Award, and an honorary doctorate from the University of Edinburgh.

There are three basic types of Society-level awards: paper, career, and service. Usually, the recipients of paper awards are selected during the early part of each year and are presented at the IEEE International Conference on Communications (ICC), which is typically held in June. The recipients of career and service awards are selected in the second part of each year for presentation at the IEEE Global Communications Conference (GLOBECOM), which is typically held in late-November or early-December. The oldest of these awards is a career award — the Edwin Howard Armstrong Achievement Award — which was first awarded in 1958; the newest is a paper award — the Heinrich Hertz Award for Best Communications Letter — which was first awarded in 2010. As you can see, both of these awards honor not only the recipients, but also are named in honor of distinguished contributors to our field. This is also true of many of our other awards, which we will describe below.

THE AWARDS

COMSOC PAPER AWARDS

Best Tutorial Paper Award: This award is given to an outstanding tutorial paper published in any Communications Society magazine or journal in the previous calendar year.

The Fred W. Ellersick Prize: This award is for an outstanding paper published in any Communications Society magazine.

The Heinrich Hertz Award for Best Communications Letter: This award is for letters published in *IEEE Communications Letters* that essentially enlarge the field of communications engineering. Again, this is the newest award given by ComSoc.

The Leonard G. Abraham Prize in the Field of Communications Systems: At present, this award is restricted to a paper published in *IEEE Journal on Selected Areas in Communications*.

The Stephen O. Rice Prize in the Field of Communications Theory: At present, this award is restricted to a paper published in the *IEEE Transactions on Communications*.

The William R. Bennett Prize in the Field of Communications Networking: At present, this award is restricted to a paper published in the *IEEE/ACM Transactions on Networking*.

The Outstanding Paper on a New Communication Topic: This award is given for an outstanding new-topic paper in any Communications Society publication.

The ComSoc & Information Theory Joint Paper Award: This award is for a paper, relevant to both named Societies, which appeared in any publication of the Communications Society or the Information Theory Society within the past three years. (This is a joint award with the IEEE Information Theory Society, administered in alternate years by one of the two Societies.)

The IEEE Marconi Prize Paper Award in Wireless Communications: This is a joint award with the IEEE Signal Processing Society, and is restricted to a paper published in the *IEEE Transactions on Wireless Communications*.

Starting in 2011 the time window of eligibility for most of these awards is that the papers should have been published in the previous three calendar years. Two exceptions to the three-year rule are the Best Tutorial Paper Award, which has a five-year eligibility window, and the Outstanding Paper on a New Communication Topic,

THE PRESIDENT'S PAGE

ComSoc President Byeong Gi Lee and ComSoc Awards Chair Vince Poor present the 2010 Joseph LoCicero Award to Larry Greenstein at GLOBECOM 2010.

which currently has a one-year eligibility window. For most of the awards with current three-year windows of eligibility, this window of eligibility was recently increased from one year to three years. The purpose of this increase was to enrich the pool of nominations, to allow two award-worthy papers appearing in the same year to receive an award, and to level the playing field for papers appearing late in a given year, which had made it difficult to determine a paper's impact in time to influence the awards process under the previous restriction. In 2012, some further changes are anticipated for these eligibility conditions, as approved by the ComSoc Board of Governors at its December 2010 meeting and subject to approval by the IEEE Technical Activities Board Awards and Recognition Committee. In particular, the Abraham, Bennett, and Rice Prizes, will become "field" awards, for outstanding papers in Communication Systems, Communication Networks, and Communication Theory, respectively, without restriction to the publication in which candidate papers appeared. These latter changes were made as a response to the recognition that outstanding contributions to these fields are not limited to publication in the journals to which the awards were previously constrained. These changes also should further enrich the pool of nominees for these awards, and again create a more level playing field for all outstanding papers. Finally, also in 2012, the Outstanding Paper on a New Communication Topic award will be renamed the IEEE Communications Society Award for Advances in Communications, and will have a fifteen-year window of eligibility. The purpose for renaming this award is to reflect the intent of the award more accurately, and the enlargement of the window of eligibility is a response to the fact that the importance of a new idea is often not recognized until much later than the original publication of the idea.

With only two exceptions, each Paper Award has been given to only one paper per year. For one of these exceptions, the recipient was a two-part paper. The other exception was the first awarding in 2001 of the ComSoc & Information Theory Joint Paper Award, which was given to two separate sets of authors (coincidentally, Vince was one of these authors).

COMSOC CAREER AWARDS

The Edwin Howard Armstrong Achievement Award: This award is given to outstanding contributions over a period of years in the field of communications technology. Again, this is the oldest award given by ComSoc. The recipient must be a member of ComSoc.

The Distinguished Industry Leader Award: This award is given for executive leadership resulting in major advances and new directions in the information and communications business area.

The Industrial Innovation Award: This award is given for major industrial accomplishments, standards, deployment of important processes or products, etc., that are of substantial benefit to the public in the field of communications and information technologies and visible beyond the company or institution where the contribution was made.

The Award for Public Service in the Field of Telecommunications: This award is given for major contributions to the public welfare through work in the field of telecommunications.?

COMSOC SERVICE AWARDS

The Donald W. McClellan Meritorious Service Award: This award is given for outstanding long-term service and leadership in the welfare of the IEEE Communications Society. The recipient must be a member of ComSoc.

The Harold Sobol Award for Exemplary Service to Meetings & Conferences: This award is given for exemplary service to IEEE Communications Society meetings and conferences over a sustained period of time.

The Joseph LoCicero Award for Exemplary Service to Publications: This award is given for exemplary service to IEEE Communications Society publications over a sustained period of time.

The ComSoc/KICS Exemplary Global Service Award: This award is given for fostering successful partnerships between ComSoc and its sister societies and for encouraging a spirit of mutual support and respect in the international communications community.

The Career and Service Awards are also typically given only to one person in a given year. Exceptions are the 1973 Edwin Howard Armstrong Award, which was given to both Samuel G. Lutz and Claude E. Shannon; the 2004 Distinguished Industry Leader Award, which was given to both Ki Tai Lee and Irwin M. Jacobs; and the Donald W. McClellan Meritorious Service Award, which has been given to two recipients in several years.

The ComSoc/KICS Exemplary Global Service Award is unique in that it is co-sponsored by ComSoc and the Korea Information and Communications Society (KICS), our sister society located in Korea. This award is presented jointly by the presidents of these two societies.

SELECTION PROCESS

The quality of the awards process relies on two important member activities: nominating worthy papers or individuals for awards, and service on the Awards Committee. Nominations for paper awards are usually due on the first of March each year, with nominations for career and service awards due on the first of September. Nominations for one of the career awards, the Distinguished Industry Leader Award, can also be submitted by February 15 for consideration in the spring. In addition to nominations directly from the membership, nominations for some paper awards also come from the editorial boards of ComSoc publications or from ComSoc Technical Committees. Overall, this process, which is a tremendous service to the Society, involves a great deal of volunteer time and effort.

For 2011, the ComSoc Awards Committee consists of 13 members, serving three years in staggered terms. In choosing the members of this committee, an effort is made to represent the Society broadly in terms of geographic distribution, technical interests, and type of work affiliation, although qualities such as experience, judgment, and a spirit of service are, of course, paramount.

The Committee does much of its work in subcommittees. Initial consideration for each award, whether it is a Paper Award, a Career Award, or a Service Award, is done by a subcommittee of the Awards Committee dedicated to that particular award and appointed by the Awards Committee Chair. The recommendations of these subcommittees are then considered by the entire Awards Committee before final selections are made. Not every award is given every year, and whether an award is given in a particular year depends on the overall pool of nominations that year.

AWARDS CEREMONY

The culmination of the ComSoc awards process, and of course the most important part of it, takes place at the semi-annual ComSoc Awards Ceremonies, held at the Society's flagship conferences, the IEEE International Conference on Communications (ICC) and the IEEE Global Communications Conference (GLOBECOM). As noted above, Paper Awards are typically presented at ICC and Career and Service Awards at GLOBECOM. However, this is not an ironclad rule, as the presentation venues depend on award recipients' ability to attend these conferences. The Awards Ceremony is also a venue for honoring ComSoc members who have been elevated to the rank of IEEE Fellow, as well as for presentations of Best ComSoc Chapter awards, of the Best Paper Award of the *Journal on Communications and Networks*, of Institute-level field awards whose recipients choose this as a venue for these presentations, and of plaques honoring key ComSoc volunteers.

NOMINATE!

By far, the most critical part of the ComSoc Awards process is that of nominating worthy papers and individuals for the awards. We encourage all ComSoc members to take part in this process. Think of the papers you have read that have had an impact on your work, and of the people you know who have made major contributions to our field, and nominate them! The process is simple, and further details can be found on the ComSoc Awards webpage:

http://www.comsoc.org/about/memberprograms/comsoc-awards

You may see someone whose contributions you have nominated appear on the stage at an Awards Ceremony soon!

COMSOC 2011 ELECTION
TAKE TIME TO VOTE

Ballots were mailed and emails with election information were sent 31 May 2011 to all Higher Grade* IEEE Communications Society Members and Affiliates (excluding Students) whose memberships were effective prior to 1 May 2011.

To cast your ballot electronically you will need your IEEE Web Account username/password — which is the same account information used to access IEEE online services such as renewing your membership, myIEEE, and Xplore. If you do not recall your web account information, you may go to: www.ieee.org/web/accounts to recover. You may also email contactcenter@ieee.org or call +1 800 678 4333 (USA/Canada) or +1 732 981 0060 (Worldwide).

If you do not receive an email from ieee-comsocvote@ieee.org on 31 May 2011 or a paper ballot by 30 June, but you feel your membership was valid before 1 May 2011, you may e-mail ieee-comsocvote@ieee.org or call +1 732 562 3904 to check your member status. (Provide your member number, full name, and address.)

Please note IEEE Policy (Section 14.1) below stating IEEE mailing lists should not be used for electioneering in connection with any office within the IEEE:

> IEEE membership mailing lists, whether obtained through IEEE Headquarters or through any IEEE organizational unit, may be used only in connection with normal IEEE sponsored activities and may be used only for such purposes as are permitted under the New York Not-For-Profit Corporation Law. They may not be used for electioneering in connection with any office within the IEEE, or for political purposes, or for commercial promotion, except as explicitly authorized….

See details at url http://www.ieee.org/web/aboutus/whatis/policies/p14-1.html.

Voting for this election closes 26 July 2011 at 4:00 p.m. EDT! Please vote!

*Includes Graduate Student Members

CERTIFICATION CORNER

THE SPRING 2011 AND FUTURE EXAMS

BY ROLF FRANTZ

IEEE Communications Society
Wireless Communication Engineering Technologies

As this is written, a group of wireless professionals are preparing to gather to review the Spring 2011 WCET examination, for which the testing window recently closed. These experts do not study individual scores, but rather the overall exam, the questions, the difficulty, and the standard of performance that might be expected of qualified individuals taking the test. This process is repeated for each WCET exam: experts in the field review each question on the exam, discuss the level of knowledge and skill that would be required to answer it correctly, and develop a consensus as to how many correctly answered questions constitute an appropriate expectation for passing the exam. This is important because each offering of the WCET exam is somewhat different – questions are replaced to keep the exam current, and also to remove questions that no one (or everyone) could answer correctly. Once the passing standard is determined, the candidates who took the exam will be notified of their scaled scores and whether they passed or not. That will conclude the process for this Spring's exam and start the process for the Fall 2011 exam.

The analysis of the Spring exam may reveal that certain questions are too easy – or too difficult – to be useful in differentiating among highly skilled, somewhat competent, and poorly prepared candidates. The value of the WCET exam lies in determining that individuals have both a broad understanding of wireless communications and in-depth knowledge in specific areas. The exam must discriminate – usually a word with negative connotations, but in this case important in identifying qualified candidates and ensuring that they, and only they, earn the IEEE Wireless Communications Professional® credential. If the review of the Spring results uncovers such overly easy (or difficult) questions, they will be deleted. New questions that address the same general topic(s) will be drawn from the question bank that has been built up for future exams, and these will replace the deleted questions.

If any questions are changed, the glossary of terms and equations that is available to candidates online during the exam will also be reviewed. If the new questions require any definitions or formulas that are not currently in the glossary, those will be added and the revised glossary will be distributed along with the refreshed examination to all testing centers where the exam can be taken.

This may sound like a lot of work, and indeed it does draw upon numerous volunteers for a significant time commitment. But to meet the goals of the WCET program, and to assure employers, potential employers, and candidates alike of the value of the certification, it is essential that we continually monitor the exam, candidates' performance on it, and any issues affecting its relevance to current industry practice. We are sometimes asked why the exam is only offered during specified month-long windows in the Spring and Fall of each year. One reason is the effort just described: it takes time to analyze each exam, determine how effective it is in distinguishing between qualified and unqualified candidates, and make any needed changes prior to the next testing window.

While on the subject of the next testing window, the Fall 2011 examination window runs from 2 October through 29 October; applications must be completed no later than 16 September. For those who want a little more time to prepare, the Spring 2012 testing window has been set. It will run from 9 April through 5 May; application can be made online now, and the application window will be open through 23 March 2012. It is not too early to start thinking about improving your skills and demonstrating to your colleagues and your employer that you are qualified as an IEEE Wireless Communications Professional®.

CONFERENCE REPORT

IEEE DYSPAN 2011 EXPANDS THE RESEARCH AND DEPLOY OF NEXT GENERATION SMART RADIO TECHNOLOGIES IN AACHEN, GERMANY

The IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN 2011), the leading international conference dedicated to the advance of cutting-edge wireless technologies, held its 5th annual meeting from May 3 – 6 in Aachen, Germany with hundreds of international experts addressing the multidisciplinary challenge of delivering the next wave of DSA technologies from the laboratory to the global telecommunications marketplace.

Launched in 2005 by the IEEE Communications Society (ComSoc), IEEE DySPAN is now recognized worldwide for its influence on technology research and policy development issues throughout the United States, Europe and Asia. This includes building healthy ecosystems associated with the international commercialization of smart radio systems as well as the enhanced utilization of "white spaces" and continued investigation of decentralized spectrum access topics.

On Tuesday, May 3, IEEE DySPAN 2011 renewed this effort with a full day of tutorials dedicated to topics such as "Cognitive Wireless Networking with WARP," "Advanced Antennas for Cognitive Radio: New Foundations for Transformational System Design," "Dynamic Spectrum Access Related Standards," "Dynamic Spectrum Markets" and "Non-Contiguous Multicarrier Transmission for Spectrally Opportunistic Wireless Access: Design Decisions And Trade-Offs."

The following day, the conference proceeded with its next stage of presentations, including the morning keynote addresses of several noted international DSA policy and research experts. Reiner Liebler, the head of Division for Technical Regulations and EMC in Federal Network Agency (BNetzA) in Germany, spoke at-length about how the "Cornerstones of a Forward-Looking Regulatory Framework" should be designed to "foster investment and innovation," in addition to strengthening competition. This includes aligning regulations with competition as a means for enhancing "economical performances" and "furthering efficient investments in mobile and fixed broadband."

David Cleevely, chairman for Cambridge Radio Frequency Services (CRFS) in the United Kingdom, and Douglas Sicker, chief technology officer for the United States Federal Communications Commission (FCC), then followed with their respective thoughts on "Radio Spectrum and Innovation: Realizing the Potential" and "Dynamic Spectrum Access Policy in the U.S." In his keynote, Cleevely discussed the need for "regulatory liberalization" as a means for driving the latest technologies in a radio spectrum marketplace, which is "doubling every 30 months," while Sicker emphasized broader policies for "enabling the availability of additional spectrum for broadband services."

After a brief Q&A with each speaker, IEEE DySPAN 2011 then initiated the first of three days of technical, business and policy symposia consisting of numerous high-level business panels, nearly 60 paper presentations and more than 25 new technology posters and demonstrations. From Wednesday through Friday, these sessions also included the detailed discussions of wide-ranging DSA and cognitive radio domain subjects extending from advanced and spectrum engineering and QoS provisioning to new measurement and sharing modeling, radio resource management and broadband spectrum sensing.

On Thursday, Pearse O'Donohue, who is responsible for the development and implementation of efficient spectrum use policies as the head of the Radio Spectrum Policy Unit for the European Commission in Belgium, began the morning with his keynote on "An EU Policy: Framework for Shared Use of Spectrum" and a brief introduction of the EU's coordinated approach to frequency management. Afterwards, Jon Peha, the former assistant director of the White House Office of Science and Technology Policy, offered attendees his insights into "Emerging U.S. Spectrum Policy and the Road to Innovation." According to Peha, these priorities include "making 500 MHz of federal and non-federal spectrum available over the next 10 years," allocating billions of dollars toward ensuring 98 percent of the American population have wireless access, fostering the confidence necessary to develop secondary markets and specifying how the spectrum will be licensed, allocated and utilized by 2016.

Attendees connected with peers and presenters during the conference networking breaks.

Another Thursday highlight entailed the high-level discussion titled "Regulatory Perspectives on Next Generation Radio Systems." Throughout this panel, representatives of worldwide regulatory agencies elaborated on the architectural issues necessary for "ushering the next generation of databases, sensing technologies and TV white space applications" as well as helping industry to go from R&D to policy to the successful business stage. Among the conclusions of the panelists was the need to develop interoperable networks generated through the dynamic cooperation of standards organizations, private industry and public regulators as well as the building of new patterns that anticipate future demands, define usage rights and respond more quickly to dynamic procedures.

In the evening, attendees enjoyed the conference's annual banquet, which was held in the historic Aachen City Hall, where the kings of Germany traditionally held their coronations. General co-chair Petri Mähönen welcomed everyone and then introduced Margrethe Schmeer, the deputy mayor of Aachen City, who highlighted the Coronation Hall's long historic significance.

Afterwards, demonstration co-chair Przemyslaw Pawelczak presented the Best Demonstration Award to the entry entitled "Constructing Radio Environment Maps with Heterogeneous Spectrum Sensors." The recipients of this honor included Vladimir Atanasovski, Daniel Denkovski, Liljana Gavrilovska, Mihajlo Pavloski and Valentin Rakovic of Ss. Cyril and Methodius University, Skopje, Macedonia; Jaap van de Beek of Huawei Technologies, Sweden; Antoine Dejonghe of IMEC, Belgium; Sébastien Grimoud of France Telecom R&D; Janne Riihijärvi of RWTH Aachen University, Germany; and Berna Sayrac of Orange Labs, France.

In addition, other award honorees included Xi Zhang, Junaid Ansari, Petri Mähönen and Guangwei Yang of RWTH Aachen University, Germany, who received an Honorable Mention for the demonstration named TRUMP: Efficient Realization of PHY/MAC Protocols for Cognitive Radio Networks.

(Continued on page 16)

CONFERENCE REPORT

(Continued from page 14)

IEEE DySPAN 2011 then concluded on Friday, May 6 with another full agenda of keynotes, business panels, technology and policy track sessions. Krishan Sabnani, vice president of Networking Research at Bell Labs, began the day with his keynote titled "Spectrum and Infrastructure Virtualization for Next-Gen Cellular Networks." Sabnani's presentation pointedly highlighted the need to increase capacities based on the exponential growth of wireless, mobile traffic that is tripling in some markets every year and is also expected to entail 2.5 billion Smartphone connections by 2015.

Immediately afterwards, Victor Bahl, the principal researcher and director at Microsoft Research's Mobile Computing Research Center, followed this keynote with his own expert insights into "The Promises and Challenges of the Wireless Frontier – from 600 MHz to 60 GHz" and the development of the "world's first urban space network." According to Bahl, we must actively consider harvesting unused spectrum and developing greater connectivity options over unlicensed frequencies to better manage the spectrum needs of densely populated areas existing with a lower number of available channels.

The remainder of the morning's activities were then punctuated by the executive panels on "Perspectives on Cognitive Radio: The Past and Next 10 Years" and "Business Perspectives on Dynamic Spectrum Access." During each session, industry experts from leading research institutions and telecommunications companies worldwide provided their professional opinions for reducing wireless network management costs, increasing the useful content of transmissions and sharing spectrum in new cooperative ways. The business perspectives panel was also particularly informative due to the lively responses of the representatives from Ericsson, France Telecom, Toshiba, and Kolodzy Consulting to the numerous audience questions and moderation of Pierre de Vries from Silicon Flatirons Center. Among their recommendations for future communications were the building of "self-configuring and self-optimizing networks" that "reuse spectrums on a much more fine-grain basis" and "predict the emergent behaviors of interacting systems of multiple cognitive radios."

Other 2011 conference highlights included the introduction of "2-minute madness" as a new and novel method for helping presenters to discuss their demonstrations and posters as well as the well-attended plenary sessions, which provided valuable research details into topics such as "UHF White Space in Europe," "SenseLess: A Database-Driven White Spaces Network," "Spectrum Requirements for TV Broadcast Services using Cellular Transmitters," "Robust Cooperative Sensing via State Estimation in Cognitive Radio Networks," "Evaluating the Economic Impact of Cognitive Radio with a Three-player Oligopoly Model" and "Spatio-Temporal Spectrum Holes and the Secondary User."

As for IEEE DySPAN 2012, the 6th annual event will be held April 3 – 6, 2012 in Bellevue, Washington. For more information, including "Call for Papers details," which ends November 2, 2011, please visit www.ieee-dyspan.org/2012 or contact Heather Ann Sweeney of the IEEE Communications Society (ComSoc) at 212-705-8938 or h.sweeney@comsoc.org. In addition, interested parties are also welcome to follow IEEE DySPAN 2012 happenings or reach out to international colleagues via links to Twitter, LinkedIn and Facebook on the conference website.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

EMBEDDED WITH INNOVATION: Workload Consolidation

Traditional designs have multiple processor architectures

In traditional designs, a single network device such as a radio network controller or a wireless base station requires multiple processing architectures to handle unique workloads.

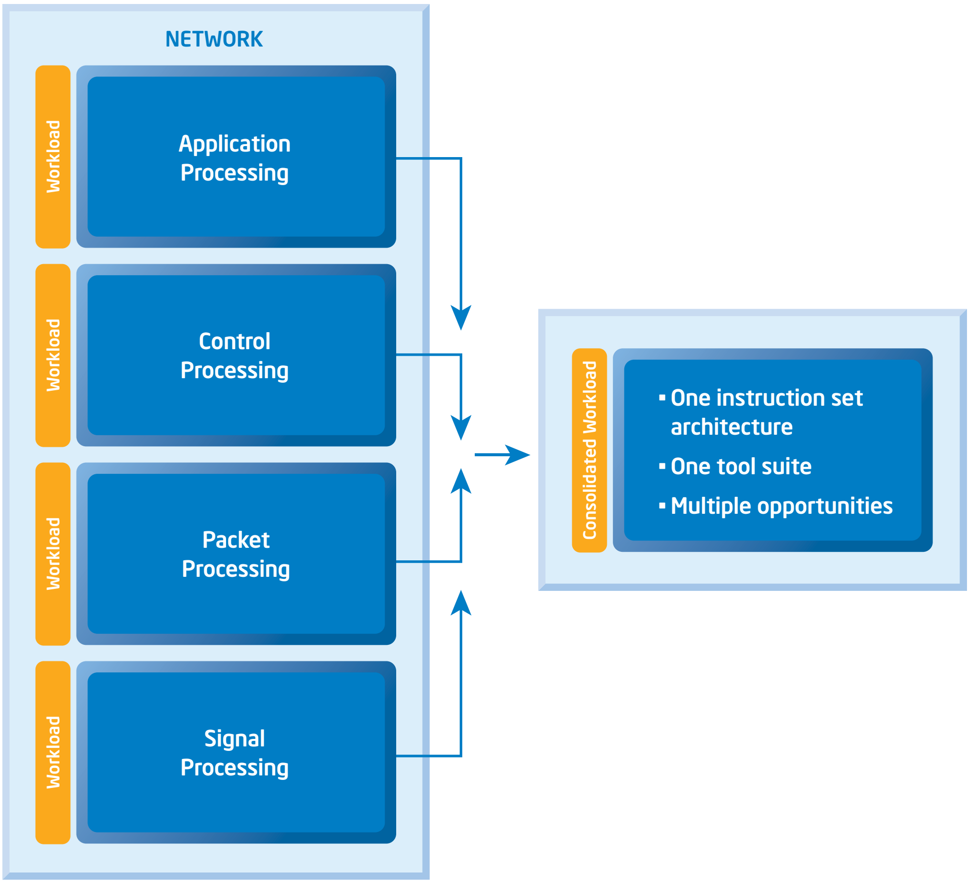
Look under the hood and you find general-purpose processors for application and control plane processing, network processors for packet processing and an alphabet soup of dedicated chips for signal processing. Just one base station can have boards with CPUs, NPUs, DSPs, FGPAs, ASICs and ASSPs, each requiring its own firmware and software.

Intel® architecture enables 4:1 workload consolidation

Intel's scalable family of code-compatible Intel® Xeon® processors has the flexibility to meet top-to-bottom performance requirements of network equipment. So now you can consolidate applications and services processing, control plane processing, packet processing and signal processing. And you can do it all on a single scalable architecture that's adaptable to a broad range of products.

>>>

The Comms Infrastructure 4:1 framework enables an Intel® processor-based platform to simultaneously execute four workloads.



READ > WATCH > SHARE >>> intel.com/go/commsinfrastructure

IEEE Communications MAGAZINE

CALL FOR PAPERS

IEEE Communications Magazine encourages the submission of technical papers and articles on topics of interest and relevance to its readers, which includes all members of the IEEE Communications Society. The scope of the articles relates to various types of telecommunications and data communications, including:

- **Technology (e.g., optical, wireless, signal processing)**
- **Networking (e.g., LAN, WAN, ad hoc, automotive)**
- **Systems**
- **Services and service management**
- **Emerging standards**
- **Development methods**
- **Deployment experiences and field trial results**
- **Market trends**
- **Regulatory and policy issues**
- **Significant global events and projects**

Articles should be tutorial in nature and written in a style comprehensible to readers outside the specialty of the article. Articles should contain a minimum of mathematical content and avoid product marketing content. The author guidelines for additional information on content and style can be found on the magazine web site http://dl.comsoc.org/ci/. The call-for-papers links for current feature topics and regular series can also be found at this web site. Articles not directly aligning with the posted calls for papers are welcome as open call submissions. All papers and articles go through a peer review process to ensure high quality content and industry relevance.

All submissions and their subsequent processing is handled through the on-line Manuscript Central tool: http://mc.manuscriptcentral.com/commag-ieee Author guidelines and forms can also be found on this site.

CONFERENCE CALENDAR

2011

JUNE

♦ **IEEE IWQoS 2011 - 18th IEEE Int'l. Workshop on Quality of Service, 5-7 June**
San Jose, California.
http://www.ieee-iwqos.org/

♦ **IEEE ICC 2011 - IEEE Int'l. Conference on Communications, 5-9 June**
Kyoto, Japan.
http://www.ieee-icc.org/2011/

• *IEEE POLICY 2011 - IEEE Int'l. Symposium on Policies for Distributed Systems and Networks, 6-8 June*
Pisa, Italy.
http://www.ieee-policy.org/

♦ **IEEE CAMAD 2011 - IEEE Int'l. Workshop on Computer-Aided Modeling Analysis and Design of Communication Links and Networks 2011, 10-11 June**
Kyoto, Japan.
http://www.nprg.ncsu.edu/camad/

♦ **IEEE HEALTHCOM 2011 - 13th IEEE Int'l. Conferece on e-Health Networking, Application & Services, 13-15 June**
Columbia, MO.
http://www.ieee-healthcom.org/

• *ConTEL 2011 - 11th Int'l. Conference on Telecommunications, 15-17 June*
Graz, Austria.
http://www.contel.hr/

• *ICUFN 2011 - 3rd Int'l. Conference on Ubiquitous and Future Networks*
Dalian, China.
http://www.icufn.org/main/

♦ **IEEE CTW 2011 - IEEE Communication Theory Workshop, 20-22 June**
Sitges, Spain.
http://www.ieee-ctw.org

♦ **IEEE SECON 2011 - 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, 27-30 June**
Salt Lake City, Utah.
http://www.ieee-secon.org/2011/

• *IEEE ITMC 2011 - IEEE Int'l. Technology Management Conference, 27-30 June*
San Jose, CA.
http://www.ieee-itmc.org/

• *SPECTS 2011 - 2011 Int'l. Symposium on Performance Evaluation of Computer and Telecommunication Systems, 27-30 June*
The Hague, Netherlands.
http://atc.udg.edu/SPECTS2011/

♦ **IEEE ISCC 2011 - 16th IEEE Symposium on Computers and Communications, 28 June-1 July**
Kerkyra, Greece.
http://www.ieee-iscc.org/2011/

• *ICL-GNSS 2011 - Int'l. Conference on Localization and GNSS 2011, 29-30 June*
Tampere, Finland.
http://www.icl-gnss.org/2011/index.php

JULY

♦ **IEEE HPSR 2011 - 12th IEEE Int'l. Conference on High Performance Switching and Routing, 4-6 July**
Cartagena, Spain.
http://www.ieee-hpsr.org/

• *OECC 2011 - 16th Opto-Electronics and Communications Conference, 4-8 July*
Kaoshlung, Taiwan.
http://www.oecc2011.org/

♦ **IEEE ICME 2011 - 2011 IEEE Int'l. Conference on Multimedia and Expo, 11-15 July**
Barcelona, Spain.
http://www.icme2011.org/

AUGUST

• *ICCCN 2011 - Int'l. Conference on Computer Communications and Networks 2011, 1-4 Aug.*
Maui, Hawaii.
http://www.icccn.org/ICCCN11/

♦ **ATC 2011 - 2011 Int'l. Conference on Advanced Technologies for Communications, 3-5 Aug.**
Da Nang City, Vietnam.
http://rev-conf.org/

• *ICADIWT 2011 - 4th Int'l. Conference on the Applications of Digital Information and Web Technologies, 4-6 Aug.*
Stevens Point, WI.
http://www.dirf.org/DIWT/

• *ITST 2011 - 11th Int'l. Conference on ITS Telecommunications, 23-25 Aug.*
St. Petersburg, Russia
http://www.itst2011.org/

♦ **IEEE P2P 2011 - IEEE Int'l. Conference on Peer-to-Peer Computing, 31 Aug.-2 Sept.**
Tokyo, Japan.
http://p2p11.org/

♦ **IEEE EDOC 2011 - 15th IEEE Int'l. Enterprise Distributed Object Computing Conference, 31 Aug.-2 Sept.**
Helsinki, Finland.
http://edoc2011.cs.helsinki.fi/edoc2011/

• *FITCE 2011 - 50th FITCE Congress - ICT: Bridging the Ever Shifting Digital Divide, 31 Aug.-3 Sept.*
Palermo, Italy.
http://www.fitce2011.org/

SEPTEMBER

• *ITC 23 2011 - 2011 Int'l. Teletraffic Congress, 6-8 Sept.*
San Francisco, CA.
http://www.itc-conference.org/2011

♦ **IEEE PIMRC 2011 - 22nd IEEE Int'l. Symposium on Personal, Indoor and Mobile Radio Communications, 11-14 Sept.**
Toronto, Canada.
http://www.ieee-pimrc.org/2011/

• *ICUWB 2011 - 2011 IEEE Int'l. Conference on Ultra-Wideband, 14-16 Sept.*
Bologna, Italy.
http://www.icuwb2011.org/

• *ICCCT 2011 - 2nd Int'l. Conference on Computer and Communication Technology, 15-17 Sept.*
Allahabad, India.
http://www.mnnit.ac.in/iccct2011/

• *SoftCOM 2011 - Int'l. Conference on Software, Telecommunications and Computer Networks, 15-17 Sept.*
Split, Croatia.
http://marjan.fesb.hr/SoftCOM/2011/index.html

♦ **IEEE GreenCom 2011 - Online Conference, 26-29 Sept.**
Virtual.
http://www.ieee-greencom.org/

♦ Communications Society portfolio events are indicated with a diamond before the listing;
• Communications Society technically co-sponsored conferences are indicated with a bullet before the listing. 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: b.erlikh@comsoc.org; fax: +1-212-705-8996. Items submitted for publication will be included on a space-available basis.

New Products

Edited by Eric Levine

Digital Oscilloscope

Rohde & Schwarz

Rohde & Schwarz has developed a digital oscilloscope that employs a real-time digital trigger. Designated the R&S®RTO Series, the advanced scope significantly enhances productivity and product performance when debugging embedded systems. The real-time digital trigger in the R&S®RTO Scope Series uses a common signal path for both the trigger and acquired data. This eliminates time and amplitude offset between the trigger and signal, enabling signals to be displayed with the least possible trigger jitter and allowing precise results to be achieved when measuring complex waveforms.

Analog triggers typically have a long re-arm cycle, and while re-arming, the instrument cannot react to trigger events, which means signal properties that should act as triggers are masked. Because the R&S®RTO Scope Series' digital trigger does not need to re-arm, every sample can trigger data acquisition to avoid missing events.

In addition, errors in serial interfaces are often caused by sporadic signal faults at the physical layer, so high acquisition rates are essential in order to detect them. Since R&S®RTO Scope Series acquires and displays signals using hardware, blind time is minimized and errors are located quickly. Also, many embedded systems employ serial data interfaces to control external devices and simply viewing the waveform is not sufficient to fully debug the operation of these systems. The R&S®RTO features an integrated trigger and decode feature that allows users to trigger on protocol-specific features and to display the captured waveform in binary, ASCII or HEX formats. The trigger and decode feature supports SPI, I2C, CAN, LIN and RS232.

http://www.rohdeschwarz.com/usa

Bare Metal Performance Tools for NetLogic Microsystems' XLP@ Multi-Core, Multithreaded Processors

ENEA

Enea® has introduced a tools solution for "bare metal" multicore implementations on the NetLogic Microsystems XLP® multi-core processor, called the Enea® Bare Metal Performance Tools (Enea BMP Tools). The term "bare metal" refers to multicore applications whereby a given processor core executes a function or application without any sort of multitasking executive, usually a simple control loop that runs in an OS-free environment for minimal overhead and maximum processing bandwidth.

Enea Bare Metal Performance Tools consists of an Eclipse-based host tools suite called Enea® Optima, a set of runtime libraries and agents for data collection, and an IPC mechanism called Enea® L1NX for transport of the collected profiling and logging data to the Optima host tool, or to an external file for later analysis. The Enea Bare Metal Performance Tools solution is designed as an extension to NetLogic Microsystems' SDK so that all characteristics of the SDK's software build, boot/load, and bare metal execution environment (NetOS) performance are maintained. Enea Bare Metal Performance Tools focus on the two most useful types of runtime tools that can aid developers for performance optimization: a) performance profiling, and b) logging.

Profiling helps developers optimize a slow performing application by visualizing runtime hardware constraints caused by the non-optimized source code. Enea Bare Metal Performance Tools provide two types of performance profiling visualization tools:

• Source Code Profiling: identifies at the source code level where constraints such as pipeline stalls, TLB misses and cache misses are causing sub-optimal performance, by matching these hardware events/counters to the source code at any level of the application function call tree, even down to individual lines of code in any given function. This includes overall CPU utilization of any function or line of code. Bare Metal Performance Tools offer complete user configuration control of all NetLogic Microsystems' XLP multi-core processor hardware counters.

• Application Profiling: creates and analyzes application software level statistics with the purpose of profiling the applications overall performance and behavior. Such statistics could be idle time, throughput statistics, extraneous hardware events, or any other user defined statistics that makes sense to the application.

http://www.enea.com

Atheros Launches HomePlug Green PHY Emulation Platform

Atheros Communications

Atheros Communications has introduced their HomePlug® Green PHY (HPGP) emulation development kit. The Atheros PL-14 HPGP development kit consists of a hardware platform based on existing HomePlug AV silicon, adapted to support serial peripheral interface (SPI) ports enabling connectivity to a vast family of low-cost and low-energy microcontrollers (MCUs), as well as an emulation firmware and software driver with application programming interfaces (APIs) to support future Atheros Green PHY products. By accurately emulating a Green PHY environment, the development kit enables Atheros customers to evaluate the technology's capabilities and develop application software for "Internet of Things" product families focused on smart grid and smart energy.

The PL-14 HPGP development kit is the first in a series of products within the Atheros Internet of Things product portfolio, which includes standards-based wired and wireless technologies to enable scalable IP infrastructures for smart grid, smart home, security, building automation, remote health and wellness monitoring, and other machine- to-machine (M2M) applications. With its extensive technology portfolio and Internet Protocol (IP) networking expertise, Atheros is uniquely positioned to deliver a variety of low-energy, standards-based communications solutions that connect potentially hundreds, if not thousands, of IP addressable devices in the home.

To enable the Internet of Things, Atheros is focused on providing technologies and solutions which are:

Scalable: enabling future growth in number of connected devices, enhanced security and reduced delay time.

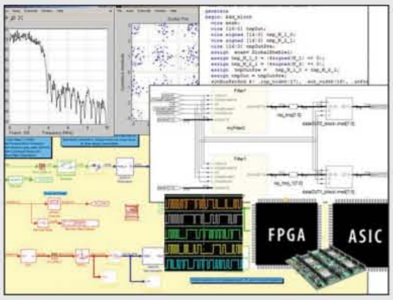
Standards-based: supporting global interoperability to improve user experience and value.

Natively supporting IP communications: leveraging the values of the IP network to connect "islands" of disparate and incompatible systems.

HomePlug Green PHY (HPGP) is a subset standard of the widely deployed HomePlug AV powerline standard. The Green PHY specification enables a low-power, low-energy and robust connectivity solution, while supporting data rates up to 10 Mbps for scalable IP connectivity. The result supports network build-out with short response times for devices used in smart grid and smart home applications, such as appliances, thermostats, electricity meters, plug-in electric vehicles (PEV), and monitoring equipment.

http://www.atheros.com

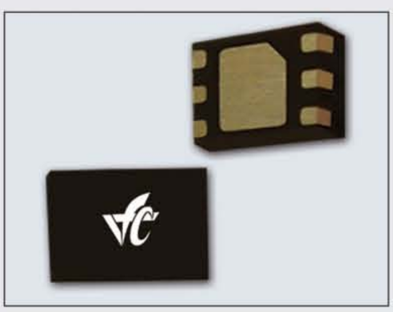
Product Spotlights



Synopsys

Designers today need high-level synthesis optimization technologies that deliver high quality of results for FPGA and ASIC while enabling rapid exploration of performance, power, and area. Synphony High-Level Synthesis (HLS) tools provide an efficient path from algorithm concept to silicon and enable greater design and verification productivity.

http://www.synopsys.com



Valpey Fisher

The VFHY100 series of 90 degree hybrids exhibit very good performance across the frequency range 600 MHz to 4.0 GHz and are offered in a miniature 1.5 x 2.0 mm leadless package. Typical performance is 30 dB of isolation, a 1.1:1 VSWR and phase balance of ±1 degree.

http://www.valpeyfisher.com

RF Micro Devices

RFMD's RF5605 front end module (FEM) for IEEE 802.11b/g/n WiFi CPE applications has an integrated three-stage linear PA, transmit harmonic filtering, and an SPDT switch. The FEM fully matches RF input and output for a 50-ohm system, incorporates matching networks optimized for linear output power (27dBm) and efficiency, and housed in a 6 x 6 mm laminate package.

http://www.RFMD.com

u2t

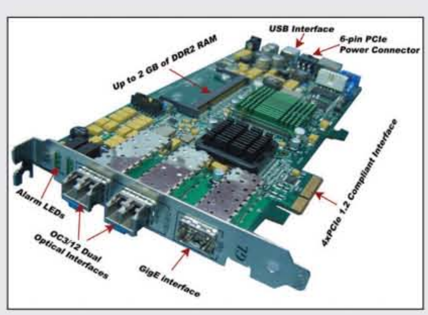
The highly compact and integrated coherent receiver CPRV1220A is suited for 100G DP QPSK applications enabling compact module and line card designs. The receiver offers balanced detection and a linear electrical interface and supports a symbol rate of up to 32Gbaud. It complies with the CCRx-MSA specification.

http://www.u2t.de

Silicon Labs

Learn how to simplify your timing design using glitch-free frequency shifting to address low-power design challenges and the complexity of generating a wide range of frequencies in consumer electronics applications such as audio, video, computing or any application that requires multiple frequencies. Download this in-depth white paper from Silicon Labs.

http://www.silabs.com/frequency-shifting



GL Communications

LightSpeed1000 is GL's new Dual OC3/12 STM1/4 PCI-Express card designed for protocol analysis and emulation of ATM, POS, RAW, and Ethernet traffic. The card is one of the most versatile and powerful cards available in the marketplace. The card sports 2 GigaBytes of DDR2 RAM, a GigE interface, and a USB 2.0 interface. Multiple cards in a rack PC permits analysis of many signals simultaneously. The card is unique in its capability to capture and transmit at wirespeed to/from hard disk on all interfaces. Its ability to capture to disk at full rate in both directions without error permits detailed offline analysis that is not possible with other test instruments. Capture is possible on both optical ports and the single Ethernet port. Other applications include monitoring, BERT, emulation, and protocol analysis.

http://www.gl.com/OC3-OC12-analysis-emulation-card.html

Global Communications Newsletter

June 2011

The Spanish International Campus of Excellence Program

By David Montoro-Mouzo and Juan Pedro Muñoz-Gea, Polytechnic University of Cartagena, Spain

The International Campus of Excellence (ICE) program, promoted by the Ministry of Education of the Spanish Government, is working to improve the quality and excellence of the Spanish universities and their environments. Similar programs have been carried out in France (with the Opération Campus program) and Germany (with the Exzellenzinitiative program). All of them are based on additional funding for the best universities, with the intention of achieving excellence and more international presence. Another common characteristic of the three previous programs is that they consider reinforcing research activities, by supporting Ph.D. schools and research institutions, the key to achieving the proposed objectives. On the other hand, other multinational organizations, like the European Commission, have also promoted similar initiatives.

The ICE program was initially staged in 2009, and it is supported with a budget of €590 for the period 2008–2010. These funds are given in the form of direct subsidies for the campuses involved, and in the form of reimbursable credits for the regional public authorities of the territory where the campus is placed. Using these funds, the campuses and regional authorities are capable of implementing their strategic plans to move toward international excellence. The ICE program is intended to solve the problems caused by the weaknesses of the Spanish university model. The main problem is the geographic atomization of universities: there are 50 public institutions and 28 private centers, scattered across 232 campuses and territorial sites. This situation leads to high maintenance cost and an extensive duplicity of offered courses. From the point of view of research, this scattering means more difficulties in achieving a critical mass of talent. These, and other aspects, have led to none of the Spanish universities being placed among the world's top 200 in the Academic Ranking of World Universities (ARWU), and only eight were in the top 500.

The objectives of the ICE program can be summarized in three main aspects. First, it is intended to improve the international visibility of the best Spanish university campuses. The excellence concept extends beyond academic excellence to excellence in teaching, excellence in research and innovation, and excellence in the physical and social environments of the campuses. Second, it is intended to promote clustering and specialization, that is, universities specializing in only one to three disciplinary areas. This policy is intended to enhance the efficiency, performance, and international competitiveness of Spanish universities. Finally, the third objective is to redefine the concept of campus. In the ICE program, it is intended to improve the physical environments of the campuses, and the manner in which the universities interact with the cities and regions where they are placed. In this context, the objective of creating European Knowledge Regions based on collaboration among Spanish, French and Portuguese universities is especially remarkable. Furthermore, the possibility of creating cross-border campuses is also contemplated.

In order to select the best projects, the institutions have to go through a serious selection process that takes place every year. First, the institutions interested in turning into an ICE have to submit a project for approval by a Technical Committee. This committee selects the best projects, which are then laid before an International Assessment Committee. This international committee of experts chooses the list of ICE projects to be adopted in that year. Finally, the winner projects are funded and developed. In the choice of winner projects different aspects are considered: improvement of teaching, adaptation to the European Higher Education Area, renovation of teaching buildings, the transference of knowledge to businesses, development of the territorial environment, internationalization, and improvement of research activities. Currently there are around 20 projects under development to convert various universities and scientific institutions into Campuses of International Excellence. Among them, there are projects in Barcelona, Madrid, Cartagena, Seville, and Valencia.

In conclusion, it is expected that the program will contribute to improve the teaching and research quality of Spanish universities, as well as to increase their international visibility. Thus far, the ICE program has been a powerful tool to stimulate the process of modernizing the Spanish universities and transforming them into a key factor of the intended shift in the Spanish social and economic model. This will improve employment, social cohesion, and local development of the territories where the ICE campuses are located by the development of a better education system, and a better research and innovation system.

VITEL 2010 International Workshop: Transition to IPv6

By Ana Robnik and Marko Jagodic, EZS – SIKOM, Slovenia

The Slovenian Society for Electronic Communications, EZS–SIKOM, a Sister Society of IEEE Communications Society, has a very long tradition of organizing national and international events on telecommunications in Slovenia. The intention of the events has always been to concentrate on the most important and relevant topics at the time, leading to more efficient and useful or simply viable telecommunications for everybody. This principle also explains the name given to these events, which is VITEL (Viable Telecommunications).

VITEL is guided by the International Advisory Commitee, composed of well-known telecommunication experts primarily from the Central and Eastern European Region and chaired by Marko Jagodic. The selection of papers and creation of the final program is prepared by the Program Committee composed of eminent telecommunication professionals from Slovenia and chaired by Alojz Hudobivnik from Iskratel. The Organizing Committee, responsible for the realization of the event, is different for each event. This one was chaired by Ana Robnik from Iskratel.

The committees decided to select "Transition to IPv6" as the appropriate theme for the VITEL 2010 International Workshop with the objective of bringing together experts from operators, service providers, academia, and industry alike to address this very important topic for the future development of telecommunications and the Internet with fast growing numbers. The event was sponsored by the IEEE Communications Society Chapter of Slovenia, the Ministry for Higher Education, Science and Technology, and the Ministry for Economy.

Multimedia-rich communications in modern converged fixed and mobile networks have triggered a rapid increase in the number of endpoints, particularly mobile users, and rapid expansion of all IP communication infrastructures, boosting demand for IP addresses well beyond the capacity of IPv4 address space. Theoretically only 4.3 million users can be connected with public IPv4 addresses. Therefore, desktops in various organizations, mobile phones, and standalone devices use private address space, connected to a single or a few public IPv4 addresses. With network address translation and usage of private addresses instead of public ones, the third-generation (3G) community and service providers solved on a short-term basis the problems with shortage of IPv4 address space; they got these from either the Internet Assigned Numbers Authority (IANA) or the Regional Internet Registry for Europe (RIPE). As a consequence, network providers have faced many problems with the complexity of managing such networks, visibility of end users in a public network, and collisions among IP addresses in case of merging such networks.

Accelerated transition from IPv4 to IPv6 has thus become one of the most important conditions for the efficient realization of the future Internet, which is also listed among the four priority areas for investment in the Digital Agenda for Europe.

The Workshop took place in the Congress Centre Brdo-Kranj on 19 and 20 April 2010. Mr. Nikolaj Simic, General Director of the Directorate for the Information Society of the Republic of Slovenia, opened the Workshop. In his opening address he discussed the extreme importance of the theme, and its short-term and long-term positive influence on the evolution of future telecommunications and Internet. Presentations for the workshop were prepared by representatives from universities, ministries, the regulator, network operators, providers of services and applications, and equipment vendors and suppliers. The most important foreign participants were Mr. Teemu Savolainen from Nokia and Mr. Marco Hogewoning, co-chair of the RIPE IPv6 Working Group.

Ana Robnik, Chairwoman of the Organizing Committee.

On the first day of the Workshop the status of the transition to IPv6 in Slovenia and around the world was discussed as well as the existing transition strategies of major equipment vendors and network operators together with encountered problems. Although using strategically different approaches for the transition to IPv6 around the world, the achieved results and experience are encouraging. What is worrisome is that the number of countries deciding to make this transition is far smaller than expected. The main reasons are rather high expenses, which are difficult to justify, and lack of interest in make the transition through the whole chain, starting with users and providers of services and applications. A big deficiency is also the lack of backward compatibility between IPv6 and IPv4. Therefore, it is extremely important that a very detailed and systematic plan be in place before the actual transition starts.

The first day ended with a roundtable entitled "Transition to IPv6 — Waiting for Godot?" The roundtable, moderated by Dr. Mitja Štular, was very provocative and inspiring. One of the most important results was the decision to prepare a document for the Directorate for the Information Society of the Republic of Slovenia, which will serve as a basis for its future activities in this area.

On the second day examples of good practice were presented and discussed, based on different strategies, different network and service environments, and different business or technology initiatives. The workshop was concluded with the presentation of transition mechanisms Dual Stack Light, 6RD, A+P, and others. More than 150 participants attended the workshop.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

1st Alexander Graham Bell Memorial Lecture by Tapan K. Sarkar in Hyderabad

By Deergha Rao Korrai, Chairman of the Communications and Signal Processing Societies Joint Chapter, IEEE Hyderabad Section

The Communications and Signal Processing Societies Joint Chapter of the IEEE Hyderabad Section was inaugurated on April 17, 2010. Since then, the Chapter has organized several Distinguished Lectures, technical lectures, and tutorials. The Hyderabad Chapter is initiating a series of lectures starting in 2011 to mark the anniversary of the birth of Alexander Graham Bell, who took up Meucci's invention of the telephone and transformed it into an industry product with telephone service and its infrastructure. The Chapter wishes these lectures to be delivered by distinguished scientists on the birthday of Alexander Graham Bell, 3 March, every year, in the broad areas of communications and signal processing.

Dr.Tapan K Sarkar, IEEE Fellow and professor in the Department of Electrical and Communications Engineering, Syracuse University, New York, delivered the 2011 memorial lecture, "A Look at Some of the Principles of Mobile Communication from an Electrical Engineering Perspective," on 3 March 2011 at Seminar Hall CR Rao Advanced Institute of Mathematics, Statistics, and Computer Science (CR Rao AIMSCS), University of Hyderabad, Hyderabad Campus.

The IEEE Signal Processing Society funded Dr. Sarkar's international air travel and accommodation for two nights in Hyderabad. His domestic air travel, travel within Hyderabad, and accommodation for one night in Hyderabad were arranged by CR Rao AIMSCS.

The audience for this lecture was 46 students, research scholars, participants from industry, and faculty from universities and colleges.

Antonio Meucci, the first electromagnetic telephone inventor.

Dr. V. U. Reddy (first from right), IEEE Life Fellow, memorial lecture subcommittee chair Dr. Tapan K Sarkar (second from right), Dr. Deergha Rao Korrai (third from right), Chapter Chair, and IEEE volunteers of the Hyderabad section after the lecture at Seminar Hall, CR Rao AIMSCS.

During the lecture Dr. Sarkar stressed that one of the basic tenets of contemporary mobile communication is directing electromagnetic energy from a transmitter to a receiver by employing an array through space-division multiple access. As the antenna beam pattern can only be characterized in the far field, it is useful to recollect where the far field of an antenna starts. For example, if a half-wave dipole is operating at 1 GHz in free space, its far field starts 0.15 m from the dipole. However, if that same dipole is placed on top of a tower 20 m above the Earth, its far field starts approximately 10 km away from the antenna, which is quite a large distance. Under a Maxwellian approach, it is difficult to visualize what antenna beamforming means in a near field environment and increasing channel capacity under this environment through antenna diversity. Since in a near field environment power is complex, it is difficult to visualize what a complex value for power in the expression of Shannon channel capacity will signify. However, through application of the reciprocity principle, it is possible to direct the signal from a transmit array to a preselected receiver and produce minimal signals simultaneously at other undesired receivers, even in a near field environment. It has been shown that the propagation measurements made by Okamura, and the development of the model by Hata and the plethora of refinements of this model can all be explained quite simply using a commercial electromagnetic simulation tool for analyzing antennas over an imperfect ground plane using the Sommerfeld formulation. Such an analysis can duplicate the slope of the propagation measurements with remarkable accuracy, and can thus perform propagation modeling.

Finally, he discussed broadband systems where dispersionless communication can take place over a 40 GHz band width. Experimental results were provided to illustrate how that can be accomplished. However, to arrive at this goal, we need to reorient our way of thinking. The conventional wisdom that the radiation pattern of an antenna when it is operating as a transmitter is equal to the same pattern when it is operating as a receiving antenna does not hold in the time domain. The transmit impulse response of an antenna is the time derivative of the receive impulse response of the same antenna. This is relevant for wideband baseband communication. For a certain choice of transmitting and receiving antenna systems, the entire propagation wireless channel can be made completely dispersionless over a 40 GHz bandwidth.

At the end, all the participants agreed that the lecture was excellent.

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

The Federation of Telecommunication Engineers of the European Union: The European ComSoc Sister Society

By Andrea Penza, FITCE President, Italy

The Federation of Telecommunication Engineers of the European Union (FITCE) is a forum for the information and communications technology (ICT) and media professionals community to exchange views and gain insights on the new developments and challenges related to technical, regulatory, societal, and economic aspects of ICT and media technologies and services. Since 2009 it has been one of the European ComSoc Sister Societies.

2011 is really a special year: FITCE is going to celebrate its golden anniversary, 50 years of life, a meaningful life and rich with great successes.

Since the beginning, FITCE was a technical reference point within the European Union, trying to convey the interests of all the technological worlds, universities, the business world, governments, and institutions. In the past 50 years FITCE committed itself to developing the telco culture, and stimulated the relationship between members and the stakeholders in the market. Every year FITCE has organized an important congress in a different European location, and most of the reference people of the European telecommunications bodies attended and offered great contributions to the development of the culture of the sector. For many years the yearly FITCE Congress represented the meeting point for exchanging information, ideas, and opinions in telecommunications. For many years the yearly Congress was named "The European Days of Telecommunications," outlining the great meaning and importance this event represented in the European telco community.

Since its beginning FITCE has been a valuable theatre and reference point of prestigious meetings between the most important stakeholders in the telco market, sharing opinions, points of view, and common values. After 50 years, FITCE has to be fully proud of the high-level contributions offered by all its members to the development of the telecommunication culture either in specific European countries (via National Associations) or in European organizations.

Now, after 50 years, as FITCE President I am pleased to announce that we have decided to celebrate this special birthday during our yearly congress. It will be held in Palermo, Italy, from 31 August to 3 September, with the strong support of a joint partnership with IEEE ComSoc, the most important and prestigious telecommunications society in the world. That constitutes for FITCE great motivation of pride and success. We decided to celebrate this important event with a special program where specific awards will be assigned to all the people who, since the beginning of FITCE'S history, have offered strategic and prestigious contributions to its development and growth. All those who have distinguished themselves within FITCE life have also experienced great success in their professional life, offering meaningful contributions to the evolution of the telecommunication world.

The title of the Congress, "ICT: Bridging the Ever Shifting Digital Divide," will create the opportunity to discuss the most up-to-date topics in the telecommunication market of today, always with an eye open to future trends. Mobile and fixed telco trends will receive top priority in the development of the strategic topics of the Congress.

The Congress will explore many issues: technological, economic, financial, and strategic; all these aspects of the telecommunication world will resonate deeply, and special attention will be given to the new emerging topics of sustainability and green ICT in order to discuss and consolidate ideas and contributions to the social well being of all of us.

My best wishes for FITCE's future and our partnership with IEEE ComSoc.

Global Communications Newsletter

www.comsoc.org/pubs/gcn

STEFANO BREGNI
Editor

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

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Guest Editorial

Energy Efficiency in Communications: Part II

Honggang Zhang *Andreas Gladisch* *Mario Pickavet* *Zhifeng Tao* *Werner Mohr*

This Feature Topic issue is a continuation in sequence of the previous November 2010 issue on "Energy Efficiency in Communications." Dedicated to the latest advances in various energy-efficient fixed and wireless communications and networking technologies, this second part (June 2011) of the Feature Topic further highlights the increasingly global interest in the technical potentials and approaches to mitigating energy consumption and the consequential environmental influences, and includes eight articles selected from a pool of high-quality paper submissions.

The first article, "Fundamental Trade-offs on Green Wireless Networks" by Yan Chen *et al.*, presents an insightful design framework for energy-efficiency-oriented mobile wireless networks, which consists of four fundamental trade-offs: deployment efficiency vs. energy efficiency, spectrum efficiency vs. energy efficiency, bandwidth vs. power, and delay vs. power. Within this article, the authors thoroughly analyze how to balance the deployment cost, throughput, and energy consumption in the network as a whole, how to guarantee the achievable rate while maintaining energy consumption of the system on a given available bandwidth, how to utilize the bandwidth and the power needed for transmission at a given target rate, and how to counterpoise the average end-to-end service delay and average power consumed in transmission, respectively.

Hanna Bogucka and Andrea Conti contribute the second article, "Degrees of Freedom for Energy Savings in Practical Adaptive Wireless Systems," and verify that adaptive communication techniques have degrees of freedom to potentially be exploited for energy saving; meanwhile, the target performance metrics can be satisfied as well, which depend on various system parameters such as the diversity technique, the energy partition between data and pilot symbols for channel estimation, and the constellation signaling. As a case study, the authors also investigate single-carrier as well as multicarrier communication systems applying both margin-adaptive and rate-adaptive pilot-assisted transmission to quantify the relevant energy savings opportunities.

Following the above two articles, the third one, "Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks" by Congzheng Han *et al.*, provides an in-depth overview of the ongoing Mobile VCE Green Radio project, which aims to establish novel approaches to reducing the energy consumption of wireless links, especially improving the design and operation of wireless base stations. Through the project, it has been shown that base stations can have much higher operational energy budgets than mobile terminals; therefore, appropriate modeling of the energy consumption of base stations is an important issue for decreasing the energy consumption of whole mobile communications systems.

Not coming singly but paired with the above, the fourth article, "Toward Dynamic Energy-Efficient Operation of Cellular Network Infrastructure" by Eunsung Oh *et al.*, depicts how dynamic operation of cellular base stations, in which redundant low-traffic base stations are switched off, can generate significant energy savings advantages. Based on real cellular traffic traces and information regarding base station locations, the authors discuss the first-order approximation of the percentage of power saving that can be expected by turning off base stations during low traffic periods while maintaining coverage and interoperator coordination.

As we can see, the four articles selected above are mainly focused on typical scenarios of mobile wireless communications and access networks. Next, we give space to the next four articles, concentrating on integrated (wired/wireless) or fixed core telecommunication networks. Accordingly, we begin with the fifth article, "Power Consumption in Telecommunication Networks: Overview and Reduction Strategies" contributed by Willem Vereecken *et al.*, which gives an overview of typical power consumption figures in a variety of wired and wireless networks, customer premises and core networks. Some key potential directions for power consumption are highlighted and explained.

The sixth article, "Energy Consumption in Wired and Wireless Access Networks" by Jayant Baliga *et al.*, provides detailed analyses on the corresponding energy consumptions of digital subscriber line, hybrid fiber coax networks, PONs, fiber to the node, point-to-point optical systems, UMTS (WCDMA), and WiMAX. The authors conclude that PONs and point-to-point optical networks are the most energy-efficient access solutions at high access rates.

The seventh article, "Putting the Cart before the Horse: Merging Traffic for Energy Conservation" by Suresh Singh and Candy Yiu, illustrates a novel solution for linear scaling of energy usage with the traffic loads within the Internet, which involves aggregating traffic from multiple input links prior to feeding them to the switch interfaces, so as to maximize the number of interfaces put to sleep. The authors arrive at a promising result: energy consumption, measured as fraction of awaking interfaces, scales linearly with load for all loads and the proposed algorithms are actually deterministic without any packet loss.

The last article, "On the Design of Green Reconfigurable Router towards Energy Efficient Internet" by Chunming Wu *et al.*, discusses how to construct energy-efficient reconfigurable router with power-aware routing mechanism through virtual networks with advanced rate adaptation processing

GUEST EDITORIAL

inside the Internet router. In particular, by taking into account of the Internet behavior features and the modular architecture of routers, the GRecRouter (Green Reconfigurable Router) designed in this article takes advantage of various opportunities and means to greatly cut down the power dissipation at the network, node, and function levels.

Finally, we would like to thank Dr. Steve Gorshe for his kind encouragement and valuable suggestions for achieving this Feature Topic issue. We also would like to thank the publications staff of IEEE Communications Society, Jennifer Porcello, Cathy Kemelmacher, and Joe Milizzo, for their professional efforts and wonderful support. Definitely, we hope that the selected articles in this issue and the final issue (August 2011) will further encourage the readers of *IEEE Communications Magazine* to be actively involved in this significant emerging area of green communications.

BIOGRAPHIES

HONGGANG ZHANG (honggangzhang@zju.edu.cn) is a full professor in the Department of Information Science and Electronic Engineering, as well as co-director of the York-Zhejiang Lab for Cognitive Radio and Green Communications, Zhejiang University, China. He received his Ph.D. degree in electrical engineering from Kagoshima University, Japan, in 1999. Prior to that, he received his Bachelor's and Master's of Engineering degrees, both in electrical engineering, from Huazhong University of Science & Technology (HUST), China, in 1989, and Lanzhou University of Technology, China, in 1992, respectively. From October 1999 to March 2002 he was with the Telecommunications Advancement Organization (TAO) of Japan, as a TAO research fellow. From April 2002 to November 2002 he was with the Toyota IT Center. From December 2002 to August 2004 he was with the UWB Research Consortium, Communications Research Laboratory (CRL), and National Institute of Information and Communications Technology (NICT) of Japan, focusing on UWB wireless communications, IEEE 802.15 WPAN standardizations, and Wireless 1394 smart home networks. He was the principal author and contributor for proposing DS-UWB in the IEEE 802.15 WPAN standardization task group. From September 2004 to February 2008 he was with CREATE-NET, where he led its wireless group in participating in a number of European FP6 & FP7 projects (EUWB, PULSERS 2). He was Co-Chair of the IEEE GLOBECOM 2008 Symposium on Selected Areas in Communications. He was the founding Technical Program Committee (TPC) Co-Chair of CrownCom 2006 as well as a Steering Committee Member of CrownCom 2006–2009. He has served as Guest Editor for an ACM/Springer *Mobile Networks and Applications Journal* Special Issue on Cognitive Radio Oriented Wireless Networks and Communications in 2007, an Elsevier *PHYCOM* Special Issue on Cognitive Radio: Algorithms and System Design in 2008, an *IEEE Transactions on Vehicular Technology* Special Issue on Achievements and the Road Ahead: The First Decade of Cognitive Radio in 2009, and an *IET Communications Journal* Special Issue on Cognitive Communications in 2010. He is an Honorary Visiting Professor of the University of York, United Kingdom. He serves as Chair of the Technical Committee on Cognitive Networks (TCCN) of the IEEE Communications Society.

ANDREAS GLADISCH [M] (andreas.gladisch@telekom.de) received his Dipl. Ing. degree in theory of electro techniques from the Technical University of Ilmenau, Germany, in 1986 and his Ph.D. degree in optical communications from Humboldt University of Berlin, Germany, in 1990, where he was engaged in research on coherent optical communication and optical frequency control. He joined the Research Institute of Deutsche Telekom in 1991, where he was involved in projects on coherent optics, WDM systems, wavelength control, and frequency stabilization. From 1996 to 1998 he led a research group working in the field of design and management of optical networks, and in 1999 he became director of the Department on Network Architecture and System Concepts of T-Systems. Since 2009 he is responsible for the research field Broadband Network Architecture and Economics of Deutsche Telekom Laboratories. He has participated in several European funded research projects, such as DEMON, MOON, LION, NOBEL, OASE, and SPARC. His research interests are in field network architecture and how optical technologies could change the overall network architecture. In 2002 he was Guest Editor of *IEEE JSAC* on WDM-Based Network Architectures. He was responsible for projects developing the mid-term strategy of the Deutsche Telekom transport network, especially concepts for the further development of SDH, WDM, and OTN. His current research activities are in broadband access network architecture, techno-economics, and green communication. He has been a TPC member for numerous conferences on optical communication and network architecture. During the last years, he has organized workshops and sessions about green communication and energy-efficient ICT at various conferences (e.g., ICC, ECOC). He has authored or coauthored more than 100 national and international technical conference or journal papers. He is a member of Informationstechnische Gesellschaft (ITG).

MARIO PICKAVET (mario.pickavet@intec.ugent.be) is full professor at the Department of Information Technology (INTEC) of Ghent University, Belgium. He received an M.Sc. degree in electrical engineering, specialized in telecommunications, in 1996 from Ghent University. His graduation thesis, *Topological Planning of Telecommunication Networks using Genetic Algorithms*, covered the development of a genetic algorithm enhanced with some deterministic optimization routines for a joint topology-capacity design of telecommunication transport networks. From 1996 until 1999, he worked as a research fellow of the Flemish Fund for Scientific Research (FWO-V) in the Broadband Communications Networks Group. In 1999 he received a Ph.D. degree in electrical engineering from the same university. His Ph.D. thesis, *Use of Heuristic Techniques for Global Design and Planning of Telecommunication Networks*, covered the design of mesh-based transport networks and the multi-period planning of survivable networks. Since 2000 he has been a professor at Ghent University where he teaches courses on discrete mathematics, multimedia networks, and network modeling. His current research interests are related to broadband communication networks (WDM, IP, [G-]MPLS, Ethernet, OPS, and OBS), and include design, long-term planning, techno-economical analysis, and energy efficiency of core and access networks. Special attention goes to operations research techniques that can be applied for routing and network design. In this context, he is currently involved in several European and national projects, such as the Network of Excellence "Building the Future Optical Network in Europe" (BONE), DICONET, ECODE, ALPHA, SPARC, TREND, and OASE. He has published about 200 international publications, both in journals (*IEEE Communications Magazine*, *IEEE JSAC*, *OSA Journal of Lightwave Technology*, *European Transactions on Telecommunications*, *Photonic Network Communication*, *Journal of Heuristics*, etc.) and in proceedings of various conferences. He has been a TPC member for numerous conferences. During the last years, he has organized several workshops and sessions about energy efficiency at various communication network conferences (e.g., ECOC 2008, OFC 2009). He is co-author of the book *Network Recovery: Protection and Restoration of Optical, SONET-SDH, IP, and MPLS*. He is a holder of a bronze medal at the International Mathematical Olympiad (Sweden, 1991).

ZHIFENG TAO (jeffrey.tao@gmail.com) obtained his B.S. from Xian Jiaotong University, P.R. China. He received M.S. and Ph.D. degrees in electrical engineering from Polytechnic Institute of New York University, Brooklyn, New York. He joined Mitsubishi Electric Research Laboratories in September 2006 and currently holds the position of principal member of technical staff. His research interests include electric vehicle networks, smart grid, cooperative communications, resource allocation, interference management, and wireless routing. He has published over 40 papers, filed over 60 U.S. patents, written several book chapters, and made over 70 technical contributions to IEEE standards. He has played a key role in the development of such standards as IEEE 802.16j and IEEE P2030 (smart grid interoperability guidelines), and received the IEEE Standard Association's recognition for Outstanding Contribution to the Development of IEEE Std 802.16j-2009 (Multihop Relay). He is a Guest Editor of an *IEEE Communications Magazine* Special Issue on Energy Efficiency in Communications, and a *EURASIP Journal on Advances in Signal Processing* Special Issue on Cooperative MIMO Multicell Networks. He was the founding Chair of the 1st IEEE International Workshop on Smart Grid Communications (affiliated with IEEE ICC 2010) and a Symposium Chair of the 1st IEEE International Conference on Smart Grid Communications. He has served as TPC Chair of the Frontier on Communications and Networking Symposium of ChinaCom 2008, Tutorial Chair of ChinaCom 2009, and Publication Chair of the International Conference on Ambient Media and Systems (Ambi-sys) 2008. He has also been a TPC member for numerous conferences, including IEEE GLOBECOM, ICC, VTC, and PIMRC. He has delivered tutorials on next-generation wireless systems at IEEE GLOBECOM and ICC, and served as a panelist on Green Communications Panel in ICC 2009. He is also a co-recipient of the IEEE GLOBECOM 2009 Best Paper Award.

WERNER MOHR [SM] (werner.mohr@nsn.com) graduated from the University of Hannover, Germany, with a Master's degree in electrical engineering in 1981 and a Ph.D. degree in 1987. He joined Siemens AG, Mobile Network Division in Munich, Germany, in 1991. He was involved in several EU funded projects and ETSI standardization groups on UMTS and systems beyond 3G. Since December 1996 he was project manager of the European ACTS FRAMES Project until the project finished in August 1999. This project developed the basic concepts of the UMTS radio interface. Since April 2007 he has been with Nokia Siemens Networks GmbH & Co. KG in Munich Germany, where he is head of Research Alliances. He was the coordinator of the WINNER Project in Framework Program 6 of the European Commission, Chairman of WWI (Wireless World Initiative) and the Eureka Celtic project WINNER+. The WINNER project laid the foundation for the radio interface for IMT-Advanced and provided the starting point for the 3GPP LTE standardization. In addition, he was Vice Chair of the eMobility European Technology Platform in the period 2008-2009 and he is now eMobility (now called Net!Works) Chairperson for the period 2010–2011. He was Chair of the "Wireless World Research Forum - WWRF" from its launch in August 2001 up to December 2003. He is a member of VDE (Association for Electrical, Electronic & Information Technologies, Germany). In 1990 he received the Award of the ITG (Information Technology Society) in VDE. He was a board member of ITG in VDE, Germany from 2006 to 2008 and was re-elected for 2009 to 2011. He is co-author of the books *Third Generation Mobile Communication Systems* and *Radio Technologies and Concepts for IMT-Advanced*.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Fundamental Trade-offs on Green Wireless Networks

Yan Chen, Shunqing Zhang, and Shugong Xu, GREAT Research Project, Huawei Technologies Co., Ltd.
Geoffrey Ye Li, Georgia Institute of Technology

ABSTRACT

Traditional mobile wireless network mainly design focuses on ubiquitous access and large capacity. However, as energy saving and environmental protection become global demands and inevitable trends, wireless researchers and engineers need to shift their focus to energy-efficiency-oriented design, that is, green radio. In this article, we propose a framework for green radio research and integrate the fundamental issues that are currently scattered. The skeleton of the framework consists of four fundamental trade-offs: deployment efficiency–energy efficiency, spectrum efficiency–energy efficiency, bandwidth–power, and delay–power. With the help of the four fundamental trade-offs, we demonstrate that key network performance/cost indicators are all strung together.

INTRODUCTION

WHY GREEN EVOLUTION?

The next-generation wireless networks are expected to provide high-speed Internet access anywhere and anytime. The popularity of the iPhone and other types of smartphones is doubtlessly accelerating the process and creating new traffic demands, such as mobile video and gaming. The exponentially growing data traffic and the requirement for ubiquitous access have triggered dramatic expansion of network infrastructures and fast escalation of energy demands. Hence, it becomes urgent for mobile operators to maintain sustainable capacity growth and, at the same time, limit the electric bill.

The escalation of energy consumption in wireless networks directly results in increased greenhouse gas emission, which has been recognized as a major threat to environmental protection and sustainable development. The European Union has acted as a leader in energy saving across the world and targeted a 20 percent greenhouse gas reduction. China's government has also promised to reduce the energy per unit of gross domestic product (GDP) by 20 percent and major pollution by 10 percent by 2020. The pressure of social responsibility serves as another strong driving force for wireless operators to dramatically reduce energy consumption and carbon footprint. Worldwide actions have been taken. For instance, Vodafone Group has announced to reduce its CO_2 emissions by 50 percentfromt its 2006–2007 baseline of 1.23 million tonnes by 2020.[1]

To meet the challenges raised by the high demands of wireless traffic and energy consumption, green evolution has become an urgent need for wireless networks today. As pointed out in [1], the radio access part of the cellular network is a major energy killer, which accounts for up to more than 70 percent of the total energy bill for a number of mobile operators.[2] Therefore, increasing the energy efficiency of radio networks as a whole can be an effective approach. Vodafone, for example, has foreseen energy efficiency improvement as one of the most important areas that demand innovation for wireless standards beyond Long Term Evolution (LTE) [2].

Green radio (GR), a research direction for the evolution of future wireless architectures and techniques toward high energy efficiency, has become an important trend in both the academic and industrial worlds. Before GR there were efforts devoted to energy saving in wireless networks, such as designing ultra-efficient power amplifiers, reducing feeder losses, and introducing passive cooling. However, these efforts were isolated and thus could not form a global vision of what we can achieve in five or ten years on energy saving. GR, on the other hand, targets innovative solutions based on top-down architecture and joint design across all system levels and protocol stacks, which cannot be achieved via isolated efforts.

RESEARCH ACTIVITIES

In academia, several workshops dedicated to green communications have been organized to discuss future green technologies. For instance, IEEE had two green communication workshops in 2009, in conjunction with ICC '09 and GLOBECOM '09, and at least three more in 2010, in conjunction with ICC '10, PIMRC '10, and GLOBECOM '10, respectively.[3]

On the other hand, research projects on GR have sprung up under different international research platforms in recent years. Table 1 lists some major international projects on GR

---

[1] *Information available at http://www.vodafone.com/start/media relations/news/group press releases/2007/01.html*

[2] *The figure is from the energy efficiency solution white paper of Huawei Technologies, "Improving Energy Efficiency, Lower CO_2 Emission and TCO"; http://www.huawei.com/green.do.*

[3] *ICC, GLOBECOM, and PIMRC are three international conferences of IEEE Communications Society.*

Project Name	Duration	Platform	Partners				Green Vision of 2020
OPERA-Net	2008.10 ~2010.10	CLETIC	France Telecom (FT)	Alcatel-Lucent (ALU)/ Nokia-Siemens Networks (NSN)/ Thomson Grass Valley	Cardiff/IMEC/ VTT	Free-scale	Improve energy efficiency of 20% by 2020
Green Radio	2009.1 ~2012.1	MVCE CORE 5	Vodafone/British Telecom/FT (Orange)	NSN/Huawei/ ALU/Thales	Bristol/KCL/ Swansea/ Edinburgh/ Southamon	NEC/Fujitsu/ Toshiba	Secure 100x reduction in energy requirements for delivery of high date rate services
ERATH	2010.1 ~2012.6	FP7 Call 4 IP	Telecom Italia/ NTT-Docomo	Ericsson/ALU	Surrey/TU Dresden/Oulu/ BME/IST/TULK/	NXP/ETSI/TTI/ ETH	Cut the energy use of mobile cellular networks by a factor of at least 2
Green Touch	2010.1 ~2015.1	—	Telefornica/China Mobile/Swisscom/ Portugal/Telecom/ FCM	ALU Bell Lab/Huawei/ Samsung	WSL (Stanford)/ RLE (MIT)/INRIA/ IBES (Melbourne)/ IMEC	Free-scale	Reduce energy per bit by a factor of 1000 from current levels

Table 1. *International research projects related to green radio. The information was authentic as of September 2010.*

research.[4] For instance, Optimizing Power Efficiency in Mobile Radio Networks (OPERANET), a European research project started in 2008, deals with the energy efficiency in cellular networks. In the United Kingdom, GR has been among Core 5 Programs in Mobile VCE since 2009, targeting parallel evolution of green architectures and techniques. Moreover, Energy Aware Radio and Network Technologies (EARTH) [3], one of the integrated projects under European Framework Programme 7 Call 4, started its ball rolling on develop green technologies at the beginning of 2010. Most recently, GreenTouch, a consortium of industry, academic, and non-governmental research experts set its five-year research goal to deliver the architecture, specification, and roadmap needed to reduce energy consumption per bit by a factor of 1000 from the current level by 2015.

The Target of This Article

GR research is a large and comprehensive area that covers all layers in the protocol stack of wireless access networks as well as the architectures and techniques. Instead of a survey that reaches every aspect of the matter, this article focuses on the fundamental framework for GR research and strings together the currently scattered research points using a logical "rope." We propose in this article four fundamental trade-offs to construct such a framework. As depicted in Fig. 1, they are:

- Deployment efficiency (DE)–energy efficiency (EE) trade-off: to balance the deployment cost, throughput, and energy consumption in the network as a whole
- Spectrum efficiency (SE)–EE trade-off: given bandwidth available, to balance the achievable rate and energy consumption of the system
- Bandwidth (BW)–power (PW) trade-off: given a target transmission rate, to balance the bandwidth utilized and the power needed for transmission
- Delay (DL)–PW trade-off: to balance the average end-to-end service delay and average power consumed in transmission

By means of the four trade-offs, key network performance/cost indicators are all strung together.

Fundamental Framework

In this section, we shall elaborate in detail on the four trade-offs that constitute the fundamental framework. As we can see, they actually connect the technologies toward green evolution in different research aspects, such as network planning, resource management, and physical layer transmission scheme design.

DE–EE Trade-off

DE, a measure of system throughput per unit of deployment cost, is an important network performance indicator for mobile operators. The deployment cost consists of both capital expenditure (CapEx) and operational expenditure (OpEx). For radio access networks, CapEx mainly includes infrastructure costs, such as base station equipment, backhaul transmission equipment, site installation, and radio network controller equipment. The key drivers for OpEx, on the other hand, are electric bill, site and backhaul lease, and operation and maintenance cost [4]. Usually, wireless engineers will estimate the network CapEx and OpEx during network planning. EE, defined as system throughput for unit of energy consumption, is mostly considered during network operation.

The two different metrics often lead to opposite design criteria for network planning. For example, in order to save the expenditure on site rental, base station equipment, and maintenance, network planning engineers tend to "stretch" the cell coverage as much as possible. However, the path loss between the base station and mobile users will degrade by 12 dB whenever the cell radius doubles if the path loss expo-

[4] *Detailed information about these projects can be found at the following addresses: http://www.mobilevce.com/index.htm (MVCE), http://www.greentouch.org (Green Touch), and http://www.opera-net.org/ (OPERA-NET), respectively.*

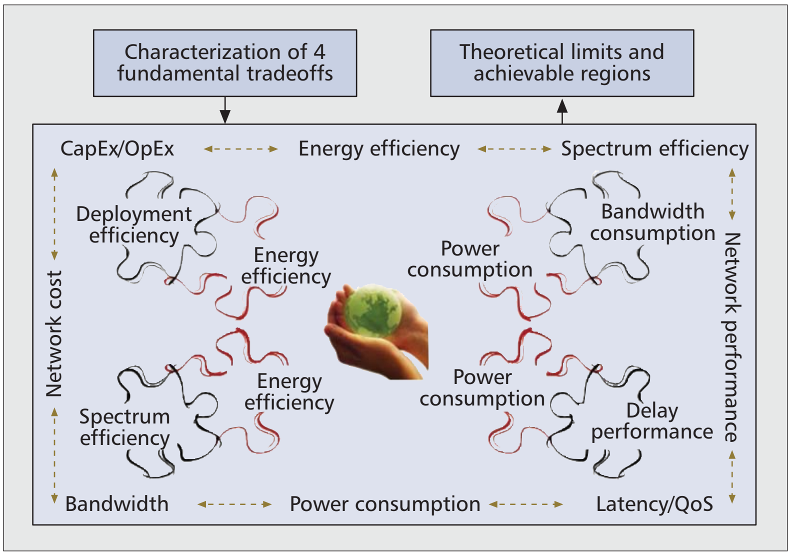


Figure 1. *Fundamental trade-offs.*

nent is four, which induces a 12 dB increase in transmit power to guarantee the same received signal strength for users at the cell edges. On the other hand, to provide cellular coverage for a given area, increasing the number of base stations will save the total network transmit power by the same factor. For example, it is shown in [5] that by shrinking the cell radius from 1000 m to 250 m, the maximum EE of a high-speed data packet access (HSDPA) network will be increased from 0.11 Mb/J to 1.92 Mb/J, corresponding to a 17.5 times gain. Therefore, to minimize energy radiation, radio resource management engineers favor small cell size deployment. From the above discussion, there should be a trade-off between DE and EE, as shown in Fig. 2a, where each point on the curve corresponds to a cell size, and should be chosen to balance specific DE and EE requirements.

However, this shape of the curve is correct when only transmission power is considered, and the deployment cost scales continuously and proportionally with the cell radius. In reality:

- There are limited types of base stations, and the equipment cost does not scale proportionally with the target cell size.
- The total network energy includes both transmit-dependent energy (e.g., power consumed by a radio amplifier) and transmit-independent energy (e.g., site cooling power consumption).

Therefore, the relation of DE and EE may deviate from the simple trade-off curve and become more complex when considering practical aspects, as shown in our recent study [6]. Figure 3 summarizes the main result of [6]. From the rightmost plot, there might not always be a trade-off between DE and EE, and the shape of a DE-EE curve depends on specific deployment scenarios. For the suburban scenario, where the path loss exponent is small (about 3.5), the network EE even increases with its DE. For the dense urban scenario, where the path loss exponent is large (about 4.5), two different EE values may result in the same DE value, corresponding to very small and very large cell radii, respectively. The former is because of the huge increase in CapEx by increasing the number of sites; the latter is due to the sharply increased electric bill in OpEx.

Since the shapes of DE-EE curves may not match our intuition, characterizing the curves with practical concerns is helpful to real-world network planning. As shown in Fig. 3, for any target network throughput and given deployment budget, we can first calculate the corresponding deployment efficiency, from which we can decide the maximum achievable energy efficiency by looking up the DE value on the DE-EE trade-off curve; then from the EE vs. cell radius curve, we get the corresponding optimal cell size.

No doubt the current results are still quite preliminary. In the future, research efforts may focus on the following two aspects:

- Improving the optimal DE-EE frontiers with advanced network architectures
- Joint architecture design with advanced transmission schemes and scheduling algorithms to improve the network DE-EE trade-off relation

For LTE-Advanced or beyond networks, heterogeneous networks (HetNet) has been approved as a work item, such as in Third Generation Partnership Project (3GPP) Release 10. With the combination of macrocells and micro/pico/femtocells, the traditionally related functionalities, coverage, and capacity provision can now be decoupled into different tiers of the network. In general, macrocells handle coverage and mobility issues, while micro/pico cells focus on local throughput. It has been shown in [7] that the network EE increases as the density of micro/picocells grows. On the other hand, the DE aspect of HetNet has been studied in [4] for different traffic distributions. From [4], a complementary hotspot layer of micro/pico cells on top of macrocells has been the most cost-effective architecture for non-uniform spatial traffic. The trade-off of DE and EE for HetNet, however, is still open.

Another promising candidate for future architectures is cooperative networks (CoopNet), where new air interface techniques, such as relay and distributed antenna systems (DAS), are employed. The newly introduced infrastructures, such as relays and remote radio heads, are of much lower cost and smaller coverage compared to macro base stations, which bring mobile users closer to the network and make deployment more flexible. However, the backhaul cost and signaling overhead may become new killers for energy consumption and system efficiency. Therefore, how much improvement the CoopNet architecture can bring to the DE-EE trade-off needs to be carefully studied.

Moreover, the incorporation of EE oriented user scheduling and radio resource management algorithms on top of HetNet and CoopNet are bound to further improve network utilization efficiency. This is especially important when the spatial traffic distribution is non-uniform and varies with time. Dynamic power control that exploits channel variations has been proven to enhance the link-level power efficiency. Similarly, by extending the idea to the network level, we may introduce dynamic coverage management to

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

exploit traffic variations. Dynamic switch off/on of coverage overlaid cells in low traffic is an example in HetNet, while dynamic relay selection or CoMP pattern selection is the counterpart in CoopNet. As it introduces no extra cost but saves redundant energy consumption, it can improve DE and EE simultaneously.

More research efforts on this topic are desired in the future.

SE-EE TRADE-OFF

SE, defined as the system throughput per unit of bandwidth, is a widely accepted criterion for wireless network optimization. The peak value of SE is always among the key performance indicators of 3GPP evolution. For instance, the target downlink SE of 3GPP increases from 0.05 b/s/Hz to 5 b/s/Hz as the system evolves from GSM to LTE. On the contrary, EE was previously ignored by most research efforts and was not considered by 3GPP as an important performance indicator until very recently. As the green evolution becomes a major trend, energy-efficient transmission becomes more and more important. Unfortunately, SE and EE are not always consistent and sometimes conflict with each other. Therefore, how to balance the two metrics in future systems deserves careful study.

To characterize the SE-EE trade-off for point-to-point transmission in additive white Gaussian noise (AWGN) channels, Shannon's capacity formula plays a key role. From Shannon's formula, the achievable transmission rate, R, under a given transmit power, P, and system bandwidth, W, is simply

$$R = W\log_2\left(1+\frac{P}{WN_0}\right),$$

where N_0 stands for the power spectral density of AWGN. According to their definitions, SE and EE can be expressed as

$$\eta_{SE} = \log_2\left(1+\frac{P}{WN_0}\right)$$

and

$$\eta_{EE} = W\log_2\left(1+\frac{P}{WN_0}\right)/(P),$$

respectively. As a result, the SE-EE relation can be expressed as

$$\eta_{EE} = \frac{\eta_{SE}}{(2^{\eta_{SE}}-1)N_0}, \quad (1)$$

which is sketched in Fig. 2a. From the above expression, η_{EE} converges to a constant, $1/(N_0 \ln 2)$ when η_{SE} approaches zero. On the contrary,

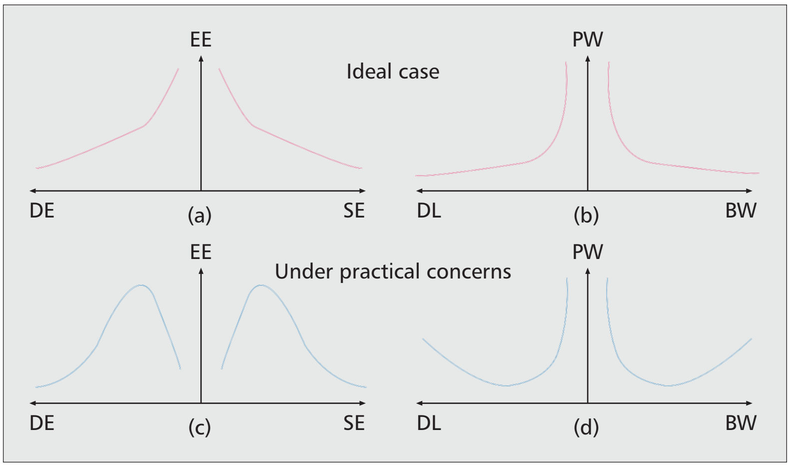


Figure 2. *Sketch of the four trade-off relations without and with practical concerns.*

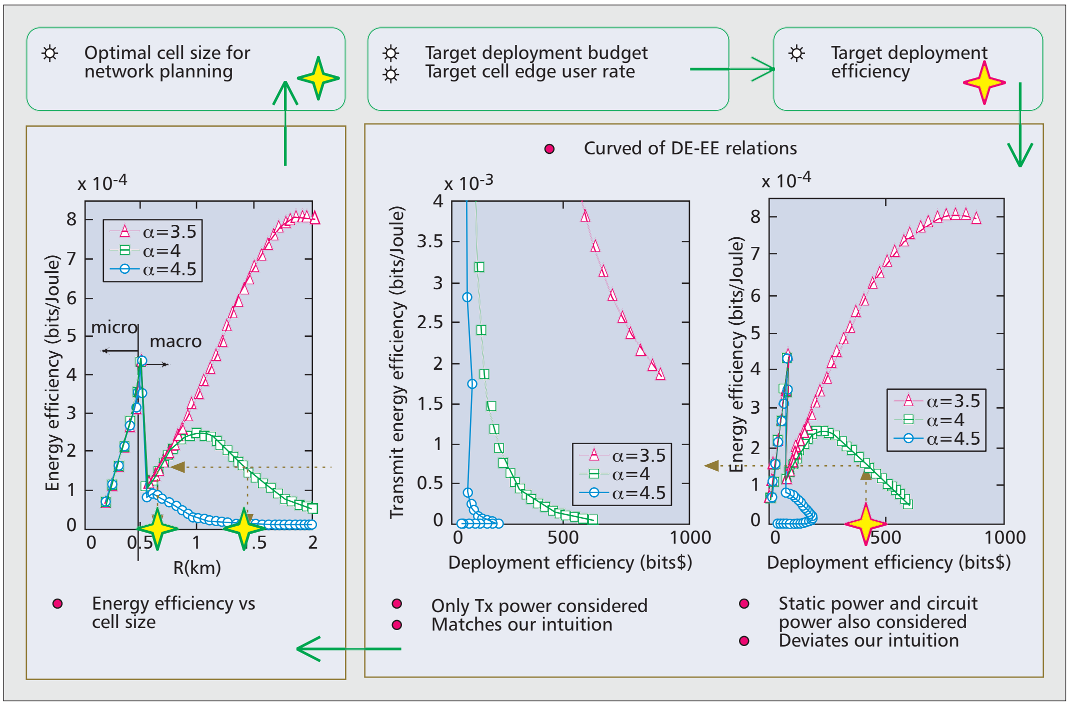


Figure 3. *Results on DE-EE relation from [6] for different path-loss exponents, α.*

In advanced network architectures, such as HetNet and CoopNet, the system may benefit even more from the joint design of physical transmission and resource management. Our recent work in presents initial results in relay-assisted cooperative systems.

η_{EE} approaches zero when η_{SE} tends to infinity.

In practical systems, however, the SE-EE relation is not as simple as the above formula. In particular, circuit power will break the monotonic relation between SE and EE as shown in [8–10]. More precisely, if circuit power is considered, the SE-EE curve will turn to a bell shape, as illustrated in Fig. 2c. From [8], we see that the transmission conditions and strategies, such as the transmission distance, modulation, and coding scheme, and resource management algorithms all have significant impact on the tradeoff of SE and EE.

Nevertheless, the SE-EE relation characterized by Eq. 1 is only for point-to-point transmission rather than for a network. Further investigation of energy-efficient transmission policies is expected to obtain more benefit and is crucial for environmental protection and sustainable development in future wireless cellular systems. Examples of future research topics may include the following aspects:

- Characterizing the SE-EE trade-off under practical hardware constraints
- Investigating the network SE-EE trade-off in multi-user/multicell environments
- Joint design of physical layer transmission schemes and resource management strategies that will improve the network SE-EE trade-off

The performance limit predicted by theoretical analysis may not be achieved in real systems due to practical hardware constraints. For instance, the typical energy conversion efficiency[5] of a power amplifier in current base stations is less than 40 percent. Moreover, the limited linearity regions of power amplifiers also set a constraint on the transmitted signals, such as the peak-to-average power ratio. How these issues would affect the SE-EE trade-off is not clear yet. Therefore, a more detailed modeling of the equipment level energy consumption and practical constraints in hardware devices and transmission signals will help us to find practically achievable SE-EE regions. The gaps between the theoretical limits and achievable regions may further guide the design of future wireless networks.

For the multi-user/multicell cases, interuser interference or intercell interference may break the fundamental assumptions in the point-to-point cases. An interesting extension of the SE-EE trade-off to multicell scenarios with intercell interference was studied in [9]. From [9], the interference power generated by the neighboring cells not only reduces the maximum achievable EE but also degrades SE and EE. As can be imagined, the higher the interference level, the larger the degradation would be. In this case, the results from the simple point-to-point case are not applicable, and a systematic approach to multi-user/multicell systems shall be developed to build on the theoretical fundamentals of energy-efficient wireless transmissions.

Energy-efficient transmission, from the point of view of resource management, can be interpreted as assigning the *right* resource to transmit to the *right* user at the *right* time. Cross-layer optimization techniques, which have proven useful, may also help to design resource allocation or user scheduling algorithms that optimize the achievable SE-EE tradeoff. A comprehensive survey on the techniques for energy-efficient wireless communication from time, frequency, and spatial domains can be found in [8], and it may serve as a good tutorial. In advanced network architectures such as HetNet and CoopNet, the system may benefit even more from the joint design of physical transmission and resource management. Our recent work in [10] presents initial results in relay-assisted cooperative systems.

BW-PW TRADEOFF

Bandwidth and PW are the most important but limited resources in wireless communications. From Shannon's capacity formula, the relation between transmit power and signal bandwidth for a given transmission rate, R, can be expressed as

$$P = WN_0\left(2^{\frac{R}{W}} - 1\right). \tag{2}$$

The above expression shows a monotonic relation between PW and BW as sketched in Fig. 2b. It can easily be seen from the above expression that the minimum power consumption is as small as N_0R ln 2 if there is no bandwidth limit.

The fundamental BW-PW relation in Fig. 2b shows that for a given data transmission rate, the expansion of the signal bandwidth is preferred in order to reduce the transmit power and thus achieves better energy efficiency. In fact, the evolution of wireless systems exhibits the same trend for bandwidth demand. For example, in GSM systems, bandwidth per carrier is 200 kHz while it is 5 MHz in UMTS systems. In future wireless systems, such as LTE or LTE-Advanced, system bandwidth is 20 MHz and may even reach as wide as 100 MHz if some techniques, such as carrier aggregation (CA),[6] are used.

The BW-PW relation is also crucial to radio resource management. In [11], it has been exploited to determine the "green" transmission strategy, which first senses and aggregates the unused spectrum using cognitive radio (CR) techniques, and then adjusts the modulation order according to the available BW each time. However, in practical systems, the circuit power consumption, such as filter loss, actually scales with the system BW, which entangles the BW and PW relation as shown in Fig. 2d. Furthermore, Fig. 4 illustrates a visual example of the three-dimensional relation among PW, BW, and EE. From the figure, we have the following two observations.

- If the circuit PW scales with the transmission BW (fixed power spectrum density), full utilization of the bandwidth-power resources may not be the most energy-efficient way to provide the wireless transmission under fixed transmission rate.
- Given a target EE, the BW-PW relation is non-monotonic.

Although the BW-PW trade-off was noticed decades ago, there are still many open issues that deserve future investigation. Some of them are:

- Advanced techniques for BW-PW trade-off with practical concerns

[5] *Also known as drain efficiency, defined as the ratio of output power over input power.*

6 *Carrier aggregation (CA) is a technique that enables aggregation of multiple component carriers (basic frequency blocks) into overall wider bandwidth. CA is among the main features in LTE-Advanced.*

- Novel network architectures and algorithms to improve BW-PW trade-off

As we know, the second-generation (2G) and 3G wireless communication systems, such as GSM and UMTS, use fixed BW transmission, leaving no space for dynamic BW adjustment. With the evolution of wireless technologies, the future deployment of LTE or LTE-Advanced systems provides more flexibilityin spectrum usage so that the transmission BW can be tuned for different applications. Meanwhile, technologies such as spectrum refarming,[7] CA, and software defined radio (SDR)-based CR techniques are maturing to support the flexible use of BW. However, the implementation and integration of these technologies will incur extra overhead in practical systems. For example, CA requires multiple radio frequency (RF) chains and CR needs additional energy for sensing. Therefore, we shall pay more attention to how these technologies can be integrated efficiently.

On the other hand, the deployment of advanced network architecture may also change the shape of the BW-PW trade-off frontier. In particular, the deployment of CoopNet and HetNet introduces additional infrastructure nodes into the network; consequently, the BW and PW planning will be different from that for conventional network architectures. Hence, the BW-PW trade-off with advanced resource management algorithms under new network architectures deserves future research. In addition, with the combination of CA and CR techniques, cross-layer approaches that jointly consider dynamic BW acquisition and BW-PW trade-off will certainly play important roles in future design.

DL-PW Trade-off

In the tradeoffs described above, the metrics such as DE, SE, and BW, are either system efficiency or resource, which are more physical layer oriented. Different from these metrics, DL, also known as service latency, is a measure of QoS and user experience and is closely related to the upper layer traffic types and statistics. As a result, the design of transmission schemes shall cope with both channel and traffic uncertainties, which makes the characterization of DL-PW tradeoff more complicated.

In early mobile communication systems, such as GSM, the service type is very limited and focuses mainly on voice communications. The traffic generated in voice service is continuous and constant, so fixed rate coding and modulation schemes are good enough. In this case, the DL between the transmitter and the receiver mainly consists of signal processing time and propagation delay. Hence, there is not much we need to do. However, the types of wireless services become diverse as technologies evolve and the ability of mobile terminals enhances the popularity of mobile http service, multimedia message service, and multimedia video service. Future networks must deal with various applications and heterogeneous DL requirements. Therefore, in order to build a green radio, it is important to know when and how to trade tolerable DL for low power.

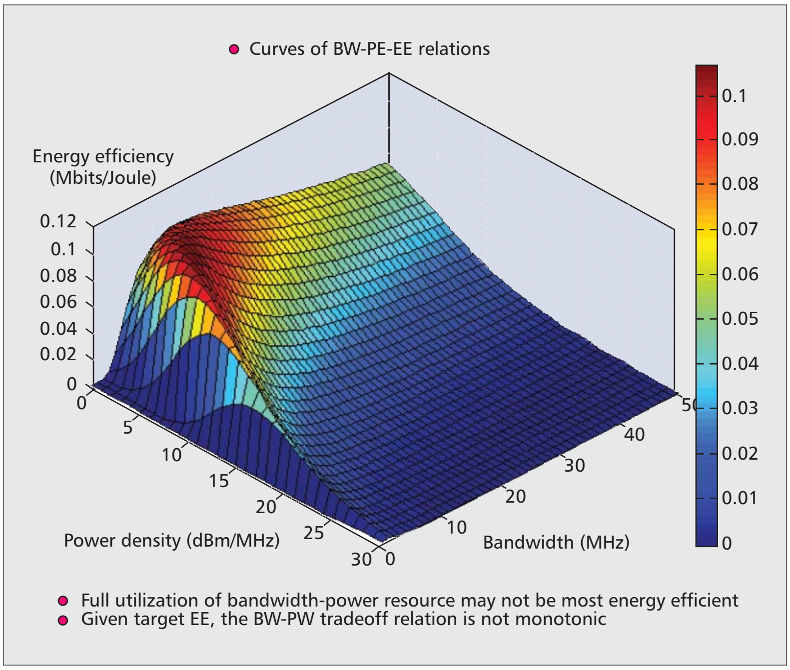


Figure 4. *Results on BW-PW-EE relation for fixed transmission rate.*

To understand the DL-PW trade-off, let us start with the simplest case first, excluding the impact of both channel and traffic dynamics. For point-to-point transmission over AWGN channels, Shannon's formula tells us that

$$R = W \log_2\left(1 + \frac{P}{WN_0}\right)$$

bits of information are transmitted each second; hence, it takes $t_b = 1/R$ s to transmit a bit. Therefore, the average power per bit can be expressed as

$$P_b = WN_0 t_b (2^{\frac{1}{t_b W}} - 1). \qquad (3)$$

The above expression shows a monotonically decreasing relation between per bit PW and DL as sketched in Fig. 2b. Also note that

$$\frac{1}{t_b W} = \frac{R}{W}$$

can be regarded as the modulation level for an uncoded communication system. Then the transmit power per bit decreases as the modulation level is reduced. However, as in all three other trade-off relations, once we take practical concerns into consideration, such as circuit power, the trade-off relation usually deviates from the simple monotonic curve and may appear like a cup shape as sketched in Fig. 2d.

The DL-PW relation with traffic dynamics is more complicated. In this case, the service DL should include both the waiting time in the traffic queue and the time for transmission; the sum of these two parts is also known as queueing DL. In addition, when traffic flow is considered, average DL per packet will be used instead of

[7] *Spectrum refarming is more like a government action to support more efficient use of wireless spectrum via reassigning 2G spectrum to 3G applications. For instance, it is now possible to deploy UMTS (3G system) on 900 MHz (2G spectrum).*

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

As the market develops, wireless networks will continue to expand in the future. Green evolution, as a result, will continue to be an urgent demand and inevitable trend for operators, equipment manufacturers, as well as other related industries.

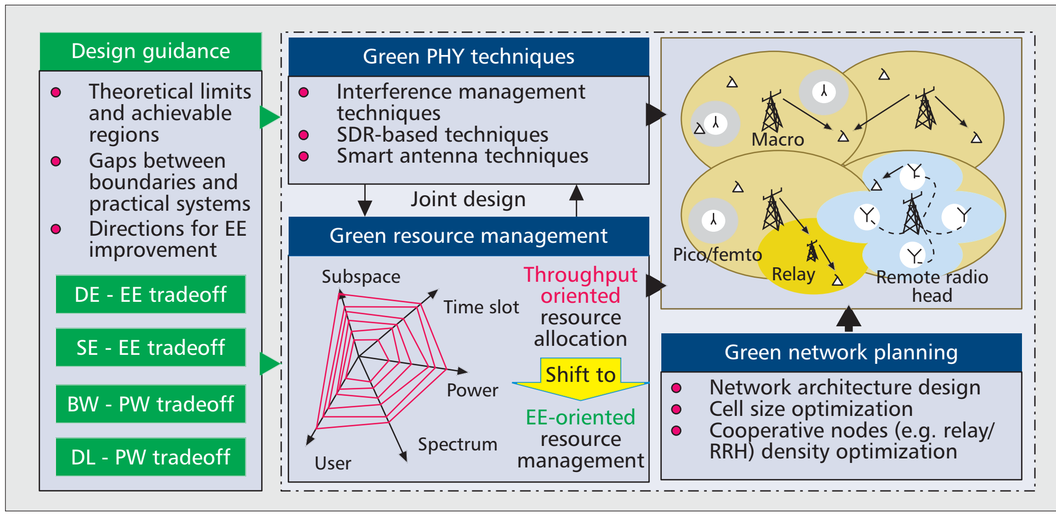


Figure 5. *An overview of how the fundamental framework guides specific system designs.*

average DL per bit. The basic trade-off in Eq. 3 has been extended to the finite packets scheduling in [12]. A lazy schedule was proposed to minimize the total transmission power while guaranteeing the transmission of all packets to be finished before a predetermined time. A benchmark paper takes both channel uncertainties and random traffic into consideration [13]. However, the mathematical model there is very complicated since both information theory and queueing theory are involved. Nevertheless, the results there are only for the point-to-point case,[8] and more open issues need to be addressed, including:

- DL-PW trade-off for heterogeneous DL requirements in multi-user/multicell scenarios
- Joint design of physical layer transmission schemes and resource management to improve DL-PW trade-off with consideration of practical concerns
- Simplified and insightful but approximate mathematical models for DL-PW relation

From queueing theory, we know that the average DL of a packet queue is determined by the statistics of the traffic arrivals and departures. Usually, the departure rates are closely related to the transmission schemes and available radio resources . In multi-user/multicell environments, however, the system resources are shared among different users and also among various application streams, which makes the departure rates of different queues correlated with each other. Consequently, the network DL-PW relation needs to be considered, and the mathematical model becomes even more complicated. In general, there is no closed form expression available to show the direct relation between DL and PW. Therefore, the investigation of simplified but approximate models is desired to provide insights for practical system design. On the other hand, due to correlation among queues, user scheduling and resource allocation algorithms are crucial to control the operation point that maximizes network power efficiency while balancing the heterogeneous DL requirements.

[8] *There has been progress on the DL-PW trade-off research in recent years. Interested readers may refer to the following link: http://ee.usc.edu/stochastic-nets/wiki/dokuwiki-2008-05-05/doku.php.*

Conclusions

In this article, we have proposed a framework for GR research to integrate the fundamental connections that are currently scattered. Four fundamental trade-offs constitute the skeleton of the framework. We have shown that, in practical systems, the trade-off relations usually deviate from the simple monotonic curves derived from Shannon's formula as summarized in Fig. 2. Moreover, most of the existing literature mainly focuses on the point-to-point single-cell case. Therefore, the trade-off relations under more realistic and complex network scenarios deserve further investigation. The insights, such as how to improve the trade-off curves as a whole and how to tune the operation point on the curve to balance the specific system requirements, are expected to guide practical system designs toward green evolution, which will be our next steps following this piece of work. Figure 5 demonstrates a whole picture of how the proposed framework will impact the green design of future systems. As the market develops, wireless networks will continue to expand in the future. Green evolution, as a result, will continue to be an urgent demand and inevitable trend for operators, equipment manufacturers, as well as other related industries. Progress in fundamental GR research, as outlined in this article, will certainly help in making a green future.

References

[1] T. Edler and S. Lundberg, "Energy Efficiency Enhancements in Radio Access Networks," *Ericsson Review*, 2004, http://www.ericsson.com/ericsson/corpinfo/publications/review/2004 01/files/2004015.pdf

[2] R. Irmer, "Evolution of LTE — Operator Requirements and Some Potential Solutions," *Proc. 5th Int'l. FOKUS IMS Wksp.*, Berlin, Germany, Nov. 2009.

[3] R. Tafazolli, "EARTH — Energy Aware Radio and Network Technologies," *Proc. Next Generation Wireless Green Networks Wksp.*, Paris, France, Nov 2009.

[4] K. Johansson, *Cost Effective Deployment Strategies for Heterogeneous Wireless Networks*, Ph.D. dissertation, KTH Info. and Commun. Tech., Stockholm, Sweden, Nov. 2007.

[5] B. Badic *et al.*, "Energy Efficiency Radio Access Architectures for Green Radio: Large Versus Small Cell Size Deployment," *Proc. IEEE VTC '09-Fall*, Anchorage, AK, Sept. 2009.

[6] Y. Chen, S. Zhang, and S. Xu, "Characterizing Energy Efficiency and Deployment Efficiency Relations for Green Architecture Design," *Proc. IEEE ICC '10*, Cape Town, South Africa, May 2010.
[7] F. Richter, A. J. Febske, and G. P. Fettweis, "Energy Efficiency Aspects of Base Station Deployment Strategies in Cellular Networks," *Proc. IEEE VTC '09-Fall*, Anchorage, AK, Sept. 2009.
[8] G. Miao *et al.*, "Cross-Layer Optimization for Energy-Efficient Wireless Communications: A Survey," *Wiley J. Wireless Commun. and Mobile Computing*, vol. 9, Apr. 2009, pp. 529–42.
[9] G. Miao *et al.*, "Interference-Aware Energy-Efficient Power Optimization," *Proc. IEEE ICC '09*, Dresden, Germany, June 2009.
[10] S. Zhang, Y. Chen, and S. Xu, "Energy Efficient Resources Allocation for Relay-Assisted Networks," *Proc. World Wireless Research Forum*, Penang, Malaysia, Apr. 2010.
[11] D. Grace *et al.*, "Using Cognitive Radio to Deliver Green Communications," *Proc. 4th Int'l. Conf. Cognitive Radio Oriented Wireless Networks and Commun.*, June 2009.
[12] E. Uysal-Biyikoglu, B. Prabhakar, and A. E. Gamal, "Energy-Efficient Packet Transmission over A Wireless Link," *IEEE/ACM Trans. Net.*, vol. 10, Aug. 2002, pp. 487–99.
[13] R. A. Berry and R. Gallager, "Communication Over Fading Channels with Delay Constraints," *IEEE Trans. Info. Theory*, vol. 48, May 2002, pp. 1135–48.

Biographies

YAN CHEN (eeyanchen@huawei.com) received her B.Sc. degree from Chu Kochen Honored College, Zhejiang University, Hangzhou, China, in 2004. She continued her further study at the Institute of Information and Communication Engineering, Zhejiang University, from which received her Ph.D. degree in 2009. During her Ph.D. study, she was a visiting researcher in the group of Prof. Vincent Lau at the Department of Electronic Engineeering, Hong Kong University of Science and Technology (HKUST), for one and a half years. After graduation, she joined Huawei Technologies (Shanghai) Co., Ltd. Currently she is working as a research engineer in the Green Radio Excellence in Architecture and Technology (GREAT), which focuses on power-efficient solutions for wireless radio access networks. Her research interests include green network information theory, energy-efficient network architecture and management, as well as the cross-layer design of radio resource scheduling algorithms that balance energy efficiency, quality of service (QoS), and signaling cost.

SHUNQING ZHANG (sqzhang @huawei.com) received his B.E. degree from Fudan University, Shanghai, China, in 2005 and his Ph.D. degree from HKUST in 2009, respectively. He joined the GREAT project at Huawei Technologies Co. Ltd. after graduation. His current research interests include energy-efficient resource allocation and optimization in cellular networks, joint baseband and radio frequency optimization, and other green radio technologies for energy saving and emission reduction.

SHUGONG XU [SM] (shugongg@huawei.com) received a B.Sc. degree from Wuhan University, China, and his M.E. and Ph.D. from Huazhong University of Science and Technology (HUST), Wuhan, China, in 1990, 1993, and 1996, respectively. He is currently director of the Access Network Technology Research Department, principal scientist, and vice director of the Communication Laboratory, Huawei Corporate Research, and chief scientist on Wireless Access Technologies in the National Key Laboratory on Wireless technologies at Huawei. He leads green research in Huawei including the GREAT project, which focuses on power-efficient solutions for wireless radio access networks. Prior to joining Huawei Technologies in 2008, he was with Sharp Laboratories of America as senior research scientist for seven years. Before starting his career in industrial research, he worked at universities including Tsinghua Univiersity, China, Michigan State University, as well as City College of New York (CCNY), where he published his most referenced paper which challenged the common assumptions and revealed the roots of performance issues of 802.11 MAC protocol in multihop network environments. In his 16+ years work on cutting edge research on wireless/mobile networking and communication, home networking, and multimedia communications, he published more than 30 peer-reviewed research papers as lead author in top international conferences and journals. He holds about 30 granted U.S. patents or patent applications. Some of these technologies have been adopted in international standards like 802.11e, DLNA, and the latest 3GPP LTE.

GEOFFREY YE LI (liye@ece.gatech.edu) received his B.S.E. and M.S.E. degrees in 1983 and 1986, respectively, from the Department of Wireless Engineering, Nanjing Institute of Technology, China, and his Ph.D. degree in 1994 from the Department of Electrical Engineering, Auburn University, Alabama. He was a teaching assistant and then a lecturer with Southeast University, Nanjing, China, from 1986 to 1991, a research and teaching assistant with Auburn University, Alabama, from 1991 to 1994, and a post-doctoral research associate with the University of Maryland at College Park from 1994 to 1996. He was with AT&T Labs — Research at Red Bank, New Jersey, as a senior and then principal technical staff member from 1996 to 2000. Since 2000 he has been with the School of Electrical and Computer Engineering at Georgia Institute of Technology as an associate and then a full professor. He has also held the Cheung Kong Scholar title at the University of Electronic Science and Technology of China since March 2006. His general research interests include statistical signal processing and telecommunications, with emphasis on OFDM and MIMO techniques, cross-layer optimization, and signal processing issues in cognitive radios. In these areas, he has published about 200 papers in refereed journals or conferences and filed about 20 patents. He also has two books. He once served or is currently serving as an editor, a member of editorial board, and a guest editor for about 10 technical journals. He organized and chaired many international conferences, including technical program vice-chair of the 2003 IEEE International Conference on Communications. He was named an IEEE Fellow for his contributions to signal processing for wireless communications in 2005, selected as a Distinguished Lecturer from 2009–2010 by IEEE Communications Society, and won the 2010 IEEE Communications Society Stephen O. Rice Prize Paper Award in the field of communications theory.

Degrees of Freedom for Energy Savings in Practical Adaptive Wireless Systems

Hanna Bogucka, Poznan University of Technology, Poland
Andrea Conti, University of Ferrara, Italy

ABSTRACT

We present a new design concept for adaptive wireless communications with new trade-offs between system performance and energy consumption. The system performance, in terms of bit error rate, outage probability, and achieved spectral efficiency, depends on constellation signaling, applied diversity, and channel estimation. Resources dedicated to channel estimation and feedback traffic contribute to the overall system and network energy consumption, and the resulting CO_2 emission. We consider the trade-offs among different methodologies and parameters toward an energy-efficient *green communication* system design. Below, we discuss the degrees of freedom in the design of communications systems with imperfect channel estimation and diversity, and investigate their energy saving options. We present the case studies of single- and multicarrier systems applying both margin-adaptive and rate-adaptive pilot-assisted transmission.

INTRODUCTION

Many communities and countries are continuously making efforts in legislation, regulations, and recommendations to all sectors of industry and agriculture, as well as in promoting ecology, to address the issues of sustainable development.[1] The information and communication technologies (ICT) domain significantly contributes to the global warming. The well-known indicative figure of 3 percent of worldwide energy is consumed by the ICT infrastructure (of which about 80 percent is consumed by base stations amounting to 5–10 million kWh/yr), which causes about 2 percent of worldwide CO_2 emissions. As an example, power consumption and CO_2 emission data for cellular networks as of today are presented in [1].[2] There, it has been shown that around 3 billion mobile stations (MSs) worldwide consume 0.2–0.4 GW of power, which results in approximately 1–2 Mt of CO_2 emission per year. Moreover, around 3 million base stations (BSs) consume 4.5 GW of power and cause approximately 20 Mt of CO_2 emission per year.

Lowering energy consumption of future wireless systems demands considerable attention at all stages of telecommunication systems. A holistic approach to wireless communication eco-sustainability assumes that attention should be given to multiple components of a network to obtain noticeable power savings in the communication network as a whole. As an example, the breakdown of conventional cellular system BS power consumption is presented in Table 1. The data provided show that 50–80 percent of the total power (energy averaged over a long period of time, e.g., a year) consumed by a cellular system BS is the power consumed by the transmit-power amplifier and the feeder. This means that the main challenge (and opportunity in the same time) lies in engineering flexible transmission techniques that are able to reduce the BS transmit-power. Similar conclusions can be derived for mobile stations.

For wireless local area networks (WLANs) one sees even more opportunities to reduce the transmitter and receiver power consumption, particularly due to the fact that radio environmental conditions do not change as quickly as they do in mobile scenarios. Table 2 presents the typical breakdown of the WLAN transmitter and receiver power consumption. Again, one can see that asignificant portion of the total consumed power goes to the transmit power amplifier and the radio frequency (RF) front-end at a transmitter, as well as to the processing of the received signal at a receiver.

Transmission methods that add flexibility to the system layers to optimize power consumption create a great deal of opportunity to reduce typical energy consumption. Novel signal processing techniques at the physical layer that increase wireless network capacity (and thus enable transmission energy savings), link adaptation and intelligent network management options for better radio resource utilization, cooperative communication techniques at all layers, and cognitive capabilities of radio devices may all be used to achieve the goal of spectral and energy efficiency.

In this article, we focus on link adaptation methods to reduce the energy consumption of next-generation wireless systems. These methods present multiple degrees of freedom (see, e.g., [4–6]), which sometimes brings a dilemma of what the designer's priorities should be: whether spectral efficiency or energy efficiency is the pri-

[1] *Since the United Nations General Assembly in December 1987 and its Resolution 42/187, sustainable development has become an issue and an aspiration for our civilization.*

[2] *Energy averaged over a long period of time, and over a number of network components, accounting for the network operation with changing traffic intensity, idle modes, and so on, is considered.*

 0163-6804/11/$25.00

Base station components	Percentage of power consumption — range	Percentage of power consumption — typical value
Power amplifier including feeder	50–80%	65%
Power supply	5–10%	7.5%
Signal processing (digital and analog)	5–15%	10%
Air conditioning	10–25%	17.5%

Table 1. *Breakdown of conventional cellular system base station power consumption (based on [2]; Alcatel-Lucent data).*

ority, and in which situations? Further technical questions arise on the proper compromise between multiple efficiency objectives, whether most of the energy assets should be used for signal processing, RF front-end processing, reconfigurable-hardware power supply (which all have an impact on global energy consumption and thus also on global warming), or minimizing energy radiation (which will additionally have a major impact on human health and so-called electromagnetic pollution). The energy assets for both mobile terminals and BSs (or network nodes in general) are usually limited, as per the motto *"All energy is only borrowed, at some point you have to return it."*

Degrees of Freedom in Adaptivity

The green communications approach touches upon the topics of adaptive and flexible radios as well as on broadly understood optimization of radio system attributes. Adaptive modulation and link adaptation techniques in general are based on the very general idea of making the best use of available resources (bandwidth, power, time, circuit, hardware, and computational) to achieve a target quality of service (QoS) and target metrics related to the quality of experience (QoE). Based on the target QoS (e.g., bit error probability or outage probability [7]) and channel state information (CSI), the optimized transmission parameters (usually a pair of the modulation constellation signaling and coding scheme) are opportunistically chosen. The choice of these parameters is often not trivial, and depends on the cost function defined (adaptation criterion) and the constraints imposed on the optimization algorithm. In addition, the overall system performance strongly depends on the accuracy of the CSI, which is affected by estimation errors and the delay of the feedback channel [5, 8].

With respect to the adaptation criteria, there are multiple possibilities and combinations of adaptation schemes, in which some particular metrics and quantities may be inputs (requirements and constraints) for one strategy and the criteria or output for another. By varying these choices, a number of variants of adaptation strategies can be obtained. For example, in the throughput-oriented (rate-adaptive) strategy an adaptation algorithm aims at provisioning of the highest bit rate (or spectral efficiency) for a required bit error rate (BER) with a given radiated power limit. This was one of the first adaptive transmission schemes, proposed by Steele and Webb for single-carrier quadrature amplitude modulation (QAM) and narrowband fading channels [4]. Following this idea of exploitation of the time-variant channel capacity, variable coding, rate, and power schemes have been investigated, as well as many other important aspects of link adaptation. The concepts elaborated for adaptive single-carrier QAM modulation and coding have been invoked for multicarrier orthogonal frequency-division multiplexing (OFDM) QAM as well (e.g., in [9]). Further significant advances have been made in the direction of adaptive subcarrier allocation, space-time diversity, and multicode systems, as well as investigation of other key agents affecting adaptive modulation performance.

Energy-Minimizing Adaptive Transmission

A typical and popular adjustable parameter within the framework of link adaptation, especially for energy-constrained wireless networks, is the power consumption of a transceiver for a negotiated level of service (e.g., partial compensation power control for energy saving in cellular systems with a given target outage is presented in [10]). The major motivation behind this type of optimization is to prolong battery life and minimize the impact of electromagnetic radiation on human health. A terminal should be configured, and the transmission parameters adapted, to minimize both the power needed for the baseband signal processing and the power consumption of the RF front-end (including the high-power amplifier, the major source of energy consumption). Thus, the philosophy of the energy-oriented (margin-adaptive) adaptation algorithm design is to find some trade-offs in choosing the optimization criteria (e.g. minimization of the transmitted power within a given limit for the baseband processing) as well as to adopt advantageous simplifications in the link description (e.g., weak subcarriers excision or the concept of using the effective signal-to-noise ratio [SNR] or uncoded BER performance metrics for coded-transmission parameters adaptation [11]) in order to perform a less complex routine toward the suboptimum but QoS-acceptable solution. (The flexibility-adaptivity-reconfigurability concept addressing these issues was presented in [12].)

Adaptive modulation and link adaptation techniques in general are based on the very general idea of making the best use of available resources (bandwidth, power, time, circuit, hardware, and computational) to achieve a target quality of service (QoS) and target metrics related to the quality of experience (QoE).

Transmitter components	Power consumption (mW)	Percentage of power consumption	Power savings options
Power amplifier	530 mW	44%	330 mW (27%) due to power adaptation
RF front-end	288 mW	24%	210 mW (10%) due to TX flexibility
Digital signal processing and MAC	261 mW	22%	Efficient DSP algorithms
Local oscillator	47 mW	4%	—
Digital-to-analog converters	40 mW	3%	—
Baseband front-end	36 mW	3%	—
Receiver components	**Power consumption [mW]**	**Percentage of the power consumption**	**Power savings options**
Analog-to-digital converters	200 mW	27%	—
Error correction decoding	176 mW	24%	Efficient decoding algorithms
Digital signal processing and MAC	176 mW	24%	Efficient DSP algorithms
Baseband front-end	72 mW	10%	—
RF front-end	60 mW	8%	Flexibility in diversity combining
Local oscillator	47 mW	7%	—

Table 2. *Breakdown of the transmitter and receiver power consumption of WLANs (based on [3], state-of-the-art WLAN 802.11A for 54 Mb/s at 13 dBm average transmit power).*

Pilot-Assisted Adaptation with Energy Cost

Energy efficiency in adaptive schemes (e.g., bit and power loading) is typically addressed by considering:

- Margin-adaptive (MA) schemes, where the required QoS parameters such as the data rate and BER are constraints, and the power (or energy) allocated to symbol transmission (at single or multiple subcarriers) is the subject of the minimization[3]
- Rate-adaptive (RA) schemes that maximize the user's data rate and result in better utilization of the spectrum and thus also in lower transmit-power-level requirements

The MA and RA schemes have been discussed extensively in the literature [5, 13], mainly assuming perfect CSI. However, in practical systems, the problem is more complex. To assist accurate channel estimation in wireless links, pilot symbols can be transmitted over each link per coherence time of the channel. Although transmission of pilots is often inevitable, it wastes energy because pilots do not convey any information. For the purpose of green radio communications, this *energy wastage* should be minimized. Given the total energy constraint defined over a particular period of time in single-carrier modulation or for an OFDM symbol, it is up to the system designer to specify the amount of energy devoted to pilots. As seen in the next section, this energy can be adaptively allocated depending on the necessary channel estimation quality. Moreover, if a portion of the available energy is devoted to these pilot symbols, the question arises of how the pilots' energy should be distributed among them, and whether it pays to use a lower number of pilots with higher single-pilot energy or a higher number with lower energy.

Two cases can be considered to examine the influence of the pilots' energy on the BER and spectral efficiency. In the *constant single-pilot energy* case, each single pilot's energy is kept on the same level, independent of the number of pilots used. Thus, due to the total energy/power limits, an increase of the pilots number results in a decrease of the energy assigned to data (information-bearing) symbols. In the second case, called *constant total pilots energy*, the energy assigned to all pilots is constant. The energy assigned to data is also constant, and accounts for a given percentage of the whole transmit energy available over a given period of time related to the parameters adaptation rate. Therefore, if the number of pilots increases, the energy assigned to a single pilot decreases.

Let us note that adaptive schemes require a feedback channel and transmission of feedback information. From the perspectives of link, system, and whole network energy efficiency, as well as from the spectral efficiency point of view, it is desirable that the feedback traffic possibly be limited. Therefore, slow adaptive schemes, which require lower feedback rates, are favorable. In rapidly changing wireless channels, special reception-diversity combining techniques can be applied, and this results in averaging the channel small-scale fading. Application of these diversity techniques may in turn result in higher power requirements for the diversity receiver. The energy efficiency of such a slow adaptive modulation scheme proposed in [6] is addressed in the following section.

Energy-Saving Options for Adaptation: the Code-Book

In practical systems the values of parameters representing the above-mentioned options for adaptivity are derived offline from the theoretical analysis of the given transmission scheme performance, and are assembled into the lookup table (*a codebook*) for adaptive transmission. The codebook constitutes a finite countable set of vectors of the adaptation parameters' values. In conventional adaptive modulation and coding, for instance, the codebook consists of pairs of parameters, the code rate and constellation scheme that satisfy the required target BER and maximize the transmission rate. For more

[3] *Note that power consumption and minimization as discussed in most of the papers translates to energy consumption over a considered period of time, such as symbol or frame time, packet interval, channel coherence time, or another transmission period.*

advanced link adaptation schemes there may be a higher number of adaptation options (larger parameter-vectors) available in the codebook. They are presented in Fig. 1. The choice of the parameters forming the codebook depends naturally on the objectives of the overall network design; in particular, we focus on the *greenness* of communication, and thus the energy efficiency, for a given level of performance.

The above considerations are related mostly to the single-link and single-user case. In literature we find some interesting schemes concerning multi-user adaptive schemes. However, here we concentrate on single-link adaptation schemes for single- and multicarrier transmission that present multiple degrees of freedom in adaptation options resulting in significant energy savings.

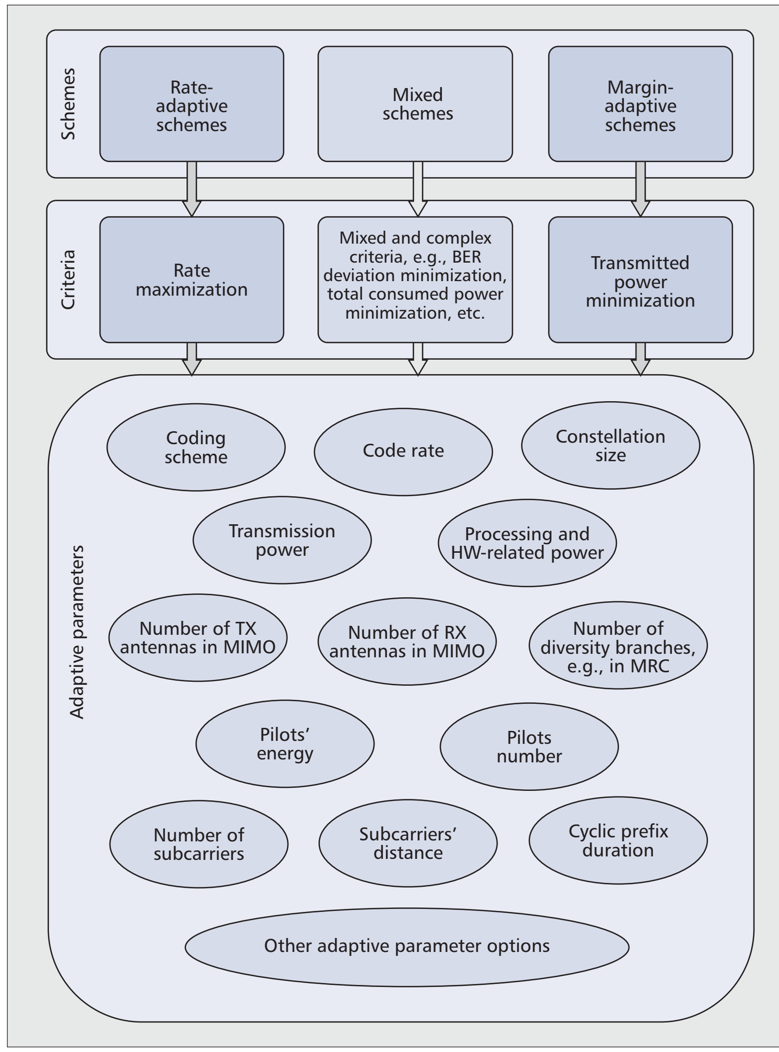


Figure 1. *Energy-saving adaptation options in a practical communication system codebook.*

Energy-Efficient Single-Carrier Adaptive QAM

Adaptive QAM systems have multiple options for adaptivity, tracking small-scale or large-scale fading variations.

Principle

It is worth noting that tracking small-scale fading as in fast adaptive modulation (FAM) with an ideal CSI provides the best performance, at the cost of frequent channel estimation and feedback. With respect to FAM, a gain in terms of feedback and complexity can be achieved by slow adaptive modulation (SAM) techniques, which adapt modulation parameters to slow channel variations (i.e., large-scale fading). Although the spectral efficiency achieved with FAM is slightly higher than that obtained with SAM, SAM achieves a significant improvement in terms of spectral efficiency and outage probability compared to a fixed modulation scheme, despite lower complexity and less frequent feedback [5, 6]. Tracking of slow channel variations in SAM and averaging of the instantaneous channel SNR is possible due to the reception-diversity technique applied. In FAM and SAM, a critical role is played by the channel estimates.

The system parameters that can be adapted are constellation signaling, the number of diversity branches, the combining technique, and the energy devoted to pilots for channel estimation (i.e., their number and energy per pilot within a frame with the duration equal to the coherence time of the tracked process in the channel). Thus, in the codebook there are at least triplets of parameters that can be adjusted for a given diversity combining technique (for coded systems, the coding technique and code rate can also be adjusted).

By adapting system parameters to channel conditions, adaptive modulation techniques allow the transmission spectral efficiency to be maximized without compromising the performance in terms of the bit error probability (BEP) and bit error outage (BEO) [7]. Note that all these triplets of parameters translate to some energy consumption, and thus to energy saving options. The use of a higher constellation size requires higher SNR; the use of higher pilots' energy results in lower energy that could be possibly devoted to data. Finally, the use of a high number of diversity branches is energy consuming, and depends on the type of diversity combining (e.g., as it may require a number of additional antennas and RF front-end chains in the applied diversity combining methods).

Let us consider a slow adaptive M-ary QAM system with a discrete set of possible constellation sizes $\{M_0, M_1, \ldots, M_J\}$. The constellation size is chosen depending on the mean SNR averaged over small-scale fading. When, even with the lowest constellation size M_0, the mean SNR $\bar{\gamma}$ is not able to guarantee the target BEP P_b^*, the system is in outage; otherwise, the transmitted data rate corresponds to $\log_2 M_J$, $j = 0, 1, \ldots, J$. The constellation signaling is chosen by comparing the value of $\bar{\gamma}$ with thresholds that guarantee the target BEP for the available constellation sizes. In particular, when the SNR value falls within the jth region (i.e., $\bar{\gamma}_j^* < \bar{\gamma} \le \bar{\gamma}_{j+1}^*$), the M_j constellation size is adopted. If $\bar{\gamma}$ up-crosses (down-crosses) the SNR threshold, the constellation size is switched to a higher (lower) level, leading to an increase (decrease) of spectral efficiency. To estimate the complex fading level,

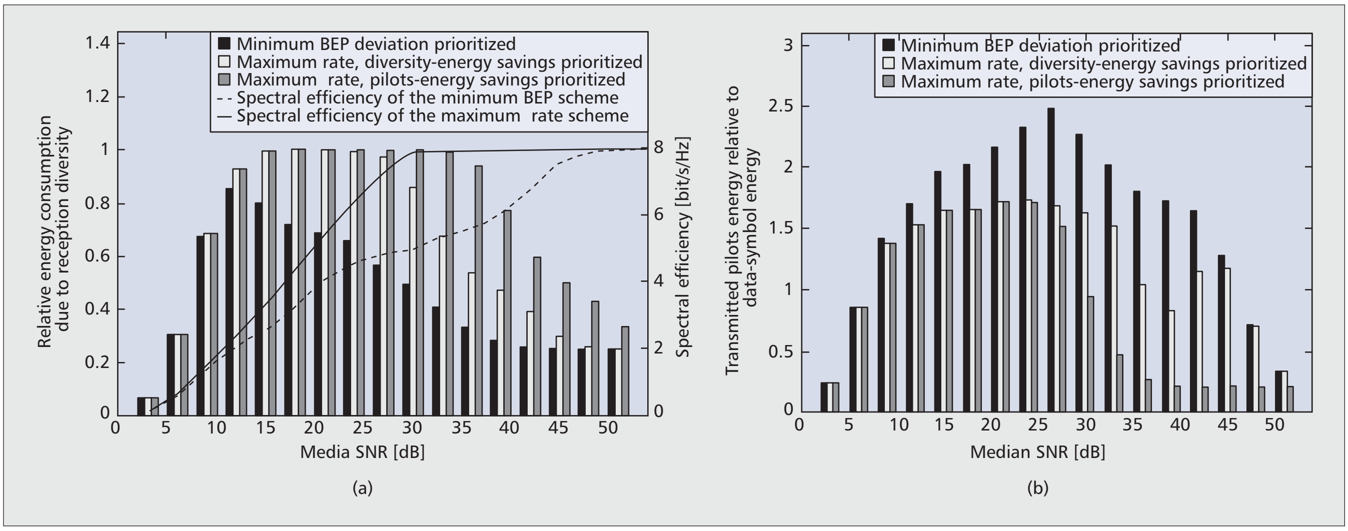


Figure 2. *Relative energy consumption by the diversity branches in use (a) and by the pilot signal (b) for the rate-maximizing scheme; target BER = 10^{-3}.*

pilot symbols are typically transmitted. In short, N_p symbols are inserted in each frame (typically with duration equal to the coherence time of the channel) and transmitted with energy E_p per pilot symbol.

The codebook for the transmission parameters adaptation (the considered lookup table) can be populated offline and based on theoretical analysis of the target BEP. As shown in [6], the BEP for the considered SAM scheme depends on the average SNR and the above-mentioned triplets of parameters. The performance of digital wireless communication systems employing *N*-branch subset diversity in the presence of non-ideal CSI can be found in [8].

Case Study: Slow Adaptive M-QAM

It is up to the system designer and the energy-saving attitude to decide the adaptation strategy option, that is, the strategy for choosing the parameter triplets in response to channel conditions. Let us consider three strategies applied to rate-adaptive schemes over Rayleigh fading and log-normal shadowing. In Figure 2 the energy consumption is presented for these rate-adaptive schemes (for a target uncoded BEP $P_b^* = 10^{-3}$), which pertains to the energy devoted to pilots relative to the data symbols energy (Fig. 2a) and the energy consumption caused by the use of a number of reception-diversity branches (Fig. 2b) relative to the energy required by the maximum considered number of branches (here, equal to 4).

The first scheme considered searches inside the codebook (designed for a given BEP) for the triplet that results in the BEP with the lowest difference from the assumed BER, and not exceeding this BER. Thus, the strategy guarantees the minimum deviation from the target BER, and adjusts all triple parameters jointly, not giving any priority to any one of them. As a result, with the increase of median SNR (equal to the expected value of SNR in dB averaged over the small-scale fading) above a certain value, the spectral efficiency increases (although not as rapidly as in the rate-maximizing schemes discussed below), and both the pilots' energy and the number of required diversity branches decrease. For a median SNR lower than the above mentioned value there is no triplet in the codebook that can satisfy the target BER. It this case we obtain some *unwanted* energy savings due to no transmission.

The second strategy maximizes the rate (and spectral efficiency) of transmission; however, once the maximum allowable rate is chosen from the codebook, the priority is given to the energy savings obtained from minimizing the number of diversity branches. Thus, the triplet from the codebook is chosen in steps, to first find the (possibly multiple) ones that present the maximum constellation size, then among them find the ones that present the lowest number of branches, and finally, if there are still multiple such triplets, choose the one that minimizes the pilots' energy.

The third rate-adaptive strategy considered also tries to maximize the spectral efficiency but with priority given to the energy savings obtained from minimizing the number of pilots. For comparison, the fixed-rate strategies guarantee a fixed transmission rate, while minimizing the values of other adaptive parameters discussed above. In Fig. 3, the energy consumption is presented for the fixed-rate schemes (again for a target BEP $P_b^* = 10^{-3}$). Numerical results are provided for the transmission of 4 and 6 b/s/Hz guaranteed, and for two cases that give priority to the energy savings due to minimization of either the number of diversity branches or the pilots' energy.

Energy-Efficient Multicarrier Adaptive QAM

Adaptive multicarrier modulation (particularly in the context of coded OFDM) has been studied extensively in the past decade. Most of the considered schemes are either RA or MA schemes [13, references therein]. Let us not repeat the vast discussion that has taken place on the spec-

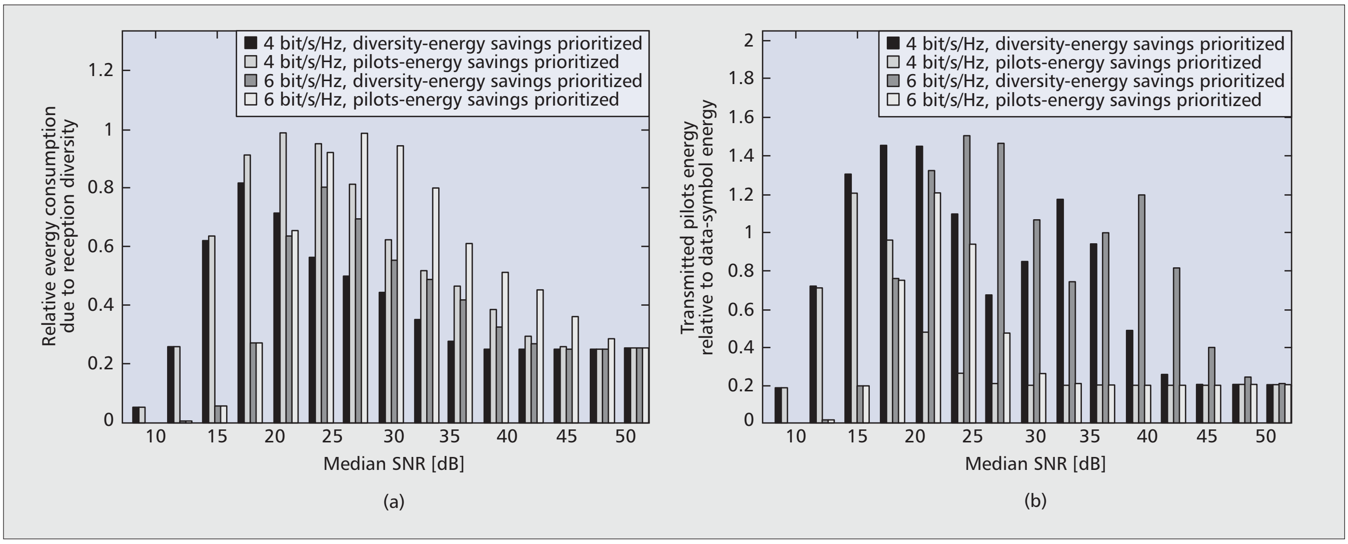


Figure 3. *Relative energy consumption by the diversity branches in use (a) and by the pilot signal (b) for the fixed-rate scheme; target BER = 10^{-3}.*

tral efficiency of the multicarrier adaptive schemes that have been considered for convolutionally and turbo coded OFDM, multiple-input multple-output (MIMO)-OFDM, multicarrier spread-spectrum transmission, and so on. Let us consider more practical adaptive schemes with imperfect CSI that require some energy for the transmission of pilots to assist with channel estimation at the receiver.

Principle

In many papers dealing with the problem of channel estimation quality and adaptive modulation, maximization of the channel capacity and minimization of the estimation error have been the main goals. For example, in [14] it has been noticed that the density of pilots in the OFDM symbol does not have to be determined for the most extreme (worst) channel conditions. Instead, adaptive pilot allocation is possible, which lowers the overhead that could have been used for data transmission. In [15], the optimal pilots-to-data-power ratio has been found for an adaptive MIMO-OFDM system, where minimization of the symbol error rate (impacted by both the channel conditions and channel estimation quality) has been the optimization goal.

Having energy efficiency in mind as our key paradigm, let us consider the influence of the pilots' energy and their density (defined as the pilots number in the OFDM symbol period relative to the total number of subcarrier symbols transmitted in that period) on the adaptive OFDM system performance and its energy consumption. As mentioned above, the constant single-pilot energy (CSPE) and constant total-pilots energy (CTPE) cases are the particular focus. In the first case, each pilot's energy is kept on the same level, independent of pilot density. As a result, under the total OFDM-symbol-energy constraint, an increase of pilot density results in a decrease of the total energy assigned to data. In the second case, the ratio between the energy assigned to pilots and the energy assigned to data in every OFDM symbol period is constant (an increase of pilot density results in lower single-pilot energy, while the data energy remains intact). In both cases, the total energy of the OFDM symbol is also subject to the constraint. (Note that from the interference management point of view, as well as the *green communications* point of view that aims at lowering the energy consumption and human electromagnetic exposure, such a constraint is required for both the data and the pilots.)

Case Study: M-QAM Adaptive OFDM

In Fig. 4, results are presented for the rate-adaptive M-QAM-based OFDM system with pilots for both the CSPE and the CTPE cases. (M equal to 2, 4, 16, 64 or 256 has been taken into account.) An example system has been considered with the number of subcarriers $N_p + N_s = 256$ (where N_p and N_s are the number of pilot symbols and the number of information data-symbols respectively in one OFDM symbol) and the two-paths Rayleigh frequency-selective channel with a delay spread of 1/4 of the OFDM symbol period, and the second path having the average power equal to –6 dB below the first one. The assumed target BER has been 10^{-3}. In the figure, we can observe the energy consumption due to the pilots usage relative to the OFDM symbol energy, as well as the resulting effective spectral efficiency (the amount of information-bearing data per second per Hertz). The results have been obtained for the adaptive uncoded OFDM system for CSPE (with $E_p = 4E_s$) and CTPE (with $N_pE_p = 0.5N_sE_s$) (where E_p and E_s are the single-pilot energy and average single data symbol energy, respectively). The pilots' density and energy have been the subject of adaptation in both cases, and data transmission rate maximization has been the optimization criterion. It would be impractical to create a codebook for all possible combinations of constellation sizes at all subcarriers, pilots' densities, and their energies. Therefore, the results presented in Fig. 4 have been obtained based on computationally complex online optimization. For a finite set of the examined pilots' densities and energies, bit and power loading has been

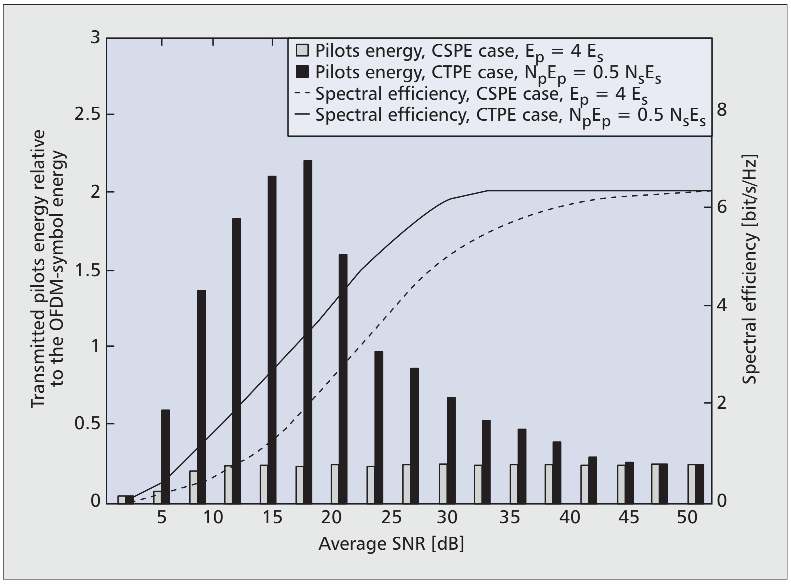


Figure 4. *Relative energy consumption by the pilot signal for the rate-adaptive OFDM scheme and the resulting spectral efficiency in the example CSPE and CTPE cases; target BER = 10^{-3}.*

performed on data subcarriers based on the Campello algorithm (maximizing the transmission rate).

In Fig. 4, one can see that as the average SNR increases, in the CSPE case, we can use pilots of lower density and lower average energy, which results in more subcarriers being available for data transmission or more energy that can be flexibly assigned to data. Consequently, the pilots energy decreases, and the spectral efficiency increases. In the CTPE case, there is less flexibility (i.e., the ratio between the pilots' energy and data energy is fixed). Thus, an increase of the average SNR translates to better channel estimation and higher spectral efficiency. The total transmission energy per OFDM symbol is assumed to be fixed as well as the minimum pilot density approximated by the channel coherence bandwidth. Let us note that for low mean SNR values we obtain some unwanted energy savings due to the transmission outage.

Similarly, as discussed in the previous section, we can also consider more degrees of freedom in pilots' energy adaptation, as well as the options of choosing the optimization criterion (data rate maximization or total transmitted power minimization), or giving a priority to pilots' energy savings or channel estimation quality. For such flexible transmission, a codebook must be created that contains pilots' energy, pilot density for the average SNRs and assumed estimation mean square error, as well as the constellation sizes for possible values of instantaneous SNRs at all subcarriers. (In order to make it practical and feasible, some simplifications in the multicarrier link description are necessary.) The choice from the codebook should take the *green* aspect of the transmission into account to come out with a green transmission model characterized by low energy "wasted" on pilots, and desirable spectral and energy efficiency.

Finally, let us note that a frequency domain method analogous to SAM elaborated for single-carrier QAM can be applied to the multicarrier case. One can consider averaging the channel in the frequency-domain (e.g., using some frequency-diversifying techniques or the *effective SNR* concept, which describes the frequency-selective channel quality with just one parameter, the effective or equivalent SNR), and adopting one modulation (and coding) scheme for all subcarriers. This option would be more practical in populating the codebook for the pilot-assisted multicarrier adaptive scheme discussed above. Alternatively, SAM can be extended to OFDM in a way that applies slow constellation adaptation separately to distinct subcarriers so that the adaptation of the transmission parameters is less frequent than every OFDM symbol interval.

Conclusions

Communication systems are often designed to satisfy target performance metrics in terms of error and outage probability, and spectral efficiency. On the other hand, energy-efficient radios should operate to optimize energy, thus reducing side effects on the climate and human health. Adaptive communication techniques have degrees of freedom to be exploited for energy saving while satisfying target performance metrics, which in turn depend on various system parameters such as the diversity technique and the number of active diversity branches, the energy partition between data and pilot symbols for channel estimation, and constellation signaling. The opportunistic choice of these parameters within a set of possible configurations (codebook) enables the system designer to move the trade-off between performance metrics, spectral efficiency, and energy efficiency. We have shown the case study for both single-carrier and multicarrier communication systems to quantify these energy savings opportunities.

Acknowledgments

The presented work has been supported by the European 7th Framework Program project NEWCOM++ (contract no. 216715). The authors wish to thank Marco Chiani, Moe Z. Win, and Leszek Koschel for helpful discussions, as well as all colleagues in the WPR.3 of NEWCOM++.

References

[1] L. Herault, "Holistic Approach for Future Energy Efficient Cellular Networks," Wireless World Research Forum 23rd Mtg., invited talk, Beijing, China, Oct. 21, 2009.
[2] M. Gruber *et al.*, "EARTH Energy Aware Radio and Network Technologies," *PIMRC 2009 Wksp. Indoor and Outdoor Femto Cells*, Tokyo, Japan, 13 Sept., 2009.
[3] L. Van der Perre *et al.*, "Software Defined Radios: Needs and Opportunities, Technological Solutions and Challenges," tutorial at CROWNCOM 2006, Mykonos Island, Greece, June 8–10, 2006.
[4] W. T. Webb and R. Steele, "Variable Rate QAM for Mobile Radio," *IEEE Trans. Commun.*, vol. 43, no. 7, July 1995, pp. 2223–30.
[5] S. T. Chung and A. J. Goldsmith, "Degree of Freedom in Adaptive Modulation: A Unified View," *IEEE Trans. Commun.*, vol. 49, no. 9, Sept. 2001, pp. 1561–71.
[6] A. Conti, M. Z. Win, and M. Chiani, "Slow Adaptive M-QAM with Diversity in Fast Fading and Shadowing," *IEEE Trans. Commun.*, vol. 55, no. 5, May 2007, pp. 895–905.

[7] A. Conti *et al.*, "Bit Error Outage for Diversity Reception in Shadowing Environment," *IEEE Commun. Lett.*, vol. 7, no. 1, Jan. 2003, pp. 15–17.
[8] A. Conti *et al.*, "Optimized Simple Bounds for Diversity Systems," *IEEE Trans. Commun.*, vol. 57, no. 9, Sept. 2009, pp. 2674–85.
[9] P. S. Chow, J. M. Cioffi, and J. A. C. Bingham, "A Practical Discrete Multitone Transceiver Loading Algorithm for Data Transmission over Spectrally Shaped Channels," *IEEE Trans. Commun.*, vol. 43, Feb./Mar./Apr. 1995, pp. 773–75.
[10] M. Chiani, A. Conti, and R. Verdone, "Partial Compensation Signal-Level-based Up-Link Power Control to Extend Terminal Battery Duration," *IEEE Trans. Vehic. Tech.*, vol. 50, no. 4, July 2001, pp. 1125–31.
[11] I. Dagres *et al.*, "Flexible Radio: A Framework for Optimized Multimodal Operation via Dynamic Signal Design," *EURASIP J. Wireless Commun. and Net.*, vol. 2005, no. 2005:3, pp. 284–97.
[12] A. Polydoros *et al.*, "WIND-FLEX: Developing A Novel Test-Bed For Exploring Flexible Radio Concepts in An Indoor Environment," *IEEE Commun. Mag.*, July 2003, vol. 41, no. 7, pp. 116–22.
[13] T. Keller and L. Hanzo, "Adaptive Modulation Techniques for Duplex OFDM Transmission," *IEEE Trans. Vehic. Tech.*, vol. 49, no. 5, Sept. 2000.
[14] G. Jiann-Ching and A. Osseiran, "Adaptive Pilot Allocation in Downlink OFDM," *IEEE WCNC 2008*, pp. 840–45.
[15] T. Kim, J.G. Andrews, "Balancing Pilot and Data Power for Adaptive MIMO-OFDM Systems," *IEEE GLOBECOM 2006*, Nov. 2006, pp. 1–5.

BIOGRAPHIES

HANNA BOGUCKA (hbogucka@et.put.poznan.pl) received M.Sc. and Ph.D. degrees in telecommunications from Poznan University of Technology (PUT), Poland, in 1988 and 1995, respectively. Since 1988 she has been employed at PUT, currently in the Chair of Wireless Communications as a professor and deputy dean for research on the Faculty of Electronics and Telecommunication. She has research interests in the area of wireless communications, and flexible, adaptive, and cognitive radio systems. She is the author of more than 100 papers and three handbooks (in Polish) in the area of radio communications.

ANDREA CONTI (a.conti@ieee.org) is an assistant professor at the University of Ferrara, Italy. His research interests involve theory and experimentation of wireless systems and networks including network localization, adaptive diversity communications, and cooperative relaying techniques. He is serving as an Editor for *IEEE Communications Letters* and served as an Associate Editor for *IEEE Transactions on Wireless Communications* (2003–2009). He has been elected Vice-Chair of the IEEE Communications Society's Radio Communications Technical Committee (2011–2012).

Green Radio: Radio Techniques to Enable Energy-Efficient Wireless Networks

Congzheng Han, Tim Harrold, and Simon Armour, University of Bristol

Ioannis Krikidis, Stefan Videv, Peter M. Grant, Harald Haas, and John S. Thompson, University of Edinburgh

Ivan Ku and Cheng-Xiang Wang, Heriot-Watt University

Tuan Anh Le and M. Reza Nakhai, Kings College London

Jiayi Zhang and Lajos Hanzo, University of Southampton

ABSTRACT

Recent analysis by manufacturers and network operators has shown that current wireless networks are not very energy efficient, particularly the base stations by which terminals access services from the network. In response to this observation the Mobile Virtual Centre of Excellence (VCE) Green Radio project was established in 2009 to establish how significant energy savings may be obtained in future wireless systems. This article discusses the technical background to the project and discusses models of current energy consumption in base station devices. It also describes some of the most promising research directions in reducing the energy consumption of future base stations.

INTRODUCTION

Given the worldwide growth in the number of mobile subscribers, the move to higher-data-rate mobile broadband, and the increasing contribution of information technology to the overall energy consumption of the world, there is a need on environmental grounds to reduce the energy requirements of radio access networks. A typical mobile phone network in the United Kingdom may consume approximately 40–50 MW, even excluding the power consumed by users' handsets. In developing countries direct electricity connections are not readily available, so Vodafone, for example, use in excess of 1 million gallons of diesel per day to power their network. Mobile communications thus contributes a significant proportion of the total energy consumed by the information technology industry.

From an operator's perspective, reducing energy consumption will also translate to lower operating expenditure (OPEX) costs. Reducing carbon emissions and OPEX for wireless cellular networks are two key reasons behind the development of the Mobile VCE Green Radio program. For example, the U.K. operators Orange and Vodafone both aim to achieve significant reductions in CO_2 emissions in the next 10 years. The Green Radio program sets the aspiration of achieving a hundredfold reduction in power consumption over current designs for wireless communication networks. This challenge is rendered nontrivial by the requirement to achieve this reduction without significantly compromising the quality of service (QoS) experienced by the network's users. In order to meaningfully measure success, appropriate measures of energy consumption must be applied. For example, a reduction in radiated power is not of benefit if it is achieved at the expense of a greater increase in power consumed in signal processing or vice versa.

The Green Radio project is pursuing energy reduction from two different perspectives. The first is to examine alternatives to the existing cellular network structures to reduce energy consumption. The second approach, discussed in detail in the present article, is to study novel techniques that can be used in base stations or handsets to reduce energy consumption in the network. We present the background to the project. We move on to discuss base station modeling, which is a critical issue for the project. We then present three case studies that describe the energy savings obtainable from different techniques that can be employed on wireless links. Finally, we present conclusions to the article.

REDUCING ENERGY CONSUMPTION IN WIRELESS NETWORKS

The specific objective of the Green Radio program is to investigate and create innovative methods for the reduction of the total energy needed to operate a radio access network and to identify appropriate radio architectures that enable such a power reduction. The typical power consumption of different elements of a

0163-6804/11/$25.00 © 2011 IEEE

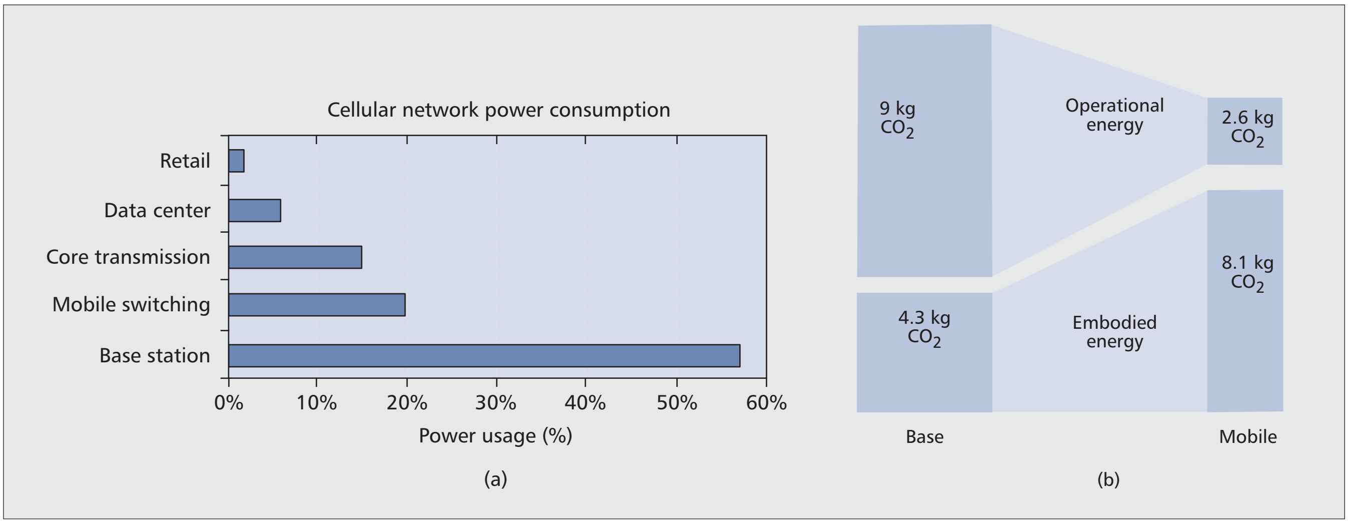


Figure 1. *a) Power consumption of a typical wireless cellular network (source: Vodafone); b) CO_2 emissions per subscriber per year as derived for the base station and mobile handset, after [1]. Embodied emissions arise from the manufacturing process rather than operation.*

current wireless network is shown in Fig. 1a. These results clearly show that reducing the power consumption of the base station or access point has to be an important element of this research program.

Studies have indicated that the mobile handset power drain per subscriber is much lower than the base station component, Fig. 1b [1]; hence, the Green Radio project will mainly focus on base station design issues. Figure 1b also shows that the manufacturing or embodied energy is a much larger component in the mobile handset than in the base station. This is because the lifetime of a base station is typically 10–15 years, compared to a typical handset being used for 2 years. In addition, the energy costs of a base station are shared between many mobile subscribers, leading to a large imbalance in the contribution of embodied energy. From the point of view of handsets, significant efforts need to be put into reducing manufacturing energy costs and increasing handset lifetime, through recycling programs, for example. The Third Generation Partnership Project (3GPP) Long Term Evolution (LTE) system has been chosen as the baseline technology for the research program; its specifications have recently been completed with a view to rolling out networks in the next two to three years [2].

The next section of this article discusses the architecture of existing base stations and identifies key parts of the system hardware where significant energy savings can be obtained.

Base Station Power Efficiency Studies

The overall efficiency of the base station, in terms of the power drawn from its supply in relation to its radio frequency (RF) power output, is governed by the power consumption of its various constituent parts, including the core radio devices.

Radio transceivers: The equipment for generating transmit signals to and decoding signals from mobile terminals.

Power amplifiers: These devices amplify the transmit signals from the transceiver to a high enough power level for transmission, typically around 5–10 W.

Transmit antennas: The antennas are responsible for physically radiating the signals, and are typically highly directional to deliver the signal to users without radiating the signal into the ground or sky.

Base stations also contain other ancillary equipment, providing facilities such as connection to the service provider's network and climate control. A major opportunity to achieve the power reduction targets of the program lies in developing techniques to improve the efficiency of base station hardware.

Analysis within the program has developed models for various base station configurations (macrocell, microcell, picocell, and femtocell) in order to establish how improvements in the hardware components will impact the overall base station efficiency. The starting point for this analysis has been the transmit chain. Near-market power consumption figures have been used in order to establish a benchmark efficiency against which improvements made as part of the project can be assessed. Target power consumption figures allow future overall base station efficiencies to be predicted.

Reference Base Station Architecture

The target system for the base station efficiency analysis is the LTE system with support for four transmit antennas. This system can exploit the space domain to achieve high data throughputs through multiple input multiple output (MIMO) techniques [2]. The reference architecture under investigation is shown in Fig. 2, this represents a macrocellular base station with three sectors, with an effective isotropic radiated power (EIRP) of 27 dBW per sector. The four transmit chains needed for the four antennas therefore require 12 power amplifiers (PAs) and antennas

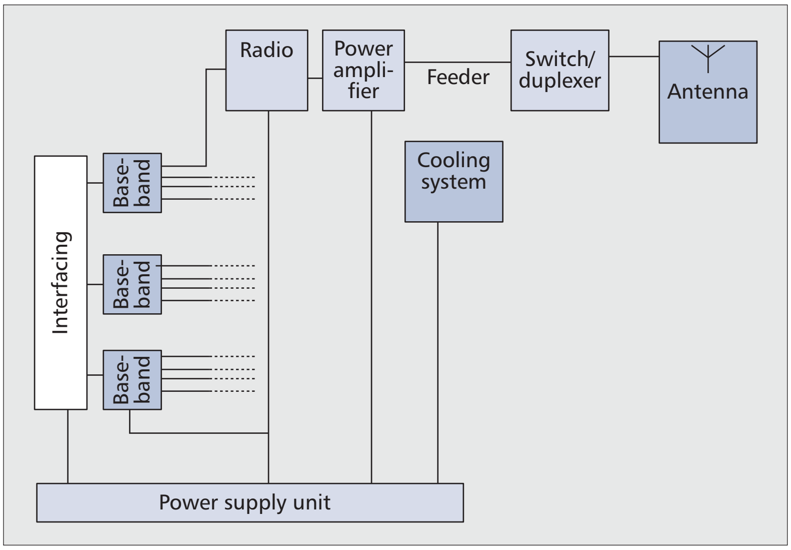


Figure 2. *Reference base station architecture for a system with three sectors and four transmit antennas per sector for MIMO capability. For clarity only one transmit chain is shown.*

per base station. For clarity, only one of the 12 transmit chains is shown in Fig. 2.

Estimated base station power consumption figures for the target system, reflecting the state of the art for the years 2010–2011, are given in Table 1. These estimates have been produced for reference purposes using efficiency figures from [3]; however, to reflect recent innovations [4], a power amplifier efficiency of 40 percent has been used. Two efficiency figures are calculated in Table 1; the top of cabinet (TOC) efficiency gives the ratio of the combined power output of the PAs to the power supply unit (PSU) power (which is used in studies such as [3]), and the radiated efficiency, which refers to the ratio of the efficiency to the total power radiated by the antenna. This second figure therefore includes antenna efficiency and feeder losses.

Target Consumption

The vision for the project is to specify an LTE compliant base station that is able to operate at much lower overall consumption, possibly sufficiently low to enable operation from renewable sources locally generated (e.g., solar or wind). Challenging power consumption targets have been set by the Green Radio program in order to achieve this aim; these target figures are given in the right column of Table 1.

The project target figures show an improvement in efficiency by reducing the power required to operate the base station by at least 50 percent. One solution reduces inefficiencies by locating the PA next to the antennas (typically both at the top of the mast) in order to minimize the power lost in feeder cables. This architecture also further reduces the need for cooling, which could arise were the PAs to be installed in cabinets in an equipment room. Additional efficiency gains are expected to come about by deactivating portions of hardware when unused.

Analysis shows that the greatest potential for increasing the overall base station efficiency comes from improving the efficiency of the PA and antenna, as well as optimizing the power transfer between them. Work underway in the program is seeking to achieve efficiency figures of 85 and 90 percent for these components, respectively. In the case of the PA, one possible approach uses the Class J amplifier [5], which relies on fundamental and second harmonic tuning to achieve high efficiencies, while maintaining the linearity required for LTE operation. In the case of the antenna, the 90 percent efficiency target is to be achieved by exploiting highly efficient dual-polarized patch antenna elements.

Case Studies for Improving Energy Efficiency in Wireless Base Stations

Earlier the power consumption of base stations was discussed and strategies to minimize power use in future base stations was described. In this section, we will move on to consider approaches which are designed around the signals that are transmitted by the base stations. In this case, the time dimension of these waveforms becomes important. In such a case, measures of energy (power × time) rather than just power become important metrics to measure system performance effectively. This section will therefore begin by discussing suitable energy metrics and then move on to discuss three case studies, based around resource allocation, interference cancellation, and the use of multihop relaying strategies.

Overview of Energy Metrics

The results in Fig. 1a of this article show the fact that base stations account for a significant proportion of the total power consumption of a wireless network. If new techniques are proposed to reduce the energy required in the network, it is important to provide meaningful metrics that identify what gains are achieved. The metrics to be used in the Green Radio project have been discussed extensively, and there are two particularly important metrics that are intended to be used during the project.

The first is an absolute measure of energy and is closely related to the industry concept of the energy consumption rating (ECR). This is typically defined as a ratio of peak power divided by the maximum data throughput for a base station transmitter. However, to be of practical use, the ECR should measure the consumed energy per information bit that is successfully transported over the network and is measured in units of joules per bit. This metric allows the absolute performance of different wireless networks to be calibrated. As a simple example, a typical LTE base station sector might operate over a bandwidth of 10 MHz with an average spectral efficiency of 1.5 b/s/Hz, thus achieving an average data rate of 15 Mb/s. If a base station antenna transmits 8 W of RF power (Table 1), the RF ECR value for this system would be 0.53 μJ/b. However, if the total power budget of the

Description	Power In (W)	Power Out (W)	Efficiency	Target Value
Radiated power (per sector)	8	501 (27dBW)	18dBi antenna gain	18dBi antenna gain
Antenna and Switch	12	8	65% efficient	85% efficient
Feeder	24	12	50% efficient	80% efficient
PA (total per sector)	60	24	40% efficient	85% efficient
PA (all sectors)	180	72		
Transceiver (all sectors)	180			70% reduction
Free Air Cooling	40			
Subtotal	400			
PSU Input	450	400	88% efficient	88% efficient
TOC Efficiency			16%	> 25%
Radiated Efficiency			5.3%	> 20%

Table 1. *Estimated power consumption for base stations in 2010–2011 and target future power consumption values for base stations.*

Resource allocation techniques that make the most efficient use of the RF amplifier have the potential to improve energy efficiency significantly. Such energy reductions could lead to further energy savings through switching off transceiver equipment and base station cooling.

base station (e.g., 450 W) is shared among 3 sectors (i.e., 150 W/sector) the ECR value for one sector would increase to 10 μJ/b.

The second metric is a relative measure rather than an absolute one and is more useful for comparing two different systems. Frequently, one may wish to compare the energy performance of a base station using a newly proposed technique (system under test) and compare to a baseline system where the approach is not deployed. The energy consumption gain (ECG) is simply the ratio (E_b/E_t), where E_b is the energy consumed by the baseline system and E_t is the energy for the system under test. The larger the value of the ECG, the more efficient the system under test becomes. However, as with the ECR metric, care needs to be taken to ensure that the energy calculations are performed in a fair manner. For example, if two base station designs are being compared, it should be ensured that both are serving the same number of users under the same traffic load conditions, in order to provide a fair comparison.

Case Study 1: Resource Allocation Strategies

RF amplifiers were identified as a key contributor to the overall energy consumption of a typical base station. In this article we use the term resource allocation to describe how the base station transmitter make the decision of how and when to transmit data to different users on the downlink (base-mobile link) within the cell it is serving. Resource allocation techniques that make the most efficient use of the RF amplifier have the potential to improve energy efficiency significantly. Such energy reductions could lead to further energy savings through switching off transceiver equipment and base station cooling.

In addition, analysis of data traffic in wireless networks show that the traffic load is typically very uneven across the cells. In the analysis of 200 cells in [2, Ch. 9], it is shown that even in peak hours, 90 percent of the data traffic is carried by only 40 percent of the cells in the network. Therefore, techniques that minimize energy consumption across varying traffic load conditions are an important research direction; here we describe two complementary techniques aimed at low and high traffic load conditions, respectively.

Under low traffic load conditions, the base station is likely to have more bandwidth available to transmit data to users than is actually required at that time. One frequency domain approach being studied in the project exploits spare bandwidth resources to reduce energy consumption. Due to the fact that channel capacity scales linearly with the available bandwidth but logarithmically with the radio transmission power, it is possible to trade spectral for energy efficiency, and achieve energy savings while retaining quality of service [6]. Rather than use a complex but spectrally efficient modulation scheme (e.g., 16-quadrature amplitude modulation [QAM]) with a narrow bandwidth, it is possible to use a simpler modulation scheme (e.g., quaternary phase shift keying [QPSK]) with a wider bandwidth.

Figure 3a shows predicted ECG gain results for this approach, as a function for the signal-to-interference-plus-noise ratio (SINR) required at the mobile receiver for a given data rate. Generally speaking, as the spectral efficiency of the data rate increases, so does the required SINR. The value of α specifies the permitted bandwidth expansion factor, and curves are shown for values of α in the range 2–6. For example, a bandwidth expansion of α = 2 would permit 16-

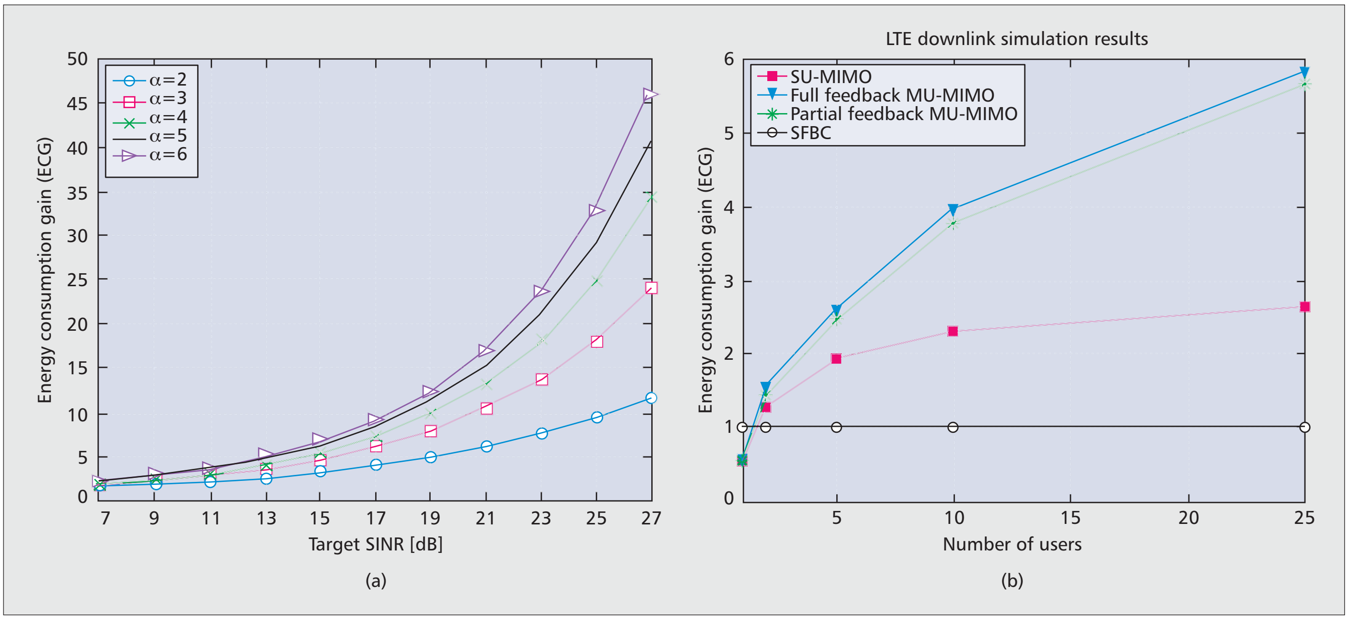


Figure 3. *a) Simulated ECG of the frequency domain bandwidth expansion as a function of the required SNR at the receiver (after [5]); b) simulated ECG of various MIMO schemes, relative to SFBC all at 3 b/s/Hz spectral efficiency (after [6]).*

QAM modulation (4 b/s/Hz maximum data rate) to be replaced by QPSK (2 b/s/Hz maximum data rate), which would require a lower SINR for reliable operation. The results show that as the SINR increases, so does the potential improvement in ECG from using the bandwidth expansion technique. Increasing the value of α beyond four is shown to provide diminishing returns in terms of ECG, except at very high values of SINR where very spectrally efficient modulation schemes would be used.

When the traffic load is high, the base station may be transmitting data to many users simultaneously, possibly using MIMO techniques. In this case, it is usually possible to exploit multiuser diversity to increase the overall multi-user capacity achieved via an opportunistic resource scheduling and allocation strategy. This is where the scheduler assigns resources according to the users' instantaneous channel conditions in the time, frequency, or/and space domains. The performance gains can be translated to further energy reduction at the transmitter. A link adaptation approach is also taken into consideration to ensure the most energy saving transmission mode is employed within the allocated resource for a required QoS level. As an example from [7], Fig. 3b shows the ECG performance of different MIMO precoding schemes compared to using the single-user MIMO diversity scheme space frequency block coding (SFBC) as the baseline case. The multi-user MIMO schemes exploiting a higher degree of diversity achieve lower cost in terms of required transmitter energy for each information bit. When the number of mobile users is large enough, performance evaluation results show that a fivefold energy gain can be achieved by multi-user MIMO through employing appropriate link adaptation and resource scheduling approaches compared to an SFBC system.

Future work in this area will study the best combination of scheduling techniques from an energy efficiency perspective across the range of traffic loads experienced in future LTE networks.

Case Study 2: Interference Management and Mitigation

Interference cancellation schemes are indispensable to combat interference in any practical communication systems where multiple base stations share the same spectrum. The impact of interference is more severe as users move closer to the boundary region between two cells, leading to significant SINR and hence data rate reduction. Most existing interference cancellation schemes have been designed to increase the spectral efficiency and data rate, while overlooking energy efficiency. However, research efforts in the Green Radio program are focused on developing energy-efficient interference cancellation schemes. If the level of interference can be reduced at mobile terminals, it will permit base stations to reduce the wireless transmission energy without compromising the SINR of the wireless link. There are two complementary strategies being considered, as shown in Fig. 4a: distributed antenna systems and receiver interference cancellation.

One way to reduce interference in cellular systems is to coordinate the multiple antennas of the adjacent base stations to form a distributed antenna system (DAS) [8]. For the resulting coordinating DAS, each and every cell edge user is collaboratively served by all of its surrounding base stations rather than only by the single best base station. This permits the interference to users on the cell edge to be effectively controlled and mitigated by coordinated transmit beamforming at all of the participating base stations. The following three schemes can be used by coordinating downlink beamforming:

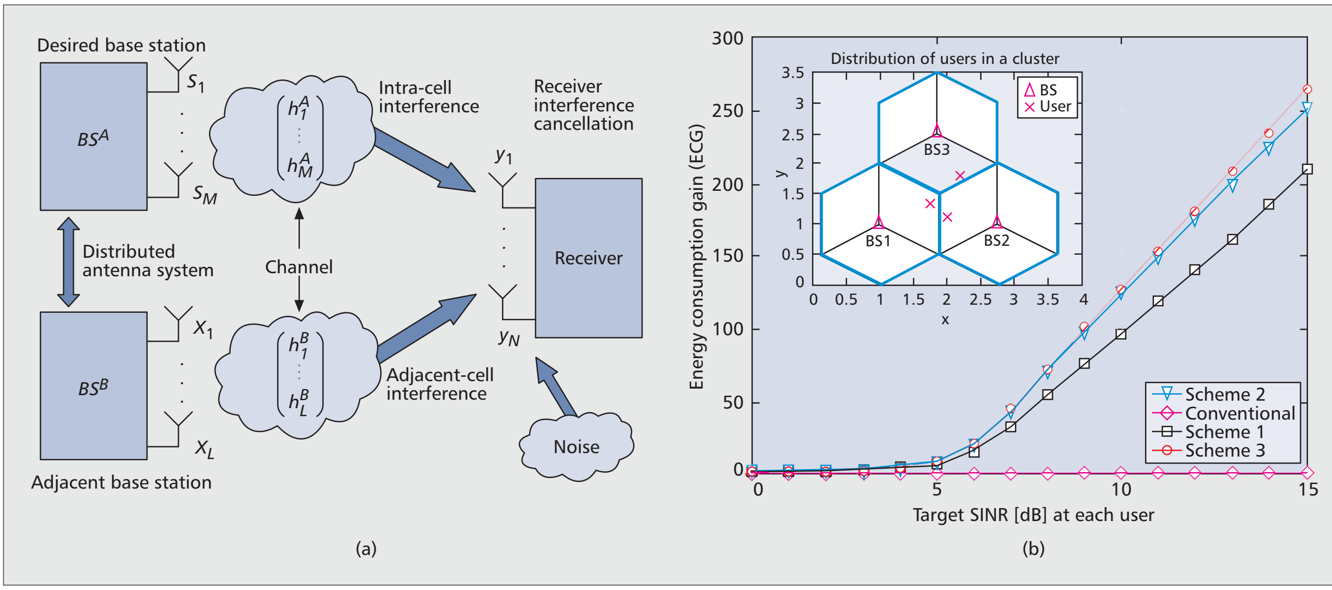


Figure 4. *a) Example scenario for transmitter and/or receiver interference cancellation showing multiple base stations transmitting to a multiple antenna receiver; and b) performance comparison of three DAS schemes, plotting ECG (relative to no DAS case) vs SNR.*

- The user is served by the base station providing the highest SINR while other base stations avoid transmitting signal energy toward that user.
- All users are served by multiple base stations using multiple antenna beamforming and coherent user-end combining (i.e., full exploitation of the interference suppression capability offered by the DAS).
- Users are allocated to one or more base stations based on their position.

These three schemes are compared in terms of ECG vs. SINR against the conventional noncooperative case in Fig. 4b for a cluster of three cells with one user per cell. The results show that all three schemes significantly outperform the conventional system at high SINRs, with schemes 2 and 3 outperforming scheme 1. However, scheme 1 may be preferable over schemes 2/3 in practical implementation, since it requires much less channel data about the users to be exchanged between the base stations and hence less energy consumption.

An alternative scheme to DAS is to apply interference cancellation techniques at a multiple-antenna receiver. The performance of different algorithms has been compared in [9] for different numbers of transmitting antennas. Linear zero forcing (ZF) and minimum mean squared error (MMSE) techniques have been compared, along with nonlinear successive interference cancellation (SIC) variants of these methods. Generally, it is observed that more transmission energy is required as the number of transmit antennas increases. This is expected as intracell interference increases with the number of transmit antennas, resulting in higher transmission energy to maintain the same SINR.

In the absence of co-channel interference from neighboring base stations, it is observed that the MMSE weight optimization approach provides better transmission energy savings than the ZF approach at the desired BS; with the SIC structure performing better than the linear receiver structure. This is because while the ZF criterion nulls out intracell interference but greatly amplifies adjacent-cell interference plus noise, the MMSE criterion jointly minimizes both intracell interference and noise, thus causing less severe amplification to the adjacent-cell interference and noise components. We also observe the same energy consumption trend when three adjacent base stations are present. The ECR values are around 3.4 times poorer than in the absence of co-channel interference for all receivers. This is because traditional interference cancellation (IC) techniques are often implemented at the link level (i.e., the point-to-point link between the desired BS and the receiver in this case). These link-level IC techniques are able to mitigate intracell interference but treat adjacent-cell interference simply as noise. More intelligent methods to cancel adjacent cell interference will be studied in future work, along with consideration of the most energy-efficient combination of IC techniques at both base stations and mobile terminals.

CASE STUDY 3: ENERGY-EFFICIENT ROUTING AND MULTIHOP

In a similar manner to the interference suppression techniques described above, the use of relays to exchange information between a base station and a mobile terminal may be an efficient way to improve base station energy efficiency. This is because the transmission distance can be reduced, increasing data rates or permitting reductions in transmission energy. Relays can enable important reductions of network energy consumption without complicated infrastructure modifications. These may be deployed in streets or buildings to provide improved signal quality to locations that might otherwise experience poor QoS.

In [10], the energy efficiency of several trans-

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

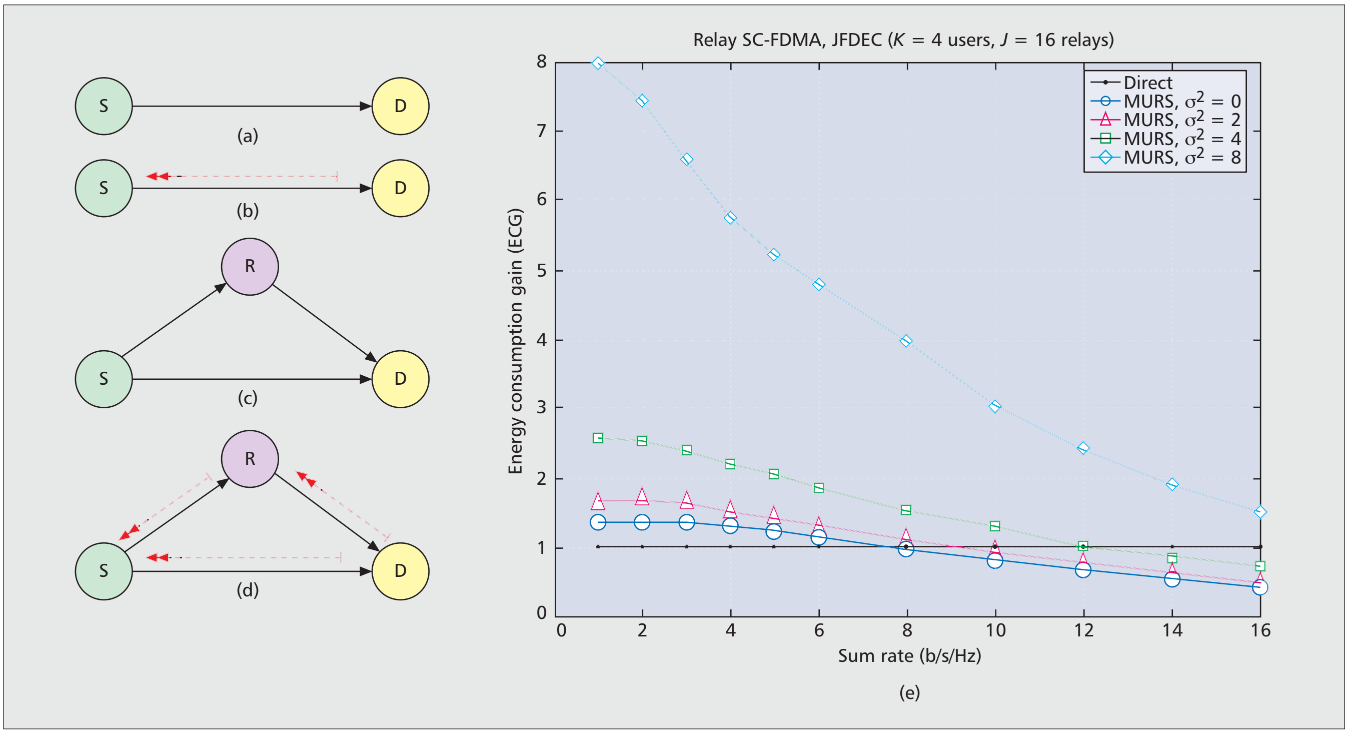


Figure 5. *a) Direct wireless link with average channel knowledge; b) direct wireless link with instantaneous feedback of channel conditions; c) relay link with average channel knowledge; d) relay link with instantaneous feedback of channel conditions; e) performance gains of relay links after [11].*

mission schemes shown in Figs. 5a–d are directly compared. Parts a and b show a conventional base station-mobile station link with average and instantaneous channel state feedback, respectively. Parts (c) and (d) show the case where a relay is present, again with average/instantaneous channel state feedback. It has been shown that the use of instantaneous channel feedback, which is the state of the art for resource allocation schemes, significantly reduces energy consumption compared to the case where only average channel state information is available. On the other hand, the impact of using a relay for communication is known to have a particularly strong impact for high signal-to-noise ratio (SNR) and low packet error rate conditions. This observation is in line with the basic conclusion from the literature that for fixed data rates, relaying is a particularly useful technique for high SNRs (or low packet error rates) because of the presence of the base station-relay-terminal path [11]; in this work, this conclusion is validated from an energy consumption perspective.

In [12], the energy efficiency of opportunistic cooperative relaying designed for the multi-user single-carrier frequency-division multiple access (SC-FDMA) uplink (mobile-to-base link) is investigated with the aid of a single-relay amplify-and-forward (AF) scheme. The AF relay estimates the received power of each subband and equalizes the power differences of the subbands, which corresponds to subband-based equalization. A joint frequency-domain equalization and combining (JFDEC) aided receiver is employed at the base station. In this scenario, there are 4 transmitting terminals and 16 available relays. The energy reduction of the proposed design is a direct benefit of spatial, frequency and selection diversity. In contrast to [11], where no terrain effects, termed shadowing, were considered, they are included in these results. In this case, the shadowing variance becomes an important parameter and expresses the variability in the environment due to buildings and other large obstacles. It may be observed in Fig. 5e that if the SNR is relatively low, the proposed multi-user relay selection (MU-RS) aided cooperative system provides an ECG of up to 8 relative to the no-relay "direct" case when experiencing a shadowing variance of 0–8 dB. However, as the operating SNR increases to a relatively high value and the target data rate increases correspondingly, the benefits of invoking an MU-RS cooperative system erode. This is not unexpected, because sharing the total transmit power between the source and relay as well as the provision of two time slots results in a throughput loss, which is not fully compensated by the relaying gain attained. It is anticipated that similar performance results will be observed for the downlink case as well.

One important future target for the work in this area is to be able to compare the energy efficiency of relay techniques with the use of femtocells. Relays provide a connection to the Internet through the nearest wireless base station. Conversely, femtocells are small low-power base stations installed in the home or office that use a wired Internet connection to provide service. Understanding the full impact of the energy consumption of these differing forms of network connection is an important but challenging task for the Green Radio project.

Conclusions

This article has described the approach being taken in the Mobile VCE project to study novel approaches to reducing the energy consumption of wireless links, particularly in improving the design and operation of wireless base stations. Analysis has shown that when accounting for manufacturing or embodied energy costs, base stations have a much higher operational energy budget than mobile terminals. Proper modeling of the energy consumption of base stations has been shown to be an important issue when trying to obtain a clear view of how different radio technologies can reduce energy consumption. Three case studies of current research in resource allocation, interference suppression, and multihop routing have also been discussed. The means by which these methods can lead to energy savings have been described, and initial results that estimate the performance benefits of these techniques have been presented. The Green Radio project is a three-year program, which started in January 2009 and is starting to deliver initial results, some of which are described and discussed here. The project is being led by industry with the expectation that the most promising research outcomes can feed into future energy-efficient wireless standards and products.

Acknowledgements

The work reported has formed part of the Green Radio Core 5 Research Program of the Virtual Centre of Excellence in Mobile & Personal Communications, http://www.mobilevce.com. This research is funded by the industrial companies who are members of Mobile VCE and by the U.K. EPSRC under grant numbers EP/G060584/1, EP/G06041X/1, EP/G062420/1, and EP/G064105/1.

The authors express their grateful thanks to the industrial leadership team, including Simon Fletcher (NEC), David Lister (Vodafone), Terence Dodgson (Nokia-Siemens Networks), and Andy Jeffries (Nortel).

References

[1] T. Edler, "Green Base Stations — How to Minimize CO_2 Emission in Operator Networks," Ericsson seminar, *Bath Base Station Conf.*, 2008.

[2] H. Holma and A. Toskala, *LTE for UMTS*, Wiley, 2009.

[3] G. Fischer, "Next-Generation Base Station Radio Frequency Architecture," *Bell Labs Tech. J.*, vol. 12, no. 2, 2007, pp. 3–18.

[4] K. Bumman, M. Junghwan, and K. Ildu, "Efficiently Amplified," *IEEE Microwave Mag.*, vol. 11, no. 5, Aug. 2010, pp. 87–100.

[5] P. Wright *et al.*, "A Methodology for Realizing High Efficiency Class-J in A Linear and Broadband PA," *IEEE Trans. Microwave Theory and Techniques*, vol. 57, 2009, pp. 3196–3204.

[6] S. Videv, H. Haas, and P. M. Grant, "Bandwidth-Energy Efficiency Trade-off with Variable Load in LTE," to be submitted to IEEE VTC Spring, Budapest, Hungary, 2011.

[7] K. C. Beh *et al.*, "Power Efficient MIMO Techniques for 3GPP LTE and Beyond," *Proc. IEEEVTC Fall*, Anchorage, AK, Sept. 2009.

[8] T. A. Le and M. R. Nakhai, "'Throughput Analysis of Network Coding Enable Wireless Backhauls," submitted to *IET Commun.*, 2010.

[9] I. Ku, C.-X. Wang, and P. M. Grant, "Impact of Receiver Interference Cancellation Techniques on Base Station Transmission Energy in MIMO Systems," submitted to IEEE ICC, Kyoto, Japan, 2011.

[10] I. Krikidis, J. S. Thompson, and P. M. Grant, "Cooperative Relaying with Feedback for Lifetime Maximization," *Proc. IEEE ICC 2010 Wksp. E2NETS*, Cape Town, South Africa, May 2010.

[11] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, "Cooperative Diversity in Wireless Networks: Efficient Protocols and Outage Behavior," *IEEE Trans. Info. Theory*, vol. 50, no. 12, Dec. 2004, pp. 3062–80.

[12] J. Zhang, L.-L. Yang, and L. Hanzo, "Power-Efficient Opportunistic Amplify-and-Forward Single-Relay Aided Multi-User SC-FDMA Uplink," *Proc. IEEE VTC Spring*, Taipei, Taiwan, May 2010.

Biographies

Congzheng Han received her B.Eng. and Ph.D. degrees in electronic and communications engineering from the University of Bristol, United Kingdom, in 2004 and 2008, respectively. She is currently a member of research staff in the Centre for Communications Research at the University of Bristol and is contributing to the Mobile VCE's Green Radio project. Her research interests include wireless LANs, future-generation standards (LTE), and power-efficient communications.

Tim Harrold is a research fellow in the Centre for Communications Research, University of Bristol. He has worked on a wide variety of wireless projects in areas including indoor/outdoor propagation, self-organising ad hoc networks, smart antennas, ultra-wideband communications, and cognitive radio. This work has been supported by organizations including the UK's Mobile VCE, BT, ERA Technology, and Kyocera. He is currently on secondment to Toshiba Research Europe's Telecommunications Research Laboratory.

Ioannis Krikidis received his diploma from the Computer Engineering and Informatics Department (CEID) of the University of Patras, Greece, in 2000, and M.Sc. and Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 2001 and 2005, respectively. From 2007 to 2010 he was a research fellow in the School of Engineering and Electronics at the University of Edinburgh, United Kingdom. He has now been elected an assistant professor at CEID.

Ivan Ku received his B.Eng. degree in electronics engineering and M.Eng.Sc. degree in communications from Multimedia University (MMU), Malaysia, in 2001 and 2006, respectively. Since April 2009 he has been a Ph.D. student at Heriot-Watt University, Edinburgh, United Kingdom. His research interests include cooperative communication, multiuser detection theory, and digital signal processing.

Tuan Anh Le received his B.Eng. and M.Sc. degrees from Hanoi University of Technology, Vietnam, in 2002 and 2004, respectively. Both are in electronics and telecommunications. He is currently a Ph.D. research student in the Centre for Telecommunications Research, King's College London, and a researcher of the Mobile VCE Core 5 Research Program on Green Radio. His interests are multicell processing, multicell beamforming, and convex optimization.

Stefan Videv got his first two degrees from Jacobs University Bremen (B.Sc. in electrical engineering and computer science, M.Sc. in communication, systems, and electronics). He is currently a Ph.D. student at the Institute for Digital Communications at the University of Edinburgh. His research interests are in cognitive radio, green radio, resource allocation, and interference avoidance.

Jiayi Zhang received his B.Eng. degree from Nanjing University of Aeronautics and Astronautics, China, and his M.Sc. degree from the University of Southampton, United Kingdom, in 2006 and 2007, respectively. He is currently continuing his study toward a Ph.D. degree within the School of Electronics and Computer Science at the University of Southampton. His research interests include SC-FDMA, MIMO-OFDM, frequency-domain equalization, turbo detection, transmitter preprocessing, cooperative relaying, multicell processing, and cross-layer optimization.

Simon Armour is a senior lecturer at the University of Bristol. He has led collaborative projects with a number of industrial partners and now leads Bristol's involvement in the MVCE's Green Radio project. His work has resulted in

The Green Radio project is a three-year program, which started in January 2009 and is starting to deliver initial results. The project is being led by industry with the expectation that the most promising research outcomes can feed into future energy-efficient wireless standards and products.

approximately 100 papers published in international journals and conferences and 11 patents. The majority of the latter are now owned by industry.

PETER M. GRANT [F] is currently a senior professorial research fellow at the University of Edinburgh. He was in 2007 appointed to be the 8th Regius Professor of Engineering at the University of Edinburgh. In 2009 he was made an officer of the Order of the British Empire (OBE) in the Queen's birthday honors list. He holds fellowships of the IET, Royal Academy of Engineering, and Royal Society of Edinburgh.

HARALD HAAS received his Ph.D. degree from the University of Edinburgh in 2001. From 2001 to 2002 he was a research project manager at Siemens in Munich. He joined Jacobs University Bremen in 2002 as associate professor before returning to the University of Edinburgh where he holds a personal Chair in Mobile Communications. His main research interests are decentralized interference management in wireless networks, multiple antenna concepts, and optical wireless communication.

LAJOS HANZO [F] received his Master's degree in electronics in 1976 and his doctorate in 1983 from the Technical University of Budapest. In 2010 he was awarded the university's highest honor, the Doctor Honaris Causa. He is also an IEEE Distinguished Lecturer of both the Communications and Vehicular Societies as well as a Fellow of the IEE/IET and Royal Academy of Engineering.

MOHAMMAD REZA NAKHAI received his Ph.D. degree in electronic engineering from King's College, University of London in 2000. He worked as a postdoctoral research fellow in the Centre for Communication Systems Research, University of Surrey. Since 2001 he has been an academic member of staff in the Centre for Telecommunications Research, King's College London. His current research interests include cognitive radio, cooperative communications, optimization, and communications theory.

JOHN THOMPSON (john.thompson@ed.ac.uk) is currently a reader at the School of Engineering at the University of Edinburgh. He currently leads the Green Radio activities at the university. He has published over 160 journal and conference papers to date. His research interests are in signal processing, multiple antenna communications, multihop techniques, and green communications. He was Technical Program Co-Chair for IEEE GLOBECOM 2010.

CHENG-XIANG WANG received his Ph.D. degree from Aalborg University, Denmark, in 2004. He joined Heriot-Watt University as a lecturer in 2005 and became a reader in 2009. His research interests include wireless channel modeling, cognitive radio networks, vehicular communication networks, green communications, and (beyond) 4G. He has published one book chapter and over 130 papers in journals and conferences. He received the IEEE GLOBECOM Best Paper Award in 2010.

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

Toward Dynamic Energy-Efficient Operation of Cellular Network Infrastructure

Eunsung Oh and Bhaskar Krishnamachari, University of Southern California
Xin Liu, University of California at Davis
Zhisheng Niu, Tsinghua University

ABSTRACT

The operation of cellular network infrastructure incurs significant electrical energy consumption. From the perspective of cellular network operators, reducing this consumption is not only a matter of showing environmental responsibility, but also of substantially reducing their operational expenditure. We discuss how dynamic operation of cellular base stations, in which redundant base stations are switched off during periods of low traffic such as at night, can provide significant energy savings. We quantitatively estimate these potential savings through a first-order analysis based on real cellular traffic traces and information regarding base station locations in a part of Manchester, United Kingdom. We also discuss a number of open issues pertinent to implementing such energy-efficient dynamic base station operation schemes, such as various approaches to ensure coverage, and interoperator coordination.

INTRODUCTION

With increasing awareness of the potential harmful effects to the environment caused by CO_2 emissions and the depletion of non-renewable energy sources, there is a growing consensus on the need to develop more energy-efficient telecommunication systems. It has been estimated that 3 percent of the world's annual electrical energy consumption and 2 percent of CO_2 emissions are caused by the information and communication technology (ICT) infrastructure. Moreover, it is estimated that ICT energy consumption is rising at 15–20 percent per year, doubling every five years [1]. According to one estimate, about a tenth of this can be attributed to cellular mobile communication systems. As of 2008, this corresponded to 60 billion kWh of electricity usage annually, about 40 million metric tons of CO_2 emissions each year. To put it in perspective this is equivalent to annual greenhouse gas emissions from about 8 million cars. Another consistent estimate states that there were over 600 thousand base stations in China deployed by three major operators which consumed about 20 billion kWh in 2007.

From the perspective of cellular network operators, reducing electrical energy consumption is not only a matter of being green and responsible, it is also very much an economically important issue. A significant portion of the operational expenditure (OPEX) of a cellular network goes to pay the electricity bill. From the above, it can be estimated that the mobile network OPEX for electricity globally is more than $10 billion dollars today.

For all these reasons, cellular network operators have been exploring ways to increase energy efficiency in all components of cellular networks, including mobile devices, base stations, and core (backhaul) networks. There has been a tremendous amount of work on mobile device energy efficiency with the objective of prolonging battery life time. Similarly, green operation of Internet has been considered and some of the techniques can be extended to the cellular backhaul networks. However, the key source of energy usage in cellular networks is the operation of base station equipment. It has been estimated that base stations contribute to 60–80 percent of the total energy consumption [2]. In this article, we therefore focus on energy-efficient operation of cellular base stations, which is referred to as green cellular operation.

Energy efficiency with respect to base stations has been considered in all stages of cellular networks, including hardware design and manufacture, deployment, and operation. A number of these efforts have focused on hardware improvements. For instance, next-generation base stations are designed to be substantially more energy efficient, for example, using more energy-efficient power amplifiers and natural resources for cooling. Others have considered collocating cellular base stations with renewable energy sources such as solar power and wind energy. In addition, cellular operators have evaluated deployment strategies that minimize the energy expenditure on base stations [3]. We primarily discuss operation, and touch on deployment.

There is room for significant improvement in cellular operation. Even when a site is experienc-

 0163-6804/11/$25.00 © 2011 IEEE

ing little or no activity, the base station consumes more than 90 percent of its peak energy; for example, a typical universal mobile telecommunications system (UMTS) BS consumes 800–1500 W and has a radio frequency (RF) output power of 40–80 W. While turning off some of the radio transceivers on a BS during low traffic periods can provide some relief, this is still not sufficient. For significant energy savings, what is called for is a more carefully coordinated dynamic approach that allows the system to shut entire base stations and transfer the corresponding load to neighboring cells during periods of low utilization.

In this article, we explore such dynamic base station management mechanisms for cellular networks to understand quantitatively the scope for potential energy savings, and also the technical challenges that arise in implementing these mechanisms, particularly to ensure that energy efficiency does not come at the expense of reduced quality of service for mobile customers. Note that although similar approaches can also be considered for reducing the energy consumption in wireless local area networks such as 802.11-based WLANs, we do not explore these in this work as the electrical energy usage and expense in such networks is an order of magnitude lower than in cellular networks (e.g., Cisco estimates that its wireless control system adaptive power management software can save 7 W/h per Aironet access point).

The article is organized as follows. We present a first-order analysis on the scope of potential energy saving through green cellular operation. We use a simple greedy algorithm to estimate the energy saving based on real cellular traffic traces and actual base station location information in the analysis. We discuss related challenges and potential solutions in green cellular operation, including maintaining coverage, enabling cooperation between operators, and providing E911 service. Last, we conclude the article.

A Case Study

The basic idea behind green cellular operation is to alleviate the inefficiencies resulting from the fact that today's base stations are typically deployed and operated continuously based on peak traffic estimates. Intuitively, it saves energy to carefully turn off unutilized and underutilized base stations during off-peak times, while maintaining coverage. Next, we use real cellular network data to estimate the degree of energy savings possible in urban areas using such a dynamic base station operation.

We look at two sets of real data. The first is a temporal traffic trace shown in Fig. 1, which was obtained from a cellular operator (the source has to be anonymized to preserve confidentiality) in a metropolitan area; and the second is base station location data from a central portion of the city of Manchester, United Kingdom, as illustrated in Fig. 2, obtained from a UK government-sponsored website [4]. We discuss these traces in detail below.

Temporal Traffic Trace

The data trace we have contains voice call information from five base stations — one central base station and its neighboring four base stations — over the course of one week. The data is plotted with a resolution of 1 s in Fig. 1. We normalize the values of load by a constant value to show the relative change. We observe a periodic sinusoidal profile of the traffic in each cell and notice that the traffic during the daytime (11 a.m.–9 p.m.) is much higher than that at nighttime (10 p.m.–9 a.m.). In addition, the traffic profile during a weekend/holiday period, even during peak hours, is much lower than that of a normal weekday.

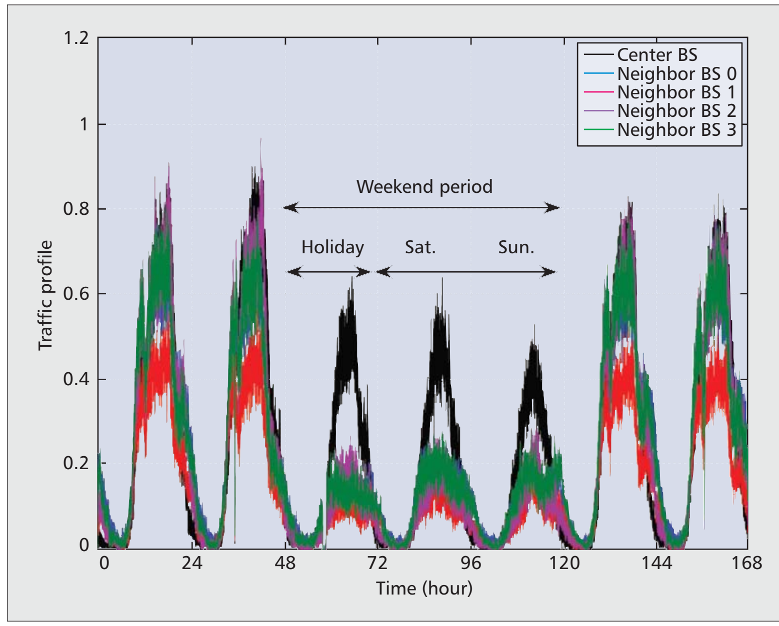


Figure 1. *Normalized traffic profile during one week from real voice call information.*

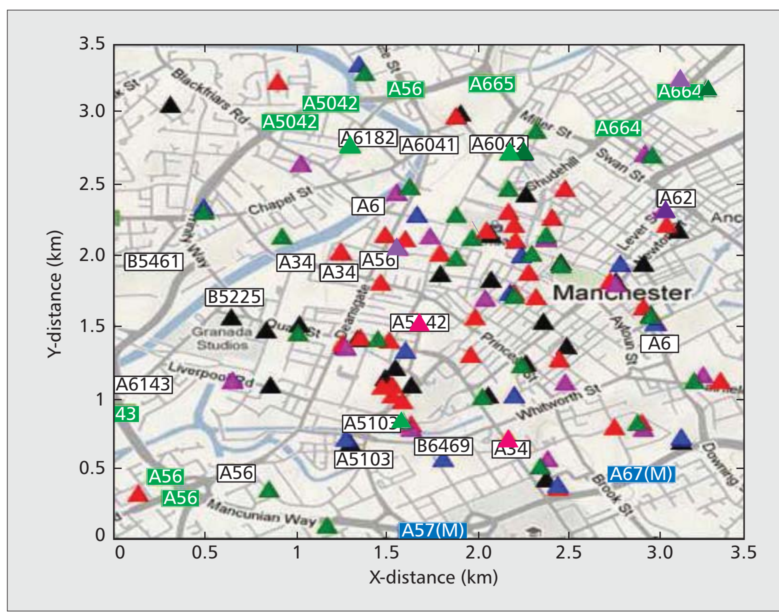


Figure 2. *Base station location data from a part of Manchester, United Kingdom. (Some base stations are collocated, in which case just one triangle is shown on the map.)*

Spatial Deployment

Figure 2 is a snapshot of the base station locations (shown as colored triangles) in part of the city of Manchester, United Kingdom. This infor-

Threshold	Weekday	Weekend	Average per week
5% of peak	23.2	29.8	25.1
10% of peak	30.2	43.3	34.0
20% of peak	38.6	75.6	49.2

Table 1. *Analysis of sample cellular traffic load profiles: percentage of time the traffic is below* x *percent weekday peak during weekdays and weekends, for* x *= 5, 10, 20.*

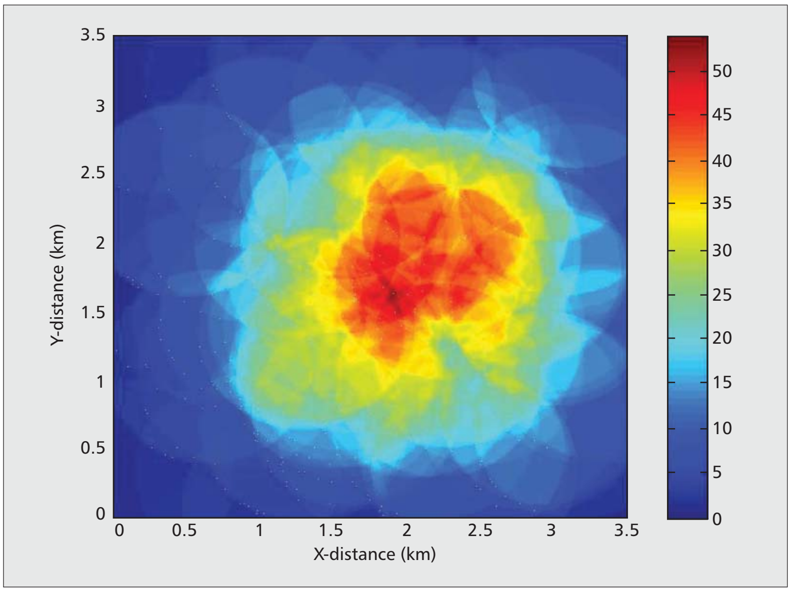


Figure 3. *Redundancy in the cellular coverage of the Manchester area at base station coverage.*

mation is obtained from a U.K. government sponsored interactive website [4] based on information provided voluntarily by mobile network operators on the location and operating characteristics of individual base stations. The information on this site includes location information, operator information, maximum transmission power, and cellular tower height and operating frequency. The information is updated by operators every three months or so [4]. We manually extract base station data from a part of Manchester in the United Kingdom from this website. This area is of size 3.5 × 3.5 km, and has a total of 139 base stations in 128 locations.

Energy Saving Estimate

We now use these real datasets to estimate the energy savings that are possible using dynamic base station operation in this part of Manchester. Note that the anonymized temporal trace data we have does not actually correspond to the base station locations in Manchester. However, our methodology for estimating the energy savings from dynamic base station management decouples the temporal and spatial components, as follows. We use the temporal traffic trace to generate rough estimates of the fraction of time when cells show low activity during weekday and weekend/holiday periods. Independently we analyze the spatial distribution of base stations (on a per-operator basis, as well as jointly) assuming uniform circular coverage zones to estimate the level of redundancy that can be eliminated while maintaining more than 95 percent of the coverage, for different communication ranges. We assume for now that the remaining 5 percent of coverage can be made up through the kinds of techniques discussed later. We combine these two intermediate estimates to generate the final estimate of the expected energy savings.

Based on the traffic load profiles shown in Fig. 1, we obtain the percentage of time that the traffic is below x percent of weekday peak, during weekdays and weekends, where x = 5, 10, 20. The numbers are presented in Table 1. In the following discussion, for ease of exposition, we shall focus on time when traffic is less than 10 percent of the peak (further work is needed to better understand exactly what fraction of peak traffic should be chosen as the threshold for dynamic base station operation). We find that during weekdays, about 30 percent of the time the traffic is less than 10 percent of the peak. During weekends and holidays, this low activity period increases to 43 percent of the time. Assuming two weekend days in a typical week, we estimate that the average low activity period is about 34 percent of the time.

We next consider the level of redundancy present in the spatial base station deployment. As shown in Fig. 2, the deployment shows non-homogeneous density for each operator, with greater density near the city center, which is a primarily commercial district. In the interest of a simplified first-order analysis, we assume that all cells have a common maximum coverage range. (In order to minimize intercell interference, some base stations may be operated below this range when all base stations are activated. We assume that the range of these base stations can be extended to the maximum level when necessary.) If all base stations were to operate at maximum range, there would be significant redundant overlap in coverage. For instance, Fig. 3 shows the coverage of all base stations when a common range of 700 m is chosen for each base station, with colors indicating the extent of redundancy. Near the city center, we find that more than 50 base stations from various operators overlapping if operated at this maximum range.

We first treat each operator separately. For a given maximum range, we use a greedy algorithm to identify a minimal number of base stations such that the net area coverage of these base stations is more than 95 percent of the area covered when all base stations are active. In the greedy algorithm, we sequentially switch off the base station with the minimum distance to its nearest active base station so long as the coverage condition is met. (We discuss later how any remaining coverage holes may be filled by techniques such as range extension, multihop relay, and multipoint coordination). This is plotted in Fig. 4. As expected, the level of redundancy is higher when the maximum range is higher. For example, when the maximum range is 700 m, depending on the operator, from 25 to more than 65 percent of the base stations can be deactivated while maintaining more than 95 percent of the original coverage.

We also plot in this figure the percentage of base stations that could be switched off while preserving 95 percent of the original coverage if all operators were able to share the base station resources. Clearly the overall energy savings are highest in this case (note that from the dataset we have, it appears that some sites are already jointly shared by multiple operators; in this analysis we have assumed that there is a single base station shared by all active operators at such sites). For the 700 m range, we observe that more than 85 percent of the base station can be shut down while keeping greater than 95 percent coverage with minimal overlap in this case. Comparing the total turned-off base stations in the case when operators act individually, we see that base station sharing increases the energy saving by about 35 percent during off-peak hours.

Putting together the temporal and spatial analysis above, we estimate that individual operators can save between about 8 percent to 22 percent of energy in such an urban deployment. Sharing base station resources together, we get a total reduction of about 29 percent for the energy expended on base station operation. Thus an important collateral finding of our analysis is that greater cooperation among operators is essential for substantial savings. We shall discuss this issue further next.

This percentage of energy saving corresponds to between 32 and 60 kWh of absolute energy savings for the roughly 12 km^2 area of Manchester we have considered (assuming the single base station power is between 800 and 1500 W). This in turn translates to about $42,000 to $78,000 annually for the electricity bill for this set of base stations, or about 200 to 375 metric tons of annual CO_2 emissions. This is a substantial reduction in greenhouse gas emission as well as cost of operation.

Some caveats are in order. Our study has not considered heterogeneous networks. Also, an analysis of savings requires better radio propagation modeling. These can be considered in future work.

Challenges and Future Directions

In the above analysis we focused on obtaining a coarse-grained estimate of energy savings if "redundant" base stations could be switched off during periods of low activity. However, this analysis has glossed over many important details pertaining to how such dynamic base station operation can be implemented in practice. A number of questions arise: What are the mechanisms by which coverage can be maintained when subsets of the base stations are switched off? What changes may be needed to make the mobile units more "cognitive" in order to enable more agile base station management across multiple spectrum bands? What is the temporal granularity at which base station operation decisions should be made? At what locality level of the cellular network hierarchical architecture should these decisions be implemented? Is there really an incentive for different operators to cooperate in base station operation? If so, how can such cooperation be realized? We now explore these issues in greater detail, in the process identifying avenues for future research and development in this area. In particular, we focus on coverage extension, then address other technical issues, and consider the location estimation problem that arises in some cellular deployments.

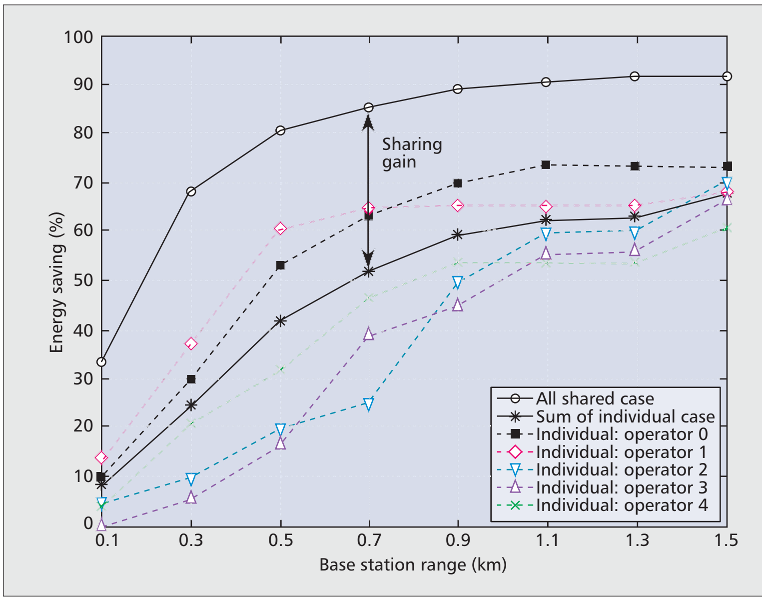


Figure 4. *Energy saving during off-peak times vs. the base station range in the Manchester area.*

Maintaining Coverage using Power Control

The primary constraint of green cellular operation is to preserve coverage and service quality when certain cellular base stations are turned off in the case of low load. Techniques for preserving coverage and maintaining quality of service have much similarity. To be concise, we focus the following discussion on maintaining coverage.

Power control will play an important role in balancing coverage and interference. Power control has been extensively used in cellular networks for interference management and quality of service provisioning. In the case of low traffic load, the main challenge is to increase coverage, while interference management is less critical. Therefore, one can potentially increase transmission power when some base stations are turned off to increase the coverage area of the remaining base stations. Although a simple idea, it needs to be carefully evaluated for both uplink and downlink transmissions.

Cognitive Multifrequency Operation

Multi-frequency-band operation can potentially be explored in green operation. Lower frequency bands have better penetration capability and can provide better coverage under the same transmission power constraint. An example of such a band is the vacant 700 MHz TV band in the United States, auctioned in February 2009 and acquired mainly by large cellular operators. Such a lower frequency band could potentially be used by larger cells, overlapping with smaller cells on a higher frequency band. It is worth noting that frequency selection requires more advanced physical layer technology, which is available in current commercial mobile devices to some degree, with more advanced features being developed as part of the cognitive radio paradigm.

Multihop Relay for Coverage Extension

Various techniques have been considered in cellular networks to improve network coverage

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

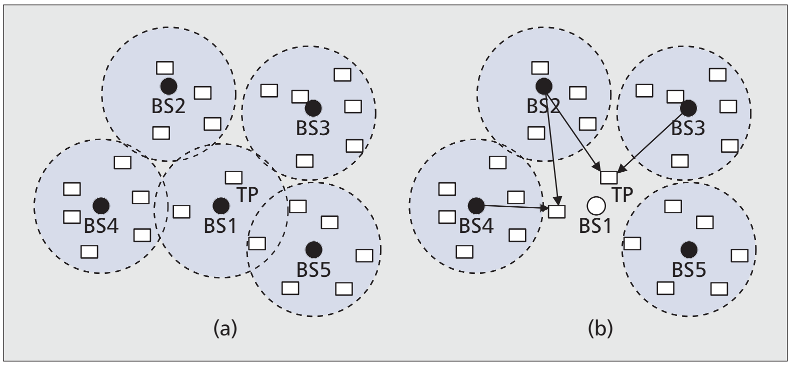


Figure 5. *Coverage extension with coordinated transmissions.*

and/or increase network throughput. Some of these techniques can be applied in the case of green cellular operation with a focus on covering the cells with shut-down base stations. The basic idea is simply to allow other devices to relay the traffic of a cellular user that receives weak or no signal directly from its nearby cellular base stations. This will clearly be useful to ensure that the dynamic shutting down of base stations, while saving energy, does not leave coverage holes. However, finding such relays can be a challenge.

Coordinated Multipoint Transmission and Reception

Another novel approach that can play a role in extending coverage is the use of coordinated multipoint transmission and reception (CoMP) being developed in the context of Long Term Evolution (LTE)-Advanced cellular networks [5]. The basic idea in CoMP, a macro-diversity scheme, is to coordinate transmission and reception at neighboring cells. As shown in Fig. 5, there are many users and five candidate base stations. The dashed lines demonstrate the transmission ranges of the base stations. If intercell cooperation is not allowed, all five base stations should operate to guarantee the coverage requirement (Fig. 5a). However, if the neighboring base stations can transmit cooperatively (e.g., BS2 and BS4, BS2 and BS3) to enlarge their coverage, all three users served by BS3 can be covered, and BS3 can be switched off to save energy (Fig. 5b). Similarly on the uplink, coordinated receptions at multiple base stations can effectively reduce the received power requirement at each individual base station. The increased power on the downlink and lowered power requirement at the uplink together provide coverage with acceptable service quality for mobile users in nearby cells whose base stations have been shut down during low activity periods [6].

Handling Heterogeneous Cell Sizes

In our simplified estimate of energy savings, we considered uniform cell sizes. In third-/fourth-generation (3G/4G) systems, because of the required high data rate, cellular towers will be more dense and have varying coverage, with more heterogeneous cells, such as microcells, picocells, and femtocells. While most existing deployment strategies focus on quality of service during peak periods, energy savings during off-peak periods should also be taken into account. In particular, heterogeneous cellular architecture should be considered in the deployment stage so that umbrella macrocells can provide overall coverage with lower data rate, and smaller microcells/picocells can provide better data rates for congested areas [7]. This can potentially be exploited for energy efficiency, since in the case of light load we can turn off (most of) the smaller cells. On the other hand, considering energy efficiency introduces additional complexity in network planning and deployment, and thus needs to be further studied.

Granularity of Control

The issue in dynamic base station switching is the temporal granularity of the control. While most of the early work in this area has primarily focused on switching off base stations once a day, it remains to be seen whether finer-grained operation (e.g., on an hourly or even more frequent basis) can improve energy efficiency, particularly in settings where traffic patterns are not predictable from day to day. However, this may require more active online monitoring of traffic and increase the complexity of coordination with other cells to ensure coverage.

Another related question is the spatial granularity of coordination. Base stations can be switched off and on in a distributed manner at either a base station controller or mobile switching center level. This granularity can affect different trade-offs between timeliness, complexity of coordination, and efficiency of the base station switching policy [8].

Cooperation between Operators

As shown in our case study earlier, there is room for greater energy savings if different operators can pool their resources. This is particularly helpful in metropolitan areas where each operator sets up a dense deployment. Such a scenario has also previously been examined by Marsan and Meo [9], who consider the setting of a single cell with two operators that cooperate with each other by accepting each others' traffic (as roaming traffic) when they are shut down and show that considerable energy savings can be obtained in such a setting. In implementing such cooperation, physical sharing may not be a substantial concern. In many cases, in urban areas operator base stations tend to be closely situated, or even on the same tower. The main challenge is the complexity in network operation, regarding issues such as cross-operator user authentication and billing, introduced by fine-granularity roaming.

This is also an interesting problem from a game theoretic perspective. Under what conditions would self-interested operators agree to cooperate with others? What kind of profit-sharing agreements will provide an adequate incentive for all participants? Cooperative game theory formulations that consider coalition formation may provide insights on these questions. From a policy perspective, however, it also needs to be examined whether such operator agreements can potentially be abused to create oligopolies that hurt customers, and how these can be ameliorated in the interest of green operation that offers other benefits to society.

Maintaining Emergency Location Service

There are other practical considerations related to the green operation of cellular infrastructure. One such issue in the United States is E911 (Enhanced 911). E911 is a North American telecommunications-based system that automatically associates the physical address with a caller's phone number, which enables the call to be routed to the most appropriate public safety answering point for that address. In cellular networks, this requires location determination of a cellular caller. In E911 Phase II, it is required that 95 percent of a network operator's in-service phones be E911 compliant, with a granularity of 300 m. Therefore, if cellular towers are used to triangulate caller locations, the 1-coverage requirement needs to be extended to 3-coverage (i.e., each caller can reach three cellular towers). However, as shown in a large-scale cellular network study by a major cellular operator in the United States, over 50 percent of the time, a caller receives information from only one base station instead of the three required for triangulation or trilateration [10]. Therefore, to achieve 3-coverage in the original cellular network, much higher cellular tower density is required. To maintain 3-coverage at low traffic volume, we might expect a similar percentage of energy saving. More detailed analysis is needed. Note that there are also alternative and complimentary techniques that address the issue of location determination, including using GPS data, combining with WiFi location information, and using Bayesian inference. Such location techniques apply in the case of green cellular operation as well.

Concluding Comments

In this article we have discussed the current trends in green cellular technology. We have focused on green cellular operation. Using real data traces, we derived a first-order approximation of the percentage of power saving one can expect by turning off base stations during low traffic periods while maintaining coverage. Our coarse-grained analysis shows promising potential. We have also presented and discussed a number of relevant challenges and solutions, such as maintaining coverage, enabling cooperation between operators, and providing E911 service.

In summary, we argue that energy efficiency is an important metric that should be used in the design and analysis of cellular networks, along with the metrics that have been considered traditionally, such as blocking/dropping probability, throughput, and delay. While a few recent papers, such as [2, 6–9], have started to address this topic, there are still many rich open problems in this domain, and we encourage the community to give it greater attention.

We argue that energy efficiency is an important metric that should be used in the design and analysis of cellular networks, along with the metrics that have been considered traditionally, such as blocking/dropping probability, throughput, and delay.

Acknowledgments

We thank Thomas Lee, Yucho Ho, and Sophia Tsang, undergraduate students at the University of California, Davis, for their help in extracting the base station information used in this article. We thank Mr. Raj Savoor from AT&T Labs for a helpful discussion regarding E911 and Mr. Donald Draper Campbell from FCC for helpful input regarding cellular coverage. The work of Dr. Oh was supported by a National Research Foundation of Korea Grant funded by the Korean Government (NRF-2009-352-D00205). The work of Dr. Niu is partially supported by the National Science Fund for Distinguished Young Scholars (60925002).

References

[1] G. Fettweis and E. Zimmermann, "ICT Energy Consumption — Trends and Challenges," *Proc. IEEE WPMC*, Lapland, Finland, Sept. 8–11, 2008.
[2] M. A. Marsan *et al.*, "Optimal Energy Savings in Cellular Access Networks," *Proc. IEEE ICC*, Dresden, Germany, June 18, 2009.
[3] "Sustainable Energy Use in Mobile Communications," White Paper, Ericsson, Aug. 2007.
[4] Ofcom, "Sitefinder: Mobile Phone Base Station Database," http://www.sitefinder.ofcom.org.uk/.
[5] S. Parkvall and D. Astely, "The Evolution of LTE Towards IMT-Advanced," *J. Commun.*, vol. 4, no. 3, Apr. 2009, pp. 146–54.
[6] D. Cao *et al.*, "Energy Saving Performance Comparison of Coordinated Multi-Point Transmission and Wireless Relaying," *Proc. IEEE GLOBECOM*, Miami, FL, Dec. 6–10, 2010.
[7] A. J. Fehske, F. Richter, and G. P. Fettweis, "Energy Efficiency Improvements through Micro Sites in Cellular Mobile Radio Networks," *Proc. IEEE GLOBECOM*, Honolulu, HI, Dec. 4, 2009.
[8] S. Zhou *et al.*, "Green Mobile Access Network with Dynamic Base Station Energy Saving," *Proc. ACM MobiCom* (Poster Paper), Beijing, China, Sept. 20–25, 2009.
[9] M. A. Marsan and M. Meo, "Energy Efficient Management of two Cellular Access Networks," *Proc. ACM SIGMETRICS*, Seattle, WA, June 15, 2009.
[10] Z. Hui, F. Baccelli, and J. Bolot, "Bayesian Inference for Localization in Cellular Networks," *Proc. IEEE INFOCOM*, San Diego, CA, Mar. 15–19, 2010.

Biographies

Eunsung Oh (eunsung.oh@usc.edu) received his B.S., M.S., and Ph.D. degrees in electrical engineering from Yonsei University, Seoul, Korea, in 2003, 2006, and 2009, respectively. He is currently a post-doctoral researcher in the Department of Electrical Engineering at the University of Southern California's Viterbi School of Engineering. His main research interests include the design and analysis of algorithms for green communication networks and next-generation wireless networks.

Bhaskar Krishnamachari (bkrishna@usc.edu) received his B.E. in electrical engineering at The Cooper Union, New York, in 1998, and his M.S. and Ph.D. degrees from Cornell University in 1999 and 2002, respectively. He is currently an associate professor and a Ming Hsieh Faculty Fellow in the Department of Electrical Engineering at the University of Southern California's Viterbi School of Engineering. His primary research interest is in the design and analysis of algorithms and protocols for next-generation wireless networks.

Xin Liu (liu@cs.ucdavis.edu) received her B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, People's Republic of China, in 1994 and 1997, respectively; and her Ph.D. degree from Purdue University in 2002. She is currently an associate professor in the Computer Science Department at the University of California, Davis. Before joining UC Davis, she was a postdoctoral research associate in the Coordinated Science Laboratory at UIUC. Her research interests include resource allocation and scheduling mechanisms in wireless networks, integration of wireless and wireline networks, sensor networks, ultra-wideband communication, and timing channels.

Zhisheng Niu [SM] (niuzhs@tsinghua.edu.cn) graduated from Northern Jiaotong University, Beijing, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. After spending two years at Fujitsu Laboratories Ltd., Kawasaki, Japan, he joined Tsinghua University, Beijing, China, in 1994, where he is now a full professor in the Department of Electronic Engineering. His current research interests include teletraffic theory, mobile Internet, radio resource management of wireless networks, and cognitive and multiple radio networks. Currently, he is a Fellow and Councilor of the IEICE, Director for Conference Publications of IEEE Communications Society, and council member of the Chinese Institute of Electronics (CIE).

Power Consumption in Telecommunication Networks: Overview and Reduction Strategies

Willem Vereecken, Ward Van Heddeghem, Margot Deruyck, Bart Puype, Bart Lannoo, Wout Joseph, Didier Colle, Luc Martens, and Piet Demeester, Ghent University

ABSTRACT

One of the main challenges for the future of information and communication technologies is reduction of the power consumption in telecommunication networks. The key consumers are the home gateways at the customer premises for fixed line access technologies and the base stations for wireless access technologies. However, with increasing bit rates, the share of the core networks could become significant as well. In this article we characterize the power consumption in the different types of networks and discuss strategies to reduce the power consumption.

INTRODUCTION

In the last decade the attention on environment-friendly solutions has drastically increased. Especially due to the debate concerning climate change, every emerging technology is scrupulously evaluated on its carbon footprint. This is also the case for information and communication technologies (ICT). It is estimated that ICT is accountable for 2–4 percent of worldwide carbon emissions. The power consumption during the use phase of the equipment accounts for roughly 40–60 percent of the carbon emissions. By 2020 these emissions are expected to double if no initiatives are taken to reduce this footprint. A significant part of these emissions, about one sixth, is attributed to telecommunication networks [1].

Worldwide, the growth rate of Internet users is about 20 percent per year. In developing countries this growth rate is closer to 40–50 percent. Thus, the share of greenfield deployments in telecommunication networks will be significant. Consequently, emerging technologies need to be evaluated on their environmental impact. Also, ICT is being regarded as a solution with the potential to eliminate about 15 percent of the global carbon footprint [2]. If the sector wishes to realize its ambitions, it will also need to demonstrate that it can reduce its own footprint. Different research efforts analyze the power consumption of telecommunication equipment [3, 4]. Next to the characterization of the power consumption, we also give an overview of the optimization strategies. In [5] the authors suggest the introduction of sleep modes, and [6] suggests component optimization and power management as power saving strategies. However, currently new approaches or variations on the suggested approaches are emerging. We introduce the different network architectures and the design parameters that define their power consumption. Based on these parameters the power consumption is then quantified. We elaborate on approaches for the reduction of power consumption.

The work described in this article was carried out with the support of the BONE project ("Building the Future Optical Network in Europe"), a Network of Excellence funded by the European Community's Seventh Framework; the IBBT-project GreenICT and the STRONGEST project, funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 247674.

NETWORK ARCHITECTURES

Figure 1 gives an overview of the different types of network architectures we consider. We make a distinction between fixed line and wireless access networks and core networks.

ACCESS NETWORK ARCHITECTURE

The purpose of access networks is to provide a connection to users through which they access the Internet. They are usually organized in tree structures. All users are connected to a central office in which the traffic is aggregated and transferred to a core network. This connection is provided through different branches of the tree. Depending on the used technologies, the tree has different aggregation levels at intermediary nodes. In the access networks we distinguish between wireless and fixed line access networks (Fig. 1).

In *fixed line access networks* the user connects through a physical wire. Three main types of technologies are currently used.

First, there is digital subscriber line (DSL), which uses the twisted pair copper cables from the old telephone lines. Several technology flavours exist, varying in bit rate and maximum range. Asymmetric DSL (ADSL) and very high bit rate DSL (VDSL) flavors are best known.

A second technology is coax cable technology on which the Data Over Cable Service Interface

0163-6804/11/$25.00 © 2011 IEEE

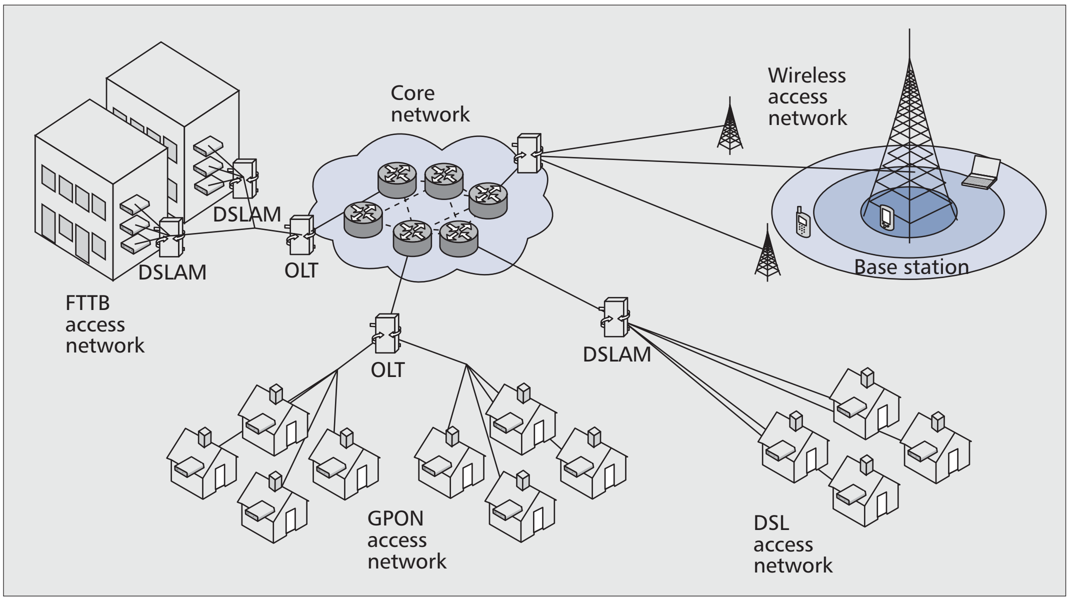


Figure 1. *Network overview.*

In fixed line access technologies the largest power consumer is the home gateway, which can easily be optimized with relatively easy approaches. For wireless access technologies the largest power consumer is the base station.

Specification (DOCSIS) standard is used. These networks are typically built starting from legacy television broadcasting networks.

Presently, optical technologies are emerging. These technologies are already used deeper in the network where higher bit rates are required. At present, optical technologies are starting to appear at the user edge of the network. They are built with either a dedicated connection to the user (point-to-point), an intermediary active splitter (active star), or intermediary passive splitters (passive optical network, PON).

Depending on the offered bit rate and traffic aggregation, the technologies can be used together in an access network. When aggregating the bit rates of VDSL, only optical technologies are able to handle the aggregated traffic load. Therefore, cable access networks are constructed with an optical backhaul and denoted hybrid fiber coax (HFC). Also, optical access networks can be terminated with a VDSL node with limited range and are called fiber to the building (FTTB), fiber to the cabinet (FTTC), or, more generally, fiber to the x (FTTx). Note, however, that FTTx also includes fiber to the home (FTTH), which denotes full optical access networks.

In this analysis we focus on the optical and DSL technologies.

Traffic in the access network is bursty and highly variable. The equipment used in the access network, on the other hand, has a power consumption that is largely constant in time and thus load-independent. Therefore, when evaluating power consumption in the access networks, we consider the power consumption per subscriber as a metric.

ADSL used to be the main access network technology, providing downstream speeds from 8 Mb/s (ADSL) to 24 Mb/s (ADSL2+) and upstream speeds of 1 Mb/s. The maximum range is between 1.5 and 5.5 km. The largest range corresponds with the lowest bit rate capacity. This range allows for large numbers of users to be aggregated in the first node of the access network. Because of this large user aggregation, the power consumption per subscriber of the devices in the backhaul of the access network will be negligible.

VDSL uses an extended frequency spectrum compared to ADSL, resulting in higher bit rates but lower ranges. Consequently, the first aggregation is closer to the user. This also means that a larger backhaul network is necessary, and the power consumption of that backhaul network is more significant.

Optical fiber technologies allow for both higher bit rates and ranges. The bit rate can go up to 10 Gb/s for a single optical fiber with a maximum range between 10 and 20 km. Currently, these bit rates are too high for a single subscriber. Therefore, point-to-point connections are mainly used in the backhaul network to aggregate large amounts of traffic. In active star and PON architectures the bit rate capacity is distributed over large numbers of users. For PONs split ratios of 32 (range = 20 km) and 64 (range = 10 km) are common. The most frequently used standard is GPON (Gigabit PON). Current implementations do not provide large numbers of fiber connections on the optical line terminal (OLT). Usually, between 4 and 72 fibers can be connected.

In *wireless access networks* the user connection is provided through a wireless link. The user's devices use radio signals to connect to a base station, which is then further connected to the central office through a backhaul network. Different technologies are available varying in transmission power, transmission frequency, modulation scheme and multiplexing technique and thus providing different access bit rates to the users.

The three main emerging wireless technologies are mobile Worldwide Interoperability for Microwave Access (WiMAX), high-speed packet access (HSPA), and Long Term Evolution (LTE).

Mobile WiMAX is based on the IEEE 802.16 standard. It operates in the 2–6 GHz band, and

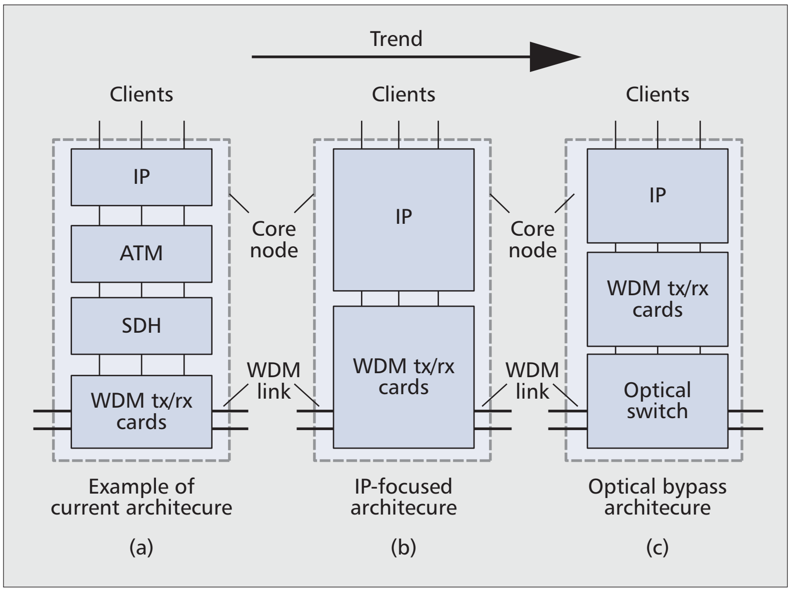


Figure 2. *Core network node architecture.*

is developed for mobile wireless applications and allows people to communicate while they are moving. The highest supported bit rate is approximately 70 Mb/s.

HSPA is the successor of the widely deployed universal mobile telecommunications system (UMTS, also known as third generation [3G]). HSPA provides increased performance by using improved modulation schemes and refining the protocols by which handsets and base stations communicate. The end-user experience is further improved by increasing the peak data rates up to 14 Mb/s in the downlink. HSPA uses the 2.1 GHz band.

LTE is the newest wireless broadband technology. LTE is marketed as the fourth generation (4G) of radio technologies. Targets for the bit rate are to have peak data rates from 10 Mb/s up to 300 Mb/s in the downlink. However, in practical implementations 300 Mb/s rates have not yet been achieved. LTE uses the 2.6 GHz band. However, in the future LTE may use the 800 MHz band.

Contrary to fixed line access networks, the determining factor in wireless access network design is the area covered by the base stations. The covered area is related to the input power of the base station antenna and the bit rate [3]. The input power determines the reach of the signal and thus the area covered by a base station. Current technologies allow different modulation schemes to be used, determined by the signal-to-noise ratio. Therefore, higher bit rates will be achieved at smaller ranges from the base station.

Customer Premises Equipment

At the customer premises the access networks connect to *customer premises equipment*. For fixed line networks this is usually a home gateway, which then further connects to other equipment such as a notebook or set-top box. For wireless access networks this equipment is more diverse. It can be a mobile phone, a wireless network interface card in a computer, or a home gateway. For wireless technologies we therefore use the more generic term *mobile station.*

Core Network Architecture

Access networks aggregate the users in a certain area. In order to interconnect these areas *core networks* are used. A core network consists of a number of core nodes that are interconnected through wavelength-division multiplexed (WDM) optical fiber links, usually in a mesh or ring topology.

Current core networks are typically a mix of several layers of technologies on top of each other, such as IP-over-ATM-over-SDH (Internet Protocol, asynchronous transfer mode, synchronous digital hierarchy), as illustrated in Fig. 2a. However, there is a trend to move to more homogenous architectures where IP is routed directly over WDM links (Fig. 2b). Given this trend, we focus on the latter architecture only. (Figure 2c will be discussed later).

From a high-level view, core nodes are optical-electrical-optical-based. This means that all optical traffic is converted to the electronic domain and processed by the node, whether the traffic is terminated at this node or not. In general, a node consists of a number of WDM transmit and receive cards, also referred to as transponders or transceivers, which are connected to an IP router. The IP router in turn can be connected to a number of access routers.

WDM fiber links carry a number of wavelengths, each typically having a capacity of 10 or 40 Gb/s. Forty to 80 wavelengths/fiber are common. Optical amplifiers are necessary at intervals of about 80 km to make up for signal attenuation. With the finite range of light paths (depending on the line rate and technology employed, this is on the order of 1000 to 4000 km), long links require regeneration of the optical signal.

Quantification of Power Consumption

In this section we quantify the power consumption of the architectures described above. The quantifications in this section are based upon equipment data sheets, own measurements and external literature sources.

Fixed Line Access Networks

In fixed line access networks each subscriber (Subs) has a dedicated connection. Thus, the power per subscriber is a stable metric. In DSL technologies the last node before the subscriber is the DSL access multiplexer (DSLAM). ADSL equipment consumes 1–2 W/Subs; VDSL equipment roughly consumes 3–5 W/Subs. VDSL equipment power consumption is slightly higher, although trends indicate this technology is being optimized.

Optical network equipment currently consumes 10–20 W/port. However, using GPON technology, this can be further distributed. The OLT consumes 0.2–0.8 W/Subs. Due to the small range of VDSL technology, it is possible that

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Technology	Range (km)	Bit rate (Mb/s)	Users/node	Minimal user density (subs/km^2)	Power/subs (with PUE) (W/subs)
ADSL ADSL2+	5.5 1.5	8[1] 24[1]	384–768	4–8 50–100	2–4
VDSL VDSL2+	1.0 0.3 0.3	26[1] 55[1] 100	16–192	5–60 50–700 50–700	6–10
GPON (32) GPON (64)	20 10	2488/32 2488/64	(4–72) * 32 (4–72) * 64	0.1–2 0.8–14	0.4–1.6
Mobile WiMAX HSPA LTE	0.340 (3 Mb/s) 0.240 (3 Mb/s) 0.470 (3 Mb/s)	1–70 1–14 1–300	272[2] 225[2] 180[2]	N/A N/A N/A	27[3] 68[3] 18[3]

[1] Downstream [2] Simultaneous Active Users [3] Modelled for 300 subscribers per km^2

Table 1. *Properties of different access network technologies.*

VDSL equipment with a small number of connections are used. For example, for a VDSL DSLAM with 16 connections, the power consumption of the optical backhaul is 0.01–0.05 W/Subs.

When evaluating the power consumption of these devices, one needs to consider that the premises where they are located often need to be cooled. Also, measures are taken in order to prevent power failure. This overhead is expressed in power usage effectiveness (PUE), which denotes the factor by which the equipment power consumption is to be multiplied in order to know the total power consumption (i.e., equipment + overhead). The PUE is typically a factor two. This means that in reality the above mentioned numbers need to be doubled to estimate the full power consumption.

The total power consumptions, including the PUE and delivered bit rates, are summarized for the different technologies in Table 1 and Fig. 3.

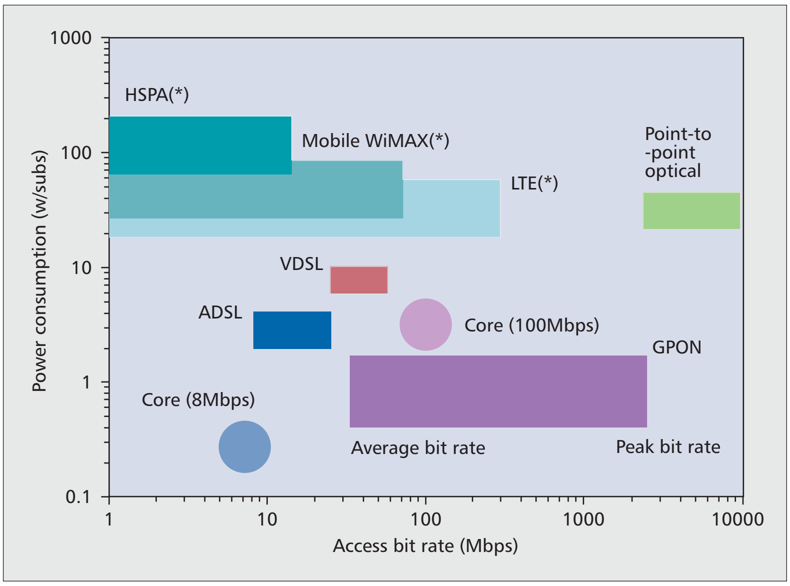


Figure 3. *Power consumption per subscriber of different network technologies.*

Wireless Access Networks

In wireless access networks, the highest power consumer is the base station. The power per subscriber is largely dependent on the subscriber density in the area covered by the base station. Hence, we first evaluate the power consumed per base station and then translate that to the power consumption per user.

A base station is here defined as the equipment needed to communicate with the mobile stations and the backhaul network. For the base stations we assume outdoor placement in a suburban environment at a height of 30 m, covering three sectors, and a mobile station at a height of 1.5 m.

In order to make a fair comparison between the considered technologies, we define a bit rate per active user of approximately 3 Mb/s. We consider the total power consumption of the base station, which includes the PUE overhead.

Mobile WiMAX has the lowest power consumption of approximately 2.9 kW/base station, and a range of 340 m. LTE has the highest power consumption, 3.7 kW/base station, and the largest range, of approximately 470 m. HSPA has the lowest range, 240 m, of all the considered technologies and a power consumption of 3.7 kW/base station, which is comparable to the power consumption of LTE.

In urban and suburban areas it is fair to consider subscriber densities between 100 and 300 users/km^2. When we assume a density of 300 users/km^2 and compare the power consumption per user, we see that LTE performs the best with a power consumption of 18 W/Subs, followed by Mobile WiMAX with a power consumption of 27 W/Subs. The power consumption per user is lower for LTE because of its larger range. HSPA has the highest power consumption per user, 68 W/Subs, caused by its lower range. Note that these numbers are related to the considered subscriber density. When

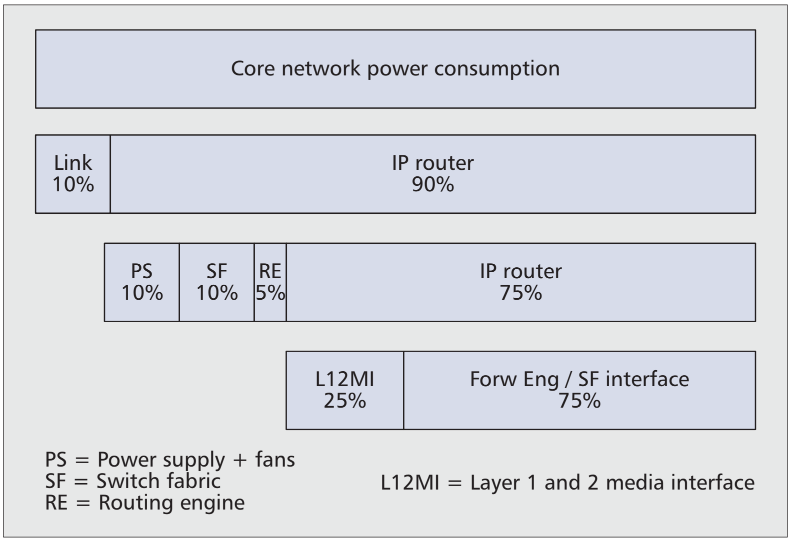


Figure 4. *Generalized core network power consumption distribution.*

we consider half of the subscriber density, the estimated power consumption per subscriber doubles.

For wireless access networks the parameters are also summarized in Table 1 and Fig. 3.

Customer Premises Equipment

Next to the power consumption of the access network, the power consumption of the customer premises equipment is important as well.

At present, for fixed line technologies, the home gateway (e.g., a DSL modem) consumes 5–10 W, which is higher than the power consumption in the access network. Home gateways for optical networks also tend to have higher energy consumption than their DSL counterparts. This is a problem since it can annihilate the potential power reduction benefit of adopting GPON technology. In wireless networks the power consumption of the mobile stations is much lower since these are designed for mobile applications, which require low power consumption for long autonomy times.

Core Networks

As shown in Fig. 4 the major share, about 90 percent, of core network power consumption is concentrated in the nodes. The WDM links, by way of optical amplifiers, make up only around 10 percent or less of the power consumption.

We used datasheets on Juniper T-series core routers to determine the power consumption distribution among the components. The line cards that provide input/output interfaces and apply the packet forwarding logic are the major power consumers. The line cards' layer 1 and 2 media interface provides framing, line speed signaling, and physical connection to a specific network media type such as ATM, synchronous optical network (SONET), SDH, or Ethernet. The purpose of core networks is to transfer traffic streams between different sites. Thus, power consumption in the core network is typically expressed in Watts per transferred bit. Figure 5 plots the maximum power consumption of a number of Juniper routers against their aggregated capacity. As can be seen, high-end routers are more energy-efficient than low-end routers: while consuming more in absolute values, the power required to transfer a bit decreases with increasing router capacity. Current routers consume between 0.1 and 0.01 W/Mb/s. On average, taking into account that near the edge of the core are a higher number of low-end routers, core routers consume about 0.05 W/Mb/s. These values already include a correction for PUE.

In [4] a calculation method is suggested to estimate the power consumption per subscriber of the Internet. Based on this method, we estimate that at ADSL access bit rates (8 Mb/s) the core consumes about 0.24 W/Subs. When accounting for an increase in access bit rate to 100 Mb/s, this will increase to approximately 3 W/Subs. We have indicated these values in Fig. 3. At present, the power consumption in core networks is significantly lower than that in the access network. With increasing bit rates due to the adoption of PON technologies, the absolute power consumption of core networks will increase. However, evaluating the power consumption in Watts per transferred bit, router technologies are expected to become more efficient. This is illustrated, for example, by Cisco's CRS router series, where its most recent member (CRS-3) appears to consume less than half of its six-year-old predecessor (CRS-1). Thus, the figure of 3 W/Subs is an upper bound, which can be lowered by the adoption of power optimized technologies. These technologies are discussed next.

Power Consumption Optimization

Switching Off Components

Presently, telecommunications networks are designed to handle peak loads, and little consideration is given to medium and low load situations. Designing adaptable networks, able to switch off elements when demand is lower, will lead to networks that consume less power.

In core networks this can be achieved with *dynamic topology optimization*. This means from the multiple possible topologies that satisfy the traffic demands, the topologies with lower overall power consumption should be preferred. Dynamic optimization typically exploits the daily or weekly alterations in traffic load, where off-peak volumes are potentially lower than 50 percent of peak volumes. When employing multilayer traffic engineering (MLTE) and changing the MLTE strategy to optimize power consumption, reductions of more than 50 percent during off-peak hours can be achieved [7]. MLTE adapts the topology to optimize power consumption, thereby increasing the number of inactive line cards, which can subsequently be switched off and thus save power.

In fixed line access networks, a similar strategy is possible by using dynamic bandwidth allocation (DBA) in PONs. DBA is currently used as a way to allow users to have increased bit rate while other users on the same medium require lower bit rate. The same strategy could also be used to create dynamically adaptable OLTs, which utilize less ports on which a higher split ratio is applied during periods of low traffic. This allows elements to be switched off in the OLT and leads to reduced power consumption.

In wireless access networks optimization can be achieved by the utilization of *hybrid hierarchical base station deployment*. When using base stations with differentiated cell sizes and wireless network technologies, a basic access network can be created that provides a low bit rate but high coverage to users. In the hierarchical layers above, base stations with smaller cell sizes but higher bit rates can be utilized to provide high-bandwidth connections when these are needed. The advantage is that the higher layers can be put to sleep and only need to be activated with high traffic demand.

For example, LTE-Advanced, the successor of LTE that is under development, will support advanced repeaters. Repeaters are active elements without full base station capabilities. Currently, repeaters are designed as always-on devices. However, in LTE-Advanded the transmission power of these repeaters will be controlled by the network and activated when users are present in the area handled by the repeater.

The WiMAX next-generation standard, 802.16m, includes handover support between femto base stations, which are designed for residential or business environments and may enhance indoor coverage, and macro base stations. They typically have a range on the order of 1 to 10 m.

Finally, it is important to optimize the power consumption of the home gateway. These are individual devices that only need to be active during periods when the user is active. At other times, it can in principle be switched off, although in reality this rarely happens. In legislation concerning standby power consumption standards of 0.5 W are emerging. Implementing this on home gateways will already lead to large optimizations.

Reducing Load

With the idle components in the telecommunication networks switched off, the next step is to reduce the load on the remaining components. This strategy will be especially important in access networks since we already pointed out it is difficult to switch off elements.

Adaptive link rate is a strategy in which different line rates are supported on a link. The lower line rates are assumed to consume less power and thus power can be saved. At the customer premises this can be used to reduce the power consumption of the home gateway. In different access network technologies this strategy is showing potential. However, mainly the higher link rates on the order of 1–10 Gb/s have significantly higher power consumption than lower link rates. Second, the algorithms for adaptive link rate use larger packet buffers. These larger buffers also require hardware that needs to be powered. In core networks it makes less sense to use adaptive link rates since the traffic shows less variation.

In core networks a promising technique to reduce power consumption is *optical bypass*, which is already in use for cost reduction and router capacity offloading (Fig. 2c). Traffic not intended for the intermediate node remains in the optical domain and is not processed by the core router. The light path is switched, using optical add/drop multiplexers (OADMs) or optical cross-connects (OXCs), from an incoming fiber link directly to the appropriate outgoing fiber link. This allows the capacity of the router and the corresponding power consumption to be reduced. Optical bypass is possible at single-wavelength or waveband granularities (requiring less ports in the OXC or OADM since multiple wavelengths are switched simultaniously). Wavelength conversion can be employed to optimize fiber utilization and resolve contention. Depending on light path utilization and network size, the power saving potential of optical bypass is up to 45 percent [8].

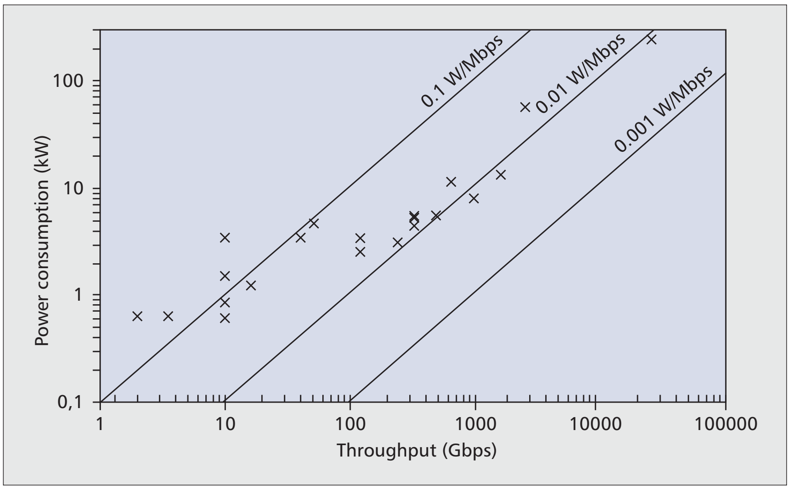


Figure 5. *Power consumption of a set of mid-tier and high-end routers.*

Optimizing Power Consumption of the Remaining Components

When the networks are optimized and the load is minimized, the power consumption of the elements should be reduced.

In next-generation PONs (NG-PONs), OLTs will be designed for higher bit rates (up to 10 Gb/s per port) and higher split ratios (up to 1:128 or even 1:256). Additionally, the range of the signal will be increased, supporting up to 60 to 100 km. In itself, a higher range will lead to higher power consumption per OLT and can require active remote nodes, which add additional power demands. However, in large operator networks there is an ongoing trend of node consolidation, reducing the number of central offices and leading to long-reach access areas. This network consolidation can enable important power consumption reductions.

The energy efficiency of wireless access networks can be improved by increasing the ranges of the base stations. Thus, larger areas can be covered by a single base station, and fewer base stations are necessary. This can be done by the use of multiple transmitting and receiving antennas. This technique is known as multiple-input multiple-output (MIMO). When using, for example, 2 transmitting and 2 receiving antennas (i.e., 2 × 2 MIMO), the range increases by 66 percent, *In wireless access networks many optimizations can still be implemented. At present, in core networks the power consumption is relatively low. Nonetheless, due to the expected increase in traffic volumes, power optimizations are welcomed here as well.*

while the power consumption increases only by 2 to 4 percent, resulting in higher energy efficiency. In the next-generation technologies, LTEAdvanced and WiMAX 802.16m, up to 8 transmitting and 8 receiving antennas can be used.

The technique of optical bypass illustrated an evolution from point-to-point WDM networks to more optical circuit-switched networks. *Optical burst switching* and *optical packet switching* take this technique further and are supposed to provide even finer switching granularity. In optical packet switching, individual packets are switched optically on the correct outgoing fiber. Since optical buffers of an appropriate size are currently infeasible, optical burst switching is proposed as an intermediate technology. A control signal is sent in advance of the packets, allowing the burst-switched router to set up a lightpath for the data, thus eliminating the need for buffering.

While optical packet switching can lead to low power consuming solutions since it eliminates power-hungry opticalel-electrical-optical conversions, it is not yet technically feasible [9]. On the other hand, it is argued that with the line card buffers and switch fabric, the two main candidates for optical implementation, consuming only about 15 percent of the total power consumption of an electronic router, potential energy savings are not as high as commonly expected [10]. A hybrid approach in which optical switches still use electronic buffering seems a more feasible low-power approach for the next decade.

It is not yet clear if the technique of optical burst switching is a viable alternative, the main issue being the relatively low throughput requiring an overbuild.

For continent-sized core networks, increasing the maximum optical path length (i.e., not requiring regeneration of the optical signal) can reduce power consumption. For a pan-European network, savings could be up to 10 percent.

Conclusion

The number of Internet users is fast increasing, and these users demand increasing bit rates. Meanwhile, the carbon footprint of ICT has to be reduced. Saving power in telecommunication networks is becoming an important challenge. Emerging technologies can lead to reduced power consumption, but the design of these technologies needs to be applied with low power consumption in mind. This means switching off components where possible, reducing the loads on the networks, and optimizing the power consumption of the network elements.

Currently, the main share of that power consumption lies near the customer. In fixed line access technologies the largest power consumer is the home gateway, which can easily be optimized with relatively easy approaches. For wireless access technologies the largest power consumer is the base station.

In fixed line access networks power consumption optimization is focused on the technology shift toward full optical networks. In particular, PONs provide low power consumption and are still being optimized. In wireless access networks many optimizations can still be implemented. At present, in core networks the power consumption is relatively low. Nonetheless, due to the expected increase in traffic volumes, power optimizations are welcomed here as well.

References

[1] M. Pickavet *et al.*, "Worldwide Energy Needs for ICT: The Rise of Power-Aware Networking," *2nd Int'l. Symp. ANTS '08*, Dec. 2008, pp. 1–3.

[2] M. Webb *et al.*, "Smart 2020: Enabling the Low Carbon Economy in the Information Age," The Climate Group, tech. rep., 2008.

[3] M. Deruyck *et al.*, "Power Consumption in Wireless Access Networks," *Euro. Wireless Conf.*, Lucca, Italy, Apr. 2010, pp. 924–31.

[4] J. Baliga *et al.*, "Energy Consumption in Optical IP Networks," *J. Lightwave Technology*, vol. 27, no. 13, July 2009, pp. 2391–2403.

[5] M. Gupta and S. Singh, "Greening of the Internet," *SIGCOMM '03: Proc. 2003 Conf. Apps., Technologies, Architectures, and Protocols for Computer Commun.*, New York, NY, USA: ACM, 2003, pp. 19–26.

[6] K. Christensen and B. Nordman, "Reducing the Energy Consumption of Networked Devices," IEEE 802.3 tutorial, 2005.

[7] B. Puype *et al.*, "Power Reduction Techniques in Multilayer Traffic Engineering," *11th Int'l. Conf. Transparent Opt. Net. 2009,* July 2009, pp. 1–4.

[8] W. Van Heddeghem *et al.*, "Energy-Efficiency in Telecommunications Networks: Link-by-Link Versus End-to-End Grooming," *ONDM '10*, Kyoto, Japan, Feb. 2010.

[9] S. Aleksic, "Analysis of Power Consumption in Future High-Capacity Network Nodes," *IEEE/OSA J. Opt. Commun. and Net.*, vol. 1, no. 3, Aug. 2009, pp. 245–58.

[10] R. Tucker *et al.*, "Evolution of WDM Optical IP Networks: A Cost and Energy Perspective," *J. Lightwave Tech.*, vol. 27, no. 3, Feb. 2009, pp. 243–52.

Biographies

Willem Vereecken (Willem.Vereecken@INTEC.ugent.be) received his M.Sc. degree in electrotechnical engineering from Ghent University, Belgium, in July 2005. He is a Ph.D. student with the INTEC Broadband Communication Networks (IBCN) research group in the Department of Information Technology at Ghent University/Interdisciplinary Institute for Broadband Technology (INTEC UGent/IBBT). His research is funded by a Dehousse Ph.D. grant. His main research interests are the environmental impact of ICT systems and the design of energy-efficient network architectures.

Ward Van Heddeghem received his Master's degree in applied engineering electromechanics from Hogeschool Gent, Belgium (1999), and his M.Sc. degree in computer science engineering from Vrije Universiteit Brussel, Belgium (2009). He is a Ph.D. student in the IBCN research group at INTEC UGent/IBBT. His research interests are in the environmental impact of ICT and energy-efficient network architectures.

Margot Deruyck received her M.Sc. degree in computer science engineering from Ghent University in July 2009. Since September 2009 she has been a research assistant in the Wireless and Cable (WiCa) research group at INTEC UGent/IBBT. Her scientific work is focused on green wireless access networks with minimal power consumption and minimal exposure of humans.

Bart Puype received his M.Sc. degree in electrotechnical engineering in 2002 and a Ph.D. degree in 2008 from Ghent University. He is working as a researcher in the IBCN research group at INTEC UGent/IBBT. His main interests are in the field of communication networks, focusing specifically on the design and evaluation of multilayer IP-over-optical networks.

Bart Lannoo received his M.Sc. degree in electrotechnical engineering from Ghent University in July 2002. He received his Ph.D. degree in May 2008. Since August 2002 he has been working with the IBCN research group of the Department of Information Technology at INTEC UGent/IBBT, where he is a postdoctoral researcher in the field of fixed and wireless access networks. He is currently active in the European projects ICT-ALPHA and ICT-OASE.

WOUT JOSEPH [M'05] received his M.Sc. degree in electrical engineering from Ghent University in July 2000. In September 2000 he became a research assistant and postdoctoral researcher in the WiCa research group at INTEC UGent/IBBT. Since October 2009 he has been a professor in the experimental characterization of wireless communication systems. His professional interests are electromagnetic field exposure assessment, propagation for wireless communication systems, antennas, and calibration.

DIDIER COLLE [M'01] received his M.Sc. degree in electrotechnical engineering (option: communications) from Ghent University in 1997. Since then, he has been working as a researcher in the IBCN research group at INTEC UGent/IBBT. His research led to a Ph.D. degree in February 2002. His work focuses on optical transport networks to support the next-generation Internet. He has been involved in several European projects like NOBEL, LASAGNE, and TBONES, and COST actions 266 and 291.

LUC MARTENS [M'92] received his M.Sc. degree in electrical engineering and his Ph.D. degree for his research on electromagnetic and thermal modeling and the development of measurement systems for that application from Ghent University in 1986 and 1990, respectively. Since January 199, he has been responsible for research on experimental characterization of the physical layer of telecommunication systems at INTEC UGent. Since April 1993 he has been a professor of electrical applications of electromagnetism.

PIET DEMEESTER [M'89, SM'98, F'09] received an M.Sc. degree in electrical engineering and a Ph.D. degree from Ghent University in 1984 and 1988, respectively. In 1992 he started a new research activity on broadband communication networks resulting in the INTEC IBCN research group. In 1993 he became a professor at Ghent University, where he is responsible for research and education on communication networks. His research focuses on several advanced research topics: network modeling, design, and evaluation; mobile and wireless networking; high-performance multimedia processing; autonomic computing and networking; service engineering; content and search management; and data analysis and machine learning. This research resulted in about 50 Ph.D.s, 1250 publications in international journals and conference proceedings, 30 international awards, and 4 spin-off companies.

Energy Consumption in Wired and Wireless Access Networks

Jayant Baliga, University of Melbourne and National ICT Australia
Robert Ayre, Kerry Hinton, and Rodney S. Tucker, University of Melbourne

ABSTRACT

Energy consumption is becoming an increasingly important issue throughout the community. For network operators in particular it is a concern as networks expand to deliver increasing traffic levels to increasing numbers of customers. The majority of the energy used by the Internet today is consumed in the access network, and this will continue to be the case for the short-to-mid-term future. Access technologies should thus be a prime focus for energy use mitigation. In this article, we present a detailed analysis of energy consumption in current and future access networks. We present the energy consumption of DSL, HFC networks, passive optical networks, fiber to the node, point-to-point optical systems, UMTS (W-CDMA), and WiMAX. Optical access networks are the most energy efficient of the available access technologies.

INTRODUCTION

The Internet has revolutionized the way in which we seek and disseminate information, transact business, educate, and entertain. Traffic growth on the consumer Internet has been high and sustained, with growing numbers of Internet customers using increasingly sophisticated applications, and using them more often. The rollout of broadband access networks has both facilitated and been driven by these increasing demands. Service providers and network operators have invested heavily in deploying and upgrading these new access networks, investing as well in large data centers and expanding core network capacity. In general, these investment decisions have been driven by the traditional design metrics of capital cost, operational cost, and capacity requirements. Energy usage has always been considered, but in the context of operational cost rather than as an issue in its own right. In today's world, the traditional network design metrics alone are no longer sustainable, and energy needs to become one of the principal design parameters for future networks and equipment.

It has been estimated that the IT industry today is responsible for a total of 2 percent of the electrical energy consumed in a typical Organization for Economic Cooperation and Development (OECD) country [1]. Within this total, the energy used in the switching, transmission, and access networks delivering the consumer Internet today has been estimated to be approximately 0.5 percent of typical national consumption, with a rising trend as customer traffic levels increase [2, 3]. Moreover, in the short- to medium-term future, the majority of the total network energy will be consumed in the access network.

This article reviews the range of access network technologies that might be used as network operators move to deliver higher-speed customer access, with a special focus on energy usage as average customer data rates increase [4]. Wise technology choices for future access networks will be an important first step in helping our industry to meet its challenges in a more energy-constrained future [5]. We focus here on the energy consumption of digital subscriber line (DSL), hybrid fiber coaxial (HFC) networks, passive optical networks (PONs), fiber to the node (FTTN), point-to-point optical (PtP) systems, WiMAX, and Universal Mobile Telecommunications System (UMTS) using wideband code-division multiple access (W-CDMA). We find that optical access networks are the most energy efficient of the available access technologies.

POWER CONSUMPTION MODEL

In this section we describe an energy model of the access network, and consider the energy consumption of a number of wired and wireless access technologies. There are several different access technologies in use today, and more are in development [4]. Figure 1 is a schematic diagram of the seven access network technologies we consider here. These technologies include DSL and HFC networks as well as a number of high-speed access technologies: PON, FTTN, PtP, WiMAX, and UMTS. In Fig. 1, thin lines indicate optical links while thick lines indicate copper links.

The energy consumption of each access network can be split into three components: the energy consumption in the customer premises equipment (i.e., the modem), the remote node

This work was supported by the Australian Research Council and by Cisco Systems.

or base station (base transceiver station, BTS), and the terminal unit (which is located in the local exchange/central office). The per-customer power consumption P_a of all seven access technologies in Fig. 1 can be expressed in the form

$$P_a = P_{CPE} + \frac{P_{RN}}{N_{RN}} + \frac{1.5 P_{TU}}{N_{TU}}, \qquad (1)$$

where P_{CPE}, P_{RN}, and P_{TU} are the powers consumed by the customer premises equipment, remote node or base station (if there is one), and terminal unit, respectively. N_{RN} and N_{TU} are the number of customers or subscribers that share a remote node and the number of customers that share a terminal unit, respectively. The last term on the right side of Eq. 1 includes a factor of 1.5 to account for additional overheads such as external power supplies, electricity distribution losses, and cooling requirements in the building that houses the terminal equipment [6, 7]. The equipment in the remote node and at the customer premises is cooled naturally by the surrounding environment.

In this article, we estimate the energy consumption of a range of access technologies, based on representative data from manufacturers' data sheets for commercial equipment. Table 1 lists commercial equipment for each of the seven access networks. The equipment listed in Table 1 is not necessarily best in class for energy efficiency, but we believe it is representative of 2010-era access network equipment.

Table 2 lists representative values of the parameters in Eq. 1 for each of the access technologies considered here. The number of users per remote node and terminal unit for the two wireless technologies (WiMAX and UMTS) correspond to per-user capacities of 0.25 Mb/s. The number of users per remote node and terminal unit for the wired technologies correspond to configurations where the ports on the remote node equipment and terminal unit are fully occupied. In the following paragraphs we describe each access technology and explain the details of the parameters used in developing an energy model for each access technology.

Digital Subscriber Line

DSL is provided through copper pairs originally installed to deliver a fixed-line telephone service [4]. A DSL modem at each customer home connects via a dedicated copper pair to a DSL access multiplexer (DSLAM) at the nearest central office (telephone exchange).

For the comparison presented here, we consider a modern ADSL2+ access service. This technology can in theory provide maximum speeds of 24 Mb/s downstream to a customer close to the central office and 1 Mb/s upstream. However, to account for the typical degradation in performance due to line length, line loss, crosstalk, and noise, we assume a maximum access rate of 15 Mb/s. We consider a typical DSLAM capable of serving 1008 customers, having a full-duplex switching capacity of 2 Gb/s, and consuming approximately 1.7 kW. The customer modem is modeled as consuming 5 W.

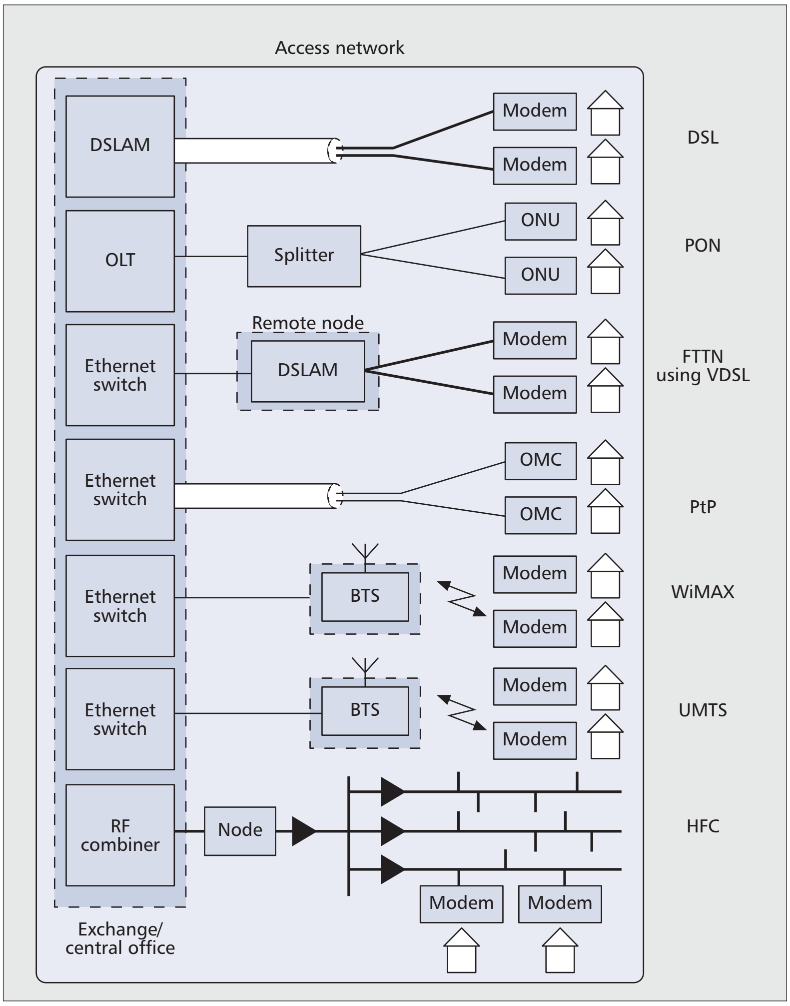


Figure 1. *Schematic of network structure with access network options including digital subscriber line (DSL) and hybrid fiber coaxial (HFC) networks as well as a number of promising candidates for future high speed access technologies - passive optical network (PON), fiber to the node (FTTN), point-to-point optical (PtP), WiMAX and UMTS. Thin lines indicate optical links and thick lines indicate copper links.*

Hybrid Fiber Coaxial Network

Cable distribution networks were initially deployed to deliver television services, and today also deliver Internet and telephony services. Typically, the television program material is compiled from national and regional sources at a headend distribution center in each regional city. This material is distributed on radio frequency (RF) modulated optical carriers through optical fiber to local nodes, where the optical signal is converted into an electrical signal. That electrical signal is then distributed to customers through a tree network of coaxial cables, with electrical amplifiers placed as necessary in the network to maintain signal quality. Hence, these networks are commonly termed hybrid fiber coaxial networks.

The electrical signal sent toward the customer on the coaxial cable includes an array of modulated RF carriers representing the individual television channels, generally extending from either 50 or 65 MHz up to a frequency of 500–900 MHz, depending on the network. A reverse channel to the node or head-end is also provided in the band below 50 MHz. Broadband

Based on the neighborhood topology, the cable is branched, the signal re-amplified, and signals for individual customers tapped off along the route. We allow for four RF carriers to be assigned to convey data signals on each of these coaxial cable links, so that each cable tree supports a total of 152 Mb/s.

	Terminal unit	Remote node	Customer premises equipment
ADSL	Alcatel Stinger FS+	N/A	D-Link DSL502
HFC	Motorola GX2	Motorola SG4000 Quad Node Motorola BLE100 RF Amplifier	Motorola SB6120
PON	Hitachi 1220	N/A	Wave7 ONT-G1000i
FTTN	Hitachi 1220	NEC AM3160	NEC VF200F6
PtP	Cisco 4503	N/A	TC Communications TC3300
WiMAX	Cisco 4503	Motorola WAP 450 Series	Alvarion BreezeMAX USB 200 Zyxel MAX-200M1
UMTS	Cisco 4503	Motorola Horizon 3G-nx	Sierra Wireless AirCard USB 306

Table 1. *Representative equipment used in model of access networks.*

	P_{TU} (kW)	N_{TU}	P_{RN} (W)	N_{RN}	P_{CPE} (W)	Technology limit	Per-user capacity
ADSL	1.7	1008	N/A	N/A	5	15 Mb/s	2 Mb/s
HFC	0.62	480	571	120	6.5	100 Mb/s	0.3 Mb/s
PON	1.34	1024	0	32	5	2.4 Gb/s	16 Mb/s
FTTN	0.47	1792	47	16	10	50 Mb/s	2 Mb/s
PtP	0.47	110	N/A	N/A	4	1 Gb/s	55 Mb/s
WiMAX	0.47	24400	1330	420	5	22 Mb/s	0.25 Mb/s
UMTS	0.47	15300	1500	264	2	20 Mb/s	0.25 Mb/s

Table 2. *Values of access network parameters used.*

Internet access is provided by using one or more of the downstream RF channels to deliver high-speed data, and one or more of the low-frequency reverse channels to send data from the customer into the network (upstream).

Each television customer has a set-top box, which demodulates the incoming signal for display on a television receiver. The data/Internet customer has a cable modem connected to his/her computer or network.

The topology of an HFC network is illustrated in more detail in Fig. 2. As before, thin lines indicate optical links while thick lines indicate copper links. In our model for Internet access via a HFC network, we include:

- "Head-end" equipment, where video RF carriers are combined in a broadband network platform (BNP) with data-supporting RF carriers onto transmission fibers
- Field-deployed node equipment, which converts the optical signals into electrical signals suitable for cable distribution
- A network of electrical RF amplifiers and splitters, so that each node can support a number of customers spread over many streets
- In each customer's premises, a cable modem; Universal Broadband Routers (UBRs) are an essential part of the HFC data network, but in this analysis we focus on the energy consumption of the access network and do not include the energy consumption of UBRs in our calculations.

We model the network using current DOCSIS-based equipment, employing 6 MHz RF channels and 256-quadrature amplitude modulation (QAM) to deliver 38 Mb/s per RF channel. Each node in the cable network receives four sets of RF data carriers on separate fibers and the video program carriers on another fiber. These data carriers are combined with video program carriers and distributed on four lines of coaxial cables. Based on the neighborhood topology, the cable is branched, the signal re-amplified, and signals for individual customers tapped off along the route. We allow for four RF carriers to be assigned to convey data signals on each of these coaxial cable links, so each cable tree supports a total of 152 Mb/s.

In the final distribution network link, we allow 15 customers to be served from a single electrical line amplifier. When the offered capacity per customer is low, the coaxial cable distribution network requires few nodes to support many customers and is highly branched; in such cases we allow one trunk amplifier to support up to eight line amplifiers. Each node requires at least one video and one data port on the BNP that combines the RF signals, and a number of RF data channels from the UBR. When modeling high data loads with low oversubscription, several UBRs may be required in a city.

The BNP consumes 620 W, while serving up to four nodes. The number of customers served by a node depends on the number of RF carriers available for data, in both downstream and upstream directions, and the per-customer traffic level. A quad node consumes 256 W; the trunk and line amplifiers each consume 35 W. In Table 2 the power consumption of the HFC remote node includes the power consumption of the node, trunk amplifier, and necessary electrical amplifiers in a typical installation.

The RF amplifiers, nodes, and head-end RF combining equipment in the HFC network are shared between data and broadcast television services; thus, the energy consumption of this equipment should be shared between the services. On the basis of the subscriber numbers for each service in one provider's network, we allocate 40 percent of energy consumption of this equipment to supporting Internet access.

We have dimensioned the network on the basis of downstream capacity delivered to customers. There are, however, many instances where the upstream capacity of the reverse channels may be limited by high ambient RF noise levels, and this limits the number of customers that can be served from a node and cable network tree. Thus, in assuming a network limited by download capacity, we offer a conservative (i.e., lower) power consumption estimate.

Passive Optical Network (PON)

Fiber to the premises installations most commonly use a PON technology, in which a single fiber from the network node feeds one or more clusters of customers through a passive splitter [4]. An optical line terminal (OLT) is located at the central office, and serves a number of access modems or optical network units (ONUs) located at each customer premises. Each customer ONU in a cluster connects via a fiber to the splitter, and from there shares the same fiber connection to the OLT. ONUs communicate with the OLT in a time multiplexed order, with the OLT assigning time slots to each ONU based on its relative demand.

The number of customers that share a connection to an OLT is generally 32 or 64. For the network energy model, we consider a gigabit PON (GPON) access network, providing asymmetric 2.4 Gb/s downstream, 1.2 Gb/s upstream from the ONU to the OLT and 32 customers sharing a connection to an OLT. The OLT equipment shelf is capable of supporting 32 GPON lines (1024 customers), has a backhaul capacity of 16 Gb/s, and draws 1.34 kW. The splitter is unpowered. The ONU is a basic model providing only data connectivity, and draws 5 W.

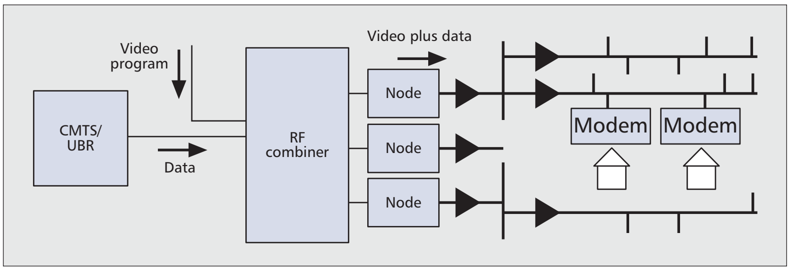


Figure 2. *Layout of an HFC network.*

Fiber to the Node Using VDSL

Fiber-to-the-node (FTTN) technology makes use of existing copper pairs [4]. Dedicated fiber is provided from a network switch to a DSLAM in a street cabinet close to a cluster of customers, and high-speed copper pair cable technologies such as very-high-speed DSL (VDSL) or ADSL2+ are used for the final feed to the customer premises. This accommodates the distance limitations of high-speed pair-cable technologies, and enables high-speed broadband service delivery without the cost of providing new cable entry to the customer premises.

In an FTTN network using VDSL, a remote node houses a VDSL DSLAM which communicates with several homes through the copper wire and connects back to an Ethernet switch in the central office/local exchange through a fiber link. A typical VDSL2 line card supports 16 customers and consumes approximately 42 W. An additional 5 W is consumed for an ONU to link the remote VDSL DLSAM to an OLT back in the central office. The VDSL2 customer modem consumes 10 W and has a peak access rate of 50 Mb/s. The Ethernet switch has 116 optical Gigabit Ethernet ports and 64 Gb/s of switching capacity, with a power consumption of 474 W. For the model, we dimension the fiber backhaul capacity to suit the customer traffic level, but set upper limits on number of customers per Ethernet switch at 1792, and the maximum number of DSLAMs per Ethernet switch at 112. The four remaining ports on the Ethernet switch are used to provide backhaul capacity.

Point-to-Point Access Optical Network

The highest access speed is achieved using a dedicated fiber between each customer premises and the network terminal unit in a PtP configuration [4]. The customer premises employs an optical media converter (OMC) to convert between the electrical signal used inside the home and the optical signal used in the access network.

For typical central office equipment we consider an Ethernet switch providing 116 optical Gigabit Ethernet ports and 64 Gb/s of switching capacity, with a power consumption of 474 W. An OMC at each home converts between the electrical signal used in the home network to an optical signal for transmission over fiber, and consumes 4 W. In this architecture there is no remote node. Each Ethernet switch connects to 110 homes, with the six remaining ports used to provide backhaul capacity.

Although the oversubscription rate applied by network providers is typically much higher for wireless access networks than for wired access networks, to facilitate a fair comparison we model the same across all access networks.

WiMAX

WiMAX is a high-speed wireless access technology. WiMAX was initially designed to provide fixed-point or nomadic wireless access services, but its design standards have since been amended to support full mobility. In our model, we focus on the use of WiMAX in a stationary setting, where each home uses an indoor modem to connect to a base station. The WiMAX base station is remotely located and uses fiber or point-to-point wireless backhauls to connect to the metropolitan and edge network. The area covered by a base station is referred to as a cell, and users in a cell share the total available bandwidth. Per-user bandwidth can be significantly increased by creating multiple sectors in a cell through the use of directional antennas. In a three-sector configuration, each antenna covers a 120° sector.

WiMAX provides access rates of up to 70 Mb/s under ideal conditions. However, typically in urban areas there is not a clear line of sight between the user and the base station, and the combination of reduced signal level and multipath interference limits access speeds to about 35 Mb/s at distances up to about 7 km, with degraded speeds at higher distances.

For the comparison in this article, we model the base station at the remote antenna site as using a point-to-point fiber link to communicate to an upstream Ethernet switch. For the base station we assume a dual-antenna multiple-input multiple-output (MIMO) system with three sectors and mast-mounted power amplifiers. Each sector is modeled as providing 35 Mb/s in total to all users in the sector, and the total base station consumption is 1330 W. The fixed point indoor customer premises unit may be a standalone modem or a USB key style modem. Standalone modems typically achieve higher throughputs than USB modems but also consume more power. We model the home modem as consuming an average of 5 W to account for the diversity of possible devices in a given coverage area.

At low average per-user traffic levels, all users within a cell coverage area will receive adequate service, subject to propagation conditions. At higher average per-user traffic levels, fewer users could be adequately served by each base station sector, and either more sectors or more base stations would be needed. This leads to a rapid increase in the equipment power consumption at higher traffic levels.

UMTS

UMTS is a cellular mobile system that can provide high-speed broadband access capability. It includes radio access to a base station, and from there connections to the core networks for data and voice. For this model we adopt the more commonly used W-CDMA variant, and focus on the broadband data access component of the network. Users may connect to a base station through their mobile phone, USB modem, or standalone modem. The base station is often located remote from its network access controller, and uses fiber or point-to-point wireless backhaul to connect to the controller. Through the radio network controller a mobile user can connect to other mobile phones, the public switched telephone network, or the Internet. As with WiMAX, capacity can be greatly increased through the use of multiple sectors.

The spectral efficiency of UMTS was greatly increased through the introduction of high-speed downlink packet access (HSDPA), high-speed uplink packet access (HSUPA), and, most recently, evolved high-speed packet access (HSPA+). HSPA+ allows for theoretical downlink speeds of 42 Mb/s and uplink speeds of 11 Mb/s. However, typically interference in urban areas limits downlink speeds to about 30 Mb/s and uplink speeds to about 6 Mb/s. For our energy consumption model, the base station connects via Ethernet to an upstream switch, and from there to the radio network controller. A typical outdoor base station consumes 1.5 kW and supports three sectors. Each sector has an average downlink throughput of 22 Mb/s. The user modem is a USB modem that consumes less than 2 W.

Oversubscription

We characterize the capacity available to each customer by the headline access rate advertised and sold to customers by the Internet service provider (ISP). However, backhaul networks connecting the access network to the metropolitan and edge networks are dimensioned by network operators to provide some lower worst-case minimum transmission rate to every customer, taking advantage of the bursty nature of customer Internet traffic. The ratio of the advertised access rate to this minimum per-user rate is referred to as the oversubscription rate. Although the oversubscription rate applied by network providers is typically much higher for wireless access networks than for wired access networks, to facilitate a fair comparison we model the same across all access networks. Note that as the use of the consumer Internet for streaming real-time services increases, high oversubscription ratios will become unsustainable.

We model each access network in terms of a headline access rate of A Mb/s per customer and an oversubscription rate M. During the busiest period of the day, the minimum capacity available to a customer is A/M, the per-user capacity. Statistical multiplexing typically occurs at the DSLAM in ADSL, at the OLT in PON and FTTN, at the UBR in HFC, at the small Ethernet switch in the case of PtP, and at the base station switch for WiMAX and UMTS.

Market Share and Take-Up Rate

In many markets, customers choose from a range of Internet access options; for example, customers may be able to choose between DSL, HFC and UMTS network providers. A network provider, when building an access network, estimates the percentage of households that will buy the service in the short to medium term, referred to as the take-up rate, but will also install additional capacity to cater for future growth in take-up.

In markets with regulated competitive access to infrastructure such as pair cable, a similar but slightly different parameter is market share. Competing ISPs commonly install DSL equipment in the same area, and customers can purchase services from a number of infrastructure-based ISPs, each of which has slightly overprovisioned to cater for future growth.

We combine these factors into one parameter and refer to it as underutilization. The power consumption of current networking equipment does not typically scale with utilization [8]; therefore, underutilization decreases the energy efficiency of equipment. To accommodate underutilization, we increase the power consumption of all access network equipment, except the customer premises equipment, by 25 percent. This increase in power consumption corresponds to network equipment utilization of 80 percent.

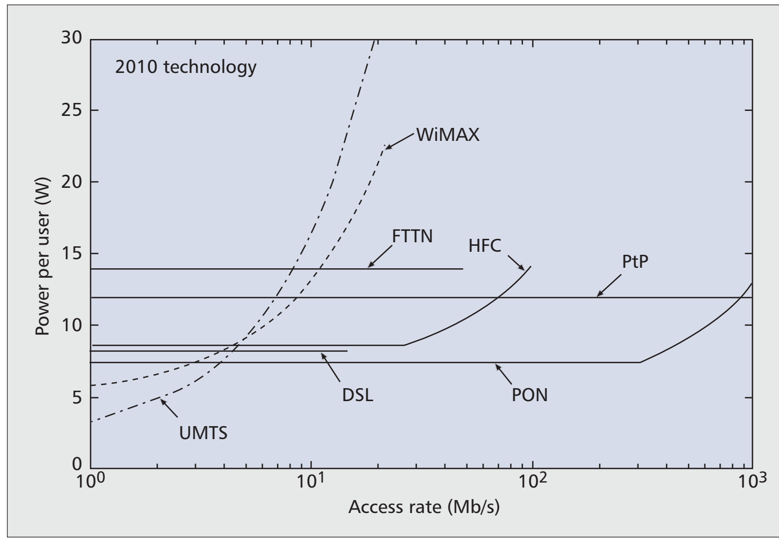


Figure 3. *Power consumption of DSL, HFC, PON, FTTN, PtP, WiMAX, and UMTS as a function of access rate with an oversubscription rate of 20. The technology used is fixed at 2010 vintage for all access rates.*

Energy Consumption

We use the model described earlier to calculate the total per-customer power consumption for typical deployments of each of the seven access networks illustrated in Fig. 1. We also use the model to project the future energy consumption of these access networks. The power consumption of each of the access networks has been calculated for a range of "headline" access service rates, with a constant oversubscription factor of 20. That oversubscription figure is low in situations where customers predominantly use traditional web services such as email and browsing, but could be considered high for future scenarios which include mass use of real-time video on demand services.

For wired access technologies such as ADSL and VDSL, we assume that all customer access ports are fully occupied. For technologies with a shared access resource such as HFC, wireless, and PON, we again assume that all physical ports are utilized, but in addition we share the resource among as many customers as could be served at the particular average service rate and oversubscription factor. As service rates increase, fewer customers can be served, and more equipment (base stations, HFC nodes, PON linecards, etc.) must be provisioned, with an increase in the per-customer power consumption.

Power Consumption Per User

Figure 3 is a plot of the per-customer power consumption of each access technology as a function of the headline access rate. Note here that this access rate is the provisioned per-user capacity multiplied by the oversubscription rate. The technology used in Fig. 3 for all access rates is the 2010-era technology described earlier. From Fig. 3, it is clear that at low access rates (< 1 Mb/s) PON, DSL, and HFC have similar power consumption. At such rates, the overall power consumption is dominated by the consumption of the customer modem. In addition, the power consumption of current networking equipment does not typically scale with utilization [8], resulting in very low efficiencies at low utilization. At an access rate of 1 Mb/s, all five wired access technologies are significantly underutilized. At these low rates, WiMAX and UMTS can flexibly share capacity among a very large number of users, and thus can achieve high efficiency and utilization. Increasing the access rate from 1 to 10 Mb/s increases the power consumption of WiMAX and UMTS by a factor of two and four, respectively, as fewer customers can use a given radio channel or base station and more resources must be provisioned to deliver the service. The power consumption of HFC services increases at a slower rate. At access rates greater than 10 Mb/s, wired access technologies are significantly more energy-efficient than wireless access technologies. HFC, DSL, and FTTN all reach technology limits in the 10–100 Mb/s range. For average service rates of a few to tens of megabits per second envisaged for mass customized streaming services such as high-definition video on demand, the PON has a clear energy advantage.

Comparison of Power Consumption in the Home and the Network

In ADSL, HFC, PON, and FTTN the customer modem or ONU consumes over 65 percent of the total power in the access network. These units would normally operate continuously, but the power consumption of the access network could be significantly reduced through the use of automated sleep modes in customer premises network equipment [9]. Assuming the Internet is used on average 8 h/day, automated sleep modes in customer premises equipment could reduce the energy consumption of the access network by up to 40 percent. Additional savings in power consumption could be realized by fast/micro sleep modes, where customer premises equipment enters a sleep mode during periods of inactivity that are shorter than a second.

	Electronics	Optical interfaces	Power amplifiers	Power conversion
Modem (DSL/HFC)	70%	0%	10%	20%
RF amplifiers	0%	0%	80%	20%
Node (HFC)	0%	20%	60%	20%
BNP	20%	60%	0%	20%
ONU	70%	10%	0%	20%
OMC	50%	30%	0%	20%
Modem (WiMAX/UMTS)	60%	0%	40%	0%
BTS (WiMAX)	69%	0%	11%	20%
BTS (UMTS)	53%	0%	27%	20%

Table 3. *Breakdown of power consumption.*

Improvement in Energy Efficiency with Time

Improvements in complementary metal oxide semiconductor (CMOS) and optical technology should lead to energy efficiency improvements in future generations of network equipment. For example, the energy efficiency improvement rate of Ethernet switches and OLTs is approximately 10 percent per annum [2, 10]. In this section we estimate the overall rate of improvement of each access network technology over time. To estimate this improvement rate we first break down the total power consumption of each item of network equipment into four subsystems: electronic, optical, power amplification, and AC/DC power conversion. We then apply standard estimates of improvement rates (given below) to these subsystems. We calculate the rate of improvement of each item of network equipment as the sum of the improvement rates of its subsystems, weighted by the proportion of total power consumed by that subsystem. This technique of estimating improvement rates is similar to the analysis performed in [10]. Table 3 lists estimates of the breakdown of total power consumption of each item of network equipment into these four subsystems. The per-annum "business as usual" improvement rates for these subsystems are:

- Electronics (26 percent)
- Optical interfaces (5 percent)
- Power conversion (0 percent)
- Power amplifiers (0 percent)

Figure 4 is a plot of the per-customer power consumption for each access technology as a function of time (bottom horizontal axis) and access rate (top horizontal axis). This plot is one scenario for future power consumption of each of the seven access network technologies, using the established "business as usual" efficiency improvement trends outlined in the previous paragraph. For this plot, the access rate is set at 5 Mb/s in 2010 and increases by 42 percent per annum (double every two years), reaching 167 Mb/s in 2020. As before, the oversubscription rate is 20. Although some ISPs today advertise access rates of 100 Mb/s or more, the oversubscription rate used in those networks is typically much greater than 20. The power consumption curves for DSL, HFC, and FTTN cease prior to 2020 because we believe these technologies have a limited ability to scale and meet future increases in bandwidth requirements.

As shown in Fig. 4, we forecast that, if electronics and optics continue to improve at current rates, a lack of improvement in power amplifiers and power conversion subsystems will result in an overall diminishing rate of improvement in all access technologies. The power consumption of HFC and UMTS falls by only 50 percent because the majority of power consumption in these access networks is in power amplifiers, which have limited scope to improve in the future. The results in Fig. 4 suggest that the per-user power consumption of most high-speed access technologies (PON, PtP, FTTN, and WiMAX) should fall by around 70 percent from 2010 to 2020. Wireless technologies will continue to consume at least 10 times more power than wired technologies when providing comparable access rates and traffic volumes. PON will continue to be the most energy-efficient access technology.

Conclusion

We have presented a model of energy consumption of current and future access networks using published specifications of representative commercial equipment. We analyzed the energy consumption of DSL, HFC, PONs, FTTN, point-to-point optical systems, UMTS (W-CDMA), and WiMAX. Passive optical networks and point-to-point optical networks are the most energy-efficient access solutions at high access rates.

Acknowledgement

The authors gratefully acknowledge the valuable discussions with B. Gathercole of the University of Melbourne and thank him for his assistance with assembling the data for the energy consumption of wireless networks.

References

[1] M. Webb *et al.*, "Smart 2020: Enabling the Low Carbon Economy in the Information Age," London, 2008.
[2] J. Baliga *et al.*, "Energy Consumption in Optical IP Networks," *J. Lightwave Tech.*, vol. 27, no. 13, July 2009, pp. 2391–403.
[3] R. S. Tucker *et al.*, "Evolution of WDM Optical IP Networks: A Cost and Energy Perspective," *J. Lightwave Tech.*, vol. 27, no. 3, Feb. 2009, pp. 243–52.
[4] P. Chanclou *et al.*, "Overview of the Optical Broadband Access Evolution: A Joint Article by Operators in the IST Network of Excellence E-Photon/One," *IEEE Commun. Mag.*, vol. 44, Aug. 2006, pp. 29–35.
[5] J. Baliga *et al.*, "Energy Consumption in Access Networks," *Opt. Fiber Commun./National Fiber Optic Engineers Conf. (OFC)/(NFOEC)*, San Diego, CA, Feb. 2008.
[6] U.S. Environmental Protection Agency, EPA Report on Server and Data Center Energy Efficiency, Aug. 2007.
[7] J. Koomey *et al.*, "Network Electricity Use Associated with Wireless Personal Digital Assistants," *ASCE J. Infrastructure Systems*, vol. 10, Aug. 2004, pp. 131–37.
[8] L. Ceuppens, "Planning for Energy Efficiency — Networking in Numbers," *OFC)/NFOEC*, San Diego, CA, Mar. 2009.

[9] M. Gupta, S. Grover, and S. Singh, "A Feasibility Study for Power Management in LAN Switches," *Int'l. Conf. Network Protocols*, Oct. 2004, pp. 361–71.

[10] O. Tamm, C. Hermsmeyer, and A. M. Rush, "Eco-Sustainable System and Network Architectures for Future Transport Networks," *Bell Labs Tech. J.*, vol. 14, no. 4, Feb. 2010, pp. 311–27.

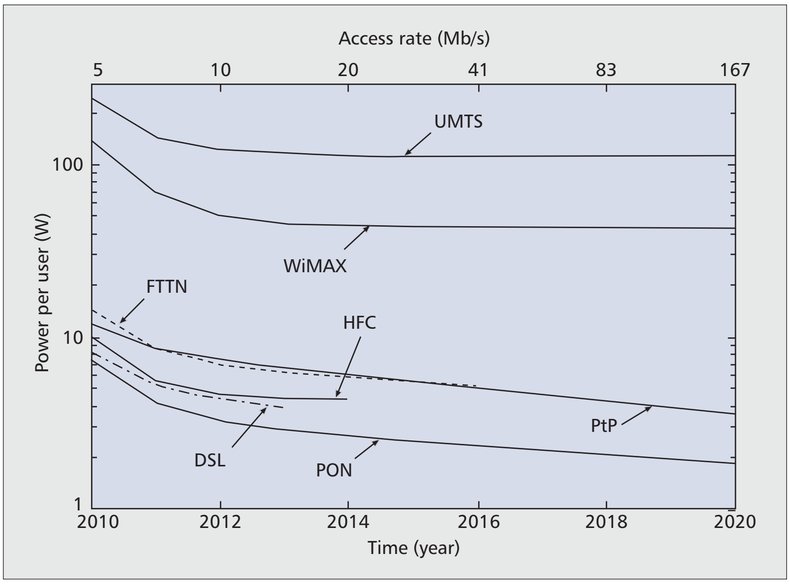


Figure 4. *Expected power consumption of latest generation DSL, HFC, PON, FTTN, PtP, WiMAX and UMTS equipment as a function of the calendar year. The base access rate in 2010 is taken as 5 Mb/s.*

BIOGRAPHIES

JAYANT BALIGA (jbaliga@ee.unimelb.edu.au) received his B.Sc. degree in computer science and B.E. degree in electrical and electronic engineering (with first class honors) in 2007 from the University of Melbourne, Australia. He is currently working toward a Ph.D. degree in electrical engineering at the same university. His research interests include energy consumption, optical network architectures, and wireless communications.

ROBERT W. A. AYRE received his B.Sc. degree in electronic engineering from George Washington University, Washington, DC, in 1967, and B.E. and M.Eng.Sc. degrees from Monash University, Melbourne, Australia, in 1970 and 1972, respectively. In 1972 he joined the Research Laboratories of Telstra Corporation, working in a number of roles primarily in the areas of optical transmission for core and access networks, and in broadband networking. In 2007 he joined the ARC Special Centre for Ultra-Broadband Networks (CUBIN) at the University of Melbourne, continuing work on networking and high-speed optical technologies.

KERRY HINTON received an Honors Bachelor of Engineering in 1978, an Honors Bachelor of Science in 1980, and a Master of Science degree in mathematical sciences in 1982, all from the University of Adelaide. He was awarded a Ph.D. in theoretical physics from the University of Newcastle Upon Tyne, United Kingdom, and a Diploma in industrial relations from the Newcastle Upon Tyne Polytechnic in 1984. In the same year he joined Telstra Research Laboratories (TRL), Victoria, Australia, and worked on analytical and numerical modeling of optical systems and components. His work has focused on optical communications devices and architectures, physical layer issues for automatically switched optical networks (ASONs), and monitoring in all-optical networks. He was also a laser safety expert within Telstra. In 2006 he joined the ARC Special Centre for Ultra-Broadband Information Networks, Australia, at the University of Melbourne, where he is undertaking research into the energy efficiency of the Internet and optical communications technologies.

RODNEY S. TUCKER [S'72, M'75, SM'85, F'90] received his B.E. degree in electrical engineering and Ph.D. degree from the University of Melbourne, Victoria, Australia, in 1969 and 1975, respectively. He is currently a Laureate Professor at the University of Melbourne, where he is director of the Institute for a Broadband-Enabled Society and the Centre for Energy-Efficient Telecommunications. He is a Fellow of the Australian Academy of Science, a Fellow of the Australian Academy of Technological Sciences and Engineering, and a Fellow of the Optical Society of America. In 1975 he was the recipient of a Harkness Fellowship by the Commonwealth Fund, New York. From 1988 to 1990 he was Editor-in-Chief of *IEEE Transactions on Microwave Theory and Techniques*. From 1991 to 1993, he was with the Management Committee of the Australian Telecommunications and Electronics Research Board, and a member of the Australasian Council on Quantum Electronics. From 1995 to 1999 and from 2009 the present, he is a member of the Board of Governors of the IEEE Lasers and Electrooptics Society. In 1995 he was the recipient of the Institution of Engineers, Australia, M. A. Sargent Medal for his contributions to electrical engineering and was named an IEEE Lasers and Electro-optics Society Distinguished Lecturer for the year 1995–1996. In 1997 he was the recipient of the Australia Prize, Australia's premier award for science and technology, for his contributions to telecommunications. From 1997 to 2006 he was an Associate Editor of *IEEE Photonics Technology Letters*. He is currently Vice-President, Publications of the IEEE Photonics Society. In 2007 he was the recipient of the IEEE Lasers and Electro-optics Society Aron Kressel Award for his pioneering contributions to high-speed semiconductor lasers.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

Putting the Cart Before the Horse: Merging Traffic for Energy Conservation

Suresh Singh and Candy Yiu, Portland State University

ABSTRACT

Reducing energy consumption in the Internet has become an increasingly important goal recently. Previous work on reducing energy consumption has primarily looked at either changing link rates or putting interfaces to sleep. Due to the unpredictable nature of traffic, the energy savings achieved have been modest, do not scale, and incur losses and delay. This article proposes a different approach to the problem, which involves aggregating traffic from multiple input links prior to feeding them to the switch interfaces. The main results we obtain are that energy consumption, measured as fraction of interfaces that are awake, scales linearly with load for all loads and the algorithms are fully deterministic yielding zero packet loss.

INTRODUCTION

Concerns over energy consumption in the Internet have prompted significant research activity in the past few years with the goal of achieving *energy proportionality*; that is, energy usage that scales with loading. Previous research work on this topic can be partitioned into two broad categories: work that examines the problem from the network level where energy consumption is part of a route optimization formulation [1, 2], and work that looks at the individual link or switch and uses sleeping or rate adaptation to save energy. The results at the link level demonstrate energy savings for loads less than 30 percent albeit with some packet loss and increased delay. The current article develops a novel solution for *linear scaling* of energy usage with all loads *without any packet loss*. The key idea is *traffic aggregation* via a *merge network* on the input to switches so that we can maximize the number of interfaces put to sleep. This network merges traffic from N input links and feeds them to $K \leq N$ input interfaces on the switch, thus enabling $N - K$ interfaces to be powered off. K scales with load such that packet loss is zero. We evaluate our solution using simulated traffic as well as traffic traces collected at a LAN switch. All cases show linear scaling with zero loss.

The remainder of the article is organized as follows. The next section summarizes related research in energy savings at the switch level. We describe our approach for saving energy, and an algorithm that uses the merge network architecture is presented. The results of our performance study are given, and then we conclude this article.

RELATED WORK

There have been many studies aimed at understanding the energy consumption in various parts of the core Internet and the access networks. Reference [1] provides a study of two Cisco routers, a GSR 12008 and a 7507. In the case of the GSR, the chassis alone consumes 200 W of power. Adding a network processor and switch fabric increases the power draw to 500 W. Adding two four-port Gigabit Ethernet cards, increases the cost to 650 W. Broadly, the energy cost splits as chassis 30 percent, fabric 46 percent, and interfaces 24 percent. In the ideal case, if energy usage of the interfaces and fabric scales with load ρ, the energy consumption of these devices could be written as $200 + \rho \times 450$ W. In this article we show how the energy cost of the interfaces can scale linearly with load. We believe that the switching fabric's cost can also scale linearly with load given that $N - K$ out of N interfaces are off. For instance, [3] provides an algorithm for powering off idle portions of a fabric dynamically.

Approaches for saving energy at the level of routers and switches have focused on exploiting periods of link inactivity to either put interfaces to sleep [4] or use loading models to change the data rate on the interface [5–7]. The results show that for loads up to 30 percent of link capacity, sleeping is possible. However, the energy savings fall rapidly and reach zero beyond these loads. The reasons are twofold: poor traffic prediction pushes us to make very conservative sleep estimates (for fear that packets will be lost at the upstream buffers), and the transition time of the interfaces to go from sleep to wake state has a dramatic impact on sleep times. If the transition time can be made close to zero, we begin to see energy savings even at loads up to 75–80 percent. Similar results hold for the adaptive link rate schemes.

RE-ARCHITECTING THE SWITCH

A fundamental problem we face in putting interfaces to sleep or changing their rates is the unpredictability of traffic on that interface.

[1] *Not all inputs may have packets at the same time, and multiple packets may be destined for the same output but the routing is still deterministic.*

Indeed, as we noted, these mechanisms show no energy savings even for loads as low as 30 percent because of the unpredictability of traffic arrivals. Our approach to this problem is to use traffic aggregation on the input to a switch in order to better use the resources. Table 1 summarizes the loading statistics for eight switch interfaces in our network. We see that the loading of individual interfaces can be very small. Indeed, the aggregate load of all these interfaces combined is 1 Gb/s, which means that, in the ideal case, all the load could be handled by one interface. The traffic pattern of P1 is very bursty with short intense periods of over 250 packets/s followed by virtually no traffic for extended periods. The servers show more steady load, and in some instances (e.g., P5) we can fit a piecewise Poisson model.

Figure 1a shows a typical switch architecture, while Fig. 1b illustrates the idea of merging traffic. In a N-port switch, each interface is connected to one end-host or the interface of some other switch. We propose the idea of merging data streams *before* feeding them to the switch. In Fig. 1b, we place an $N \times K$ merge network before the switch input interfaces. K is the number of switch input interfaces that are powered on. Thus, the traffic from the N input links is aggregated together and then fed into the K interfaces, allowing $N - K$ interfaces to sleep. It is clear that K should be a function of load and will change dynamically.

Interface	Type	Raw	Utilization
P1	Workstation	0.7 Mb/s	0.07%
P2	DMZ	304 Mb/s	30%
P3	DNS	4.96 Mb/s	0.496%
P4	Server	112 Mb/s	11.2%
P5	Server	352 Mb/s	35.2%
P6	Server	96 Mb/s	9.6%
P7	LACP	22.4 Mb/s	2.2%
P8	Server	110 Mb/s	11%

Table 1. *Captured traces (11 am, Monday).*

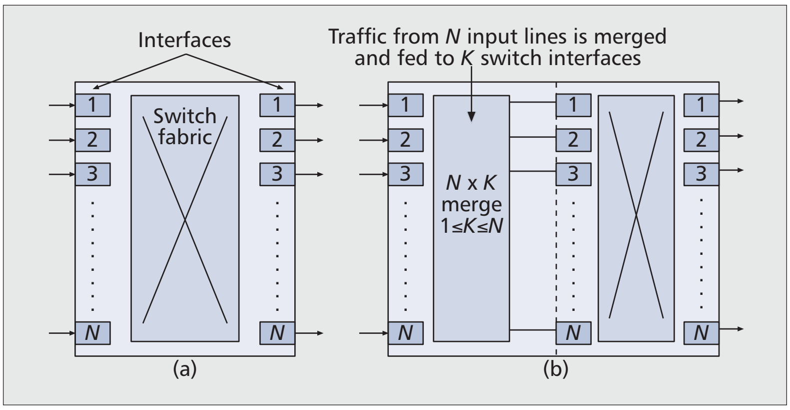


Figure 1. *Modified switch architecture: a) typical; b) with aggregation.*

Architecture of the Merge Network

The function of a merge is to send N input links to some $K \leq N$ interfaces on the switch. The challenge in designing such a merge is the need for flexibility of assignment, and for routing packets between the incoming lines and switch interfaces at low energy cost. We note that in the past a great deal of research has been done in the general area of interconnection networks. However, the model for routing there is quite different from the one we have here. In the typical interconnection network model, N inputs are to be routed to N outputs, but the output for each packet is fully deterministic. This is unlike our model where up to N input links need to be routed to a subset of switch interfaces. The routing is non-deterministic in the sense that if the number of packets arriving at N input links is less than or equal to K, no packet is lost regardless of which input link the packets come on, and the packets can be sent to any of the awake interfaces. Another important difference is that in the merge network, the routing is done entirely in analog; that is, the packets are not received and then forwarded. The reason is to keep the complexity and hence the energy cost down.

Consider, for example, $N = 4$ and $K = 2$ — for ease of explanation we label the input links using letters and the input interfaces to the switch using numbers. So input lines (a, b, c, d) are routed to input switch interfaces (1, 2). In this case, packets arriving on any of a or b or c or d can be routed to *either* 1 or 2, and this choice is dynamic. If one or two packets arrive, none of them is lost; but if three or four packets arrive, only two get sent to switch interfaces with the others lost. The key idea behind traffic aggregation is that the number of interfaces K is changed based on load. Therefore, in this example, the merge network needs to support all of 4×1, 4×2, 4×3, and 4×4 merges.

In order to enable this form of non-deterministic routing in the $N \times K$ mesh, we require a special hardware element we call the *selector* whose functional behavior is shown in Fig. 2a. There are two incoming lines and two outgoing ones. When there is just one packet arrival on either incoming line, the packet is sent out on the solid outgoing line as shown in Fig. 2a. However, if two packets arrive at the two inputs, the earlier one is sent along the solid outgoing line while the later one is sent out along the dotted line. If the links have slotted transmissions and packets arrive simultaneously, we assume some static ordering (say based on incoming line numbers) to determine which packet gets sent out over which outgoing line.

We developed a prototype of a selector for Cat 5e Ethernet wiring using four broadband absorptive single throw switches with appropriate transistor logic. The design of the selector is discussed in [8], but we note that the design we use is far from optimal and was done primarily to demonstrate feasibility.

To illustrate how the selector can help us build a merge network, consider the $4 \times K$ example above. As before, label the interfaces 1, 2, 3, and 4, and the corresponding incoming lines *a*, *b*, *c*, and *d*. Figure 2b shows a merge network that supports all possible merge combinations. The solid line emanating from each selector is the default output, while the dotted line is the deflection route. To configure the network to perform a 4 × 1 merge, we simply shut off interfaces 2–4. Thus, packets arriving at those interfaces are dropped and only one packet, the one that is routed to interface 1, makes it through. Similarly, to implement a 4 × 3 merge, we simply shut off interface 4; thus, three packets make it to interfaces 1–3, while the fourth, if there is a fourth, will be sent to interface 4 and dropped. An important observation about the merge network is that it is *passive* and does not require any logic elements. Therefore, its energy consumption is much lower than the interconnection networks that make up the internal switching fabrics of high-speed switches.

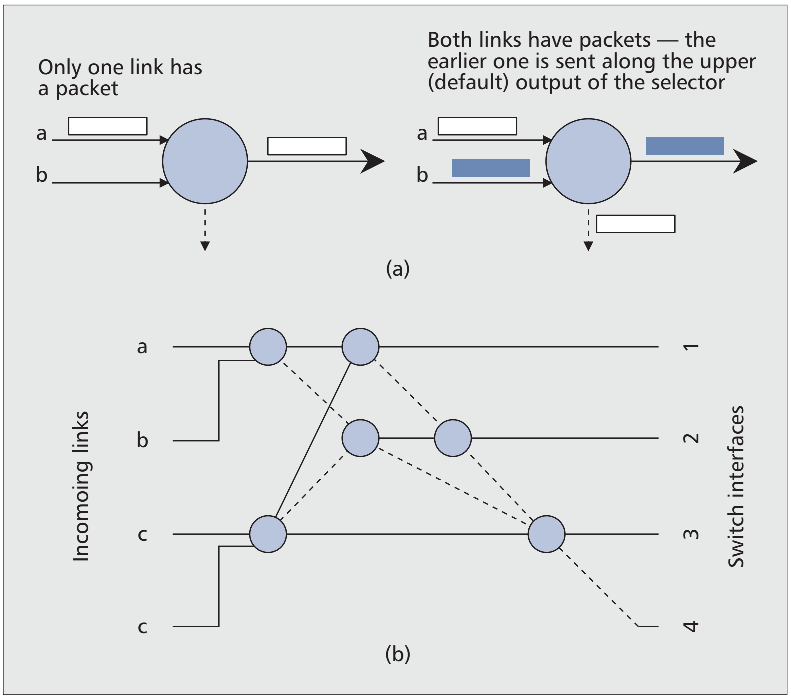


Figure 2. *Example of a merge network: a) logical operation of a selector; b) example of a merge network.*

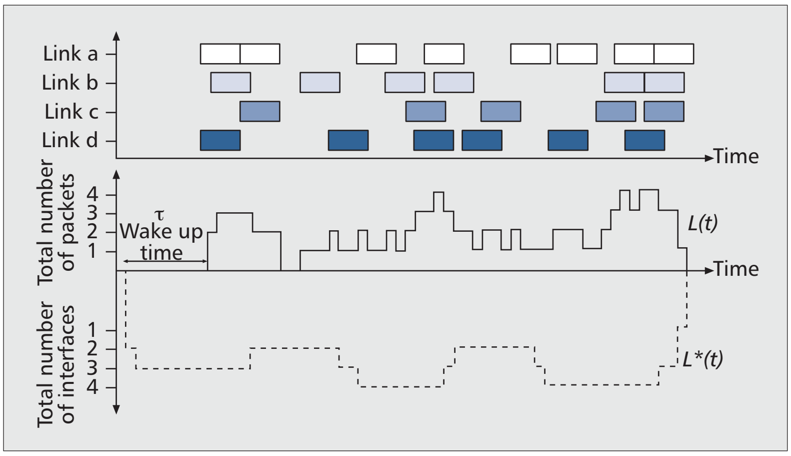


Figure 3. *Optimal aggregation of packets.*

The general structure of an $N \times K$ merge network is a simple generalization of the one shown in Fig. 2b. A log N depth binary tree made up of selectors gives us an $N \times 1$ merge. Next, take all the $N - 1$ deflected outputs (dotted line in Fig. 2a) of all the selectors and form a binary tree with those to get an $(N - 1) \times 1$ merge. This process iterates to create a complete merge network.

An important question that arises is that of the optimality of the merge network as well as measuring its complexity. The complexity of the merge network can be specified by two numbers: the *depth* of the network and the total *number* of selectors used. For the network shown in Fig. 2b, the depth is 4 while the number of selectors used is 6. Indeed, we can prove that the minimum depth of an $N \times N$ merge network is log2 $N + N - 2$, and the number of selectors needed is $N(N - 1)/2$.

Deterministic Traffic Aggregation Algorithm

In this section we begin by defining "optimality" of merging and then describe a zero loss aggregation algorithm that scales energy usage (as measured by the number of on interfaces) linearly with load.

A key parameter that affects the amount of sleeping possible is τ, the time to wake up a sleeping interface. If τ is very small, interfaces can be woken up just in time to process a packet (for the optimal case), thus resulting in perfect linear scaling of energy cost with load. The essential idea behind our definition of optimality is, given perfect traffic prediction, interfaces are woken up τ time before they need to process a packet. The energy cost therefore includes the τ wakeup time. Consider the example in Fig. 3 where we have four incoming links with packet arrivals as shown. At the bottom we show the load $L(t)$ as a function of time (solid line plot). Given a nontrivial value of τ as indicated in the figure, we can plot the number of interfaces that need to be awake or be woken up as a function of time $L^*(t)$ (dotted line plot). Note that initially we need two interfaces to process the two packets coming along links *a* and *d*; therefore, two interfaces need to be woken up τ time previously. This is why the bottom dotted line plot starts with a value of 2 for the number of interfaces. However, a short time later we need to initiate wakeup for a third interface to process the packet arriving along link *b*. After these three packets have been processed, we only need two open interfaces; therefore, the idle interface can be put back to sleep (this operation is assumed instantaneous). We denote the instantaneous load by $L(t)$ (the solid line curve in the figure) and the number of interfaces either awake or in the process of waking up as $L^*(t)$ (the dotted line curve). The amount of energy consumed is the area under $L^*(t)$.

Deterministic Aggregation Algorithm

The key idea we use here is to perform traffic shaping at the upstream interfaces by buffering packets for a short amount of time prior

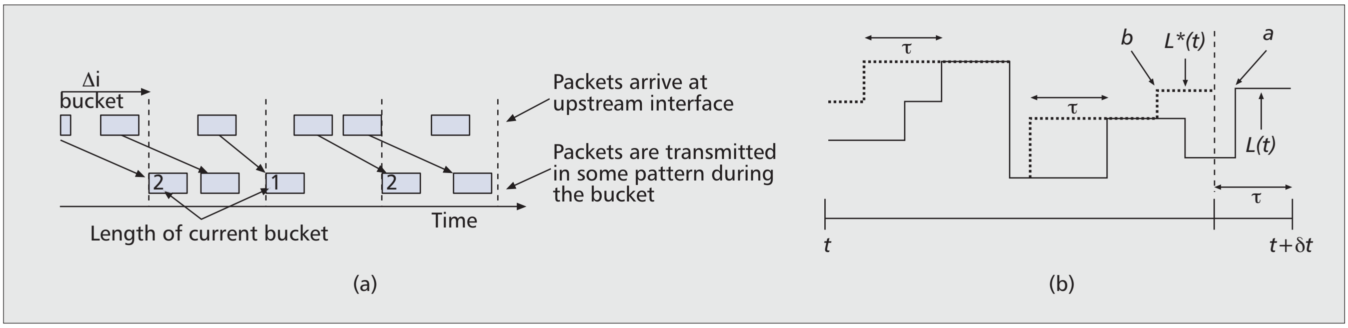


Figure 4. *Traffic aggregation algorithm: a) Δ aggregation of packets at upsteam interfaces for link* i*; b) computing* L* *given* L.

to transmitting them in an appropriately shaped burst. If the downstream merge network knows the expected load for the next burst for each link, it can intelligently open an appropriate number of interfaces and guarantee zero packet loss. We call the buffering interval a *bucket*. The size of the bucket will depend on the buffering capability of the upstream interfaces as well as on the maximum acceptable delay.

Algorithm (Δ *Aggregation*): Let us denote the bucket size of link i by Δ_i and let τ denote the time to wake up an interface. Figure 4a illustrates the bucketization process performed by the upstream interface for link i. As illustrated, packets arriving during the current bucket are buffered and then transmitted in the next bucket.

- Each bucket has one packet right at the start of the bucket, and this packet contains the load information of the bucket (i.e., number of packets) in a special field in the header — this is the only way for the upstream interface to communicate with the downstream interface regarding loading for the current bucket.
- The *pattern* of packets within a bucket is known to the downstream switch. For example, packets can be back to back or spread out within the bucket. This pattern is also communicated to the downstream interface in the header of the first packet within the bucket.

The buffer space required at the upstream interface is Δr bits when the transmission speed is r b/s and the maximum delay experienced by a packet is Δ. Note that this delay may affect the performance of protocols like TCP if it is significant as compared to the round-trip time of a given connection. However, in our study this value was less than 0.5 ms; thus, we believe that the impact on TCP will be minimal.

Let us now consider the algorithm for determining K in the $N \times K$ merge. At any time t, assuming we start t from zero and count up, the downstream switch will know the load along link i until time

$$\left\lfloor \frac{t}{\Delta_i} \right\rfloor \Delta_i + \Delta_i .$$

Given N incoming links, the time until which the downstream nodes know the *total load* is

$$t + \delta_t = \min_{i=1}^{N} \left\lfloor \frac{t}{\Delta_i} \right\rfloor \Delta_i + \Delta_i .$$

As previously, let us denote the total load as a function of time, $L(t)$. We can now state a general *lossless* algorithm for turning interfaces on and off as follows.

Algorithm $K(t)$: At any time t determine load L for interval $[t, t + \delta_t]$. Let $L_{max}(t_1, t_2)$ be defined as the maximum load in the interval $[t_1, t_2]$. In other words,

$$L_{max}(t_1, t_2) = \max_{t'=t_2}^{t_2} \{L(t')\} .$$

Let $L^*(t)$ denote the number of interfaces at time t that are either awake or in the process of waking up. We compute L^* as

$$L^*(t - \tau) = L_{max}(t - \tau, t).$$

The reasoning behind this expression is illustrated in Fig. 4b. First, note that L^* is only defined till time $t + \delta_t - \tau$ because it takes time τ to wake up an interface, and at time $t + \delta_t - \tau$ we only know the load until time $t + \delta_t$. At the time indicated by point a more interfaces need to be awake due to an increase in load. However, these interfaces need to have been woken up τ time previously at point b.

Once we have determined L^*, the algorithm itself is relatively straightforward and can be stated as follows. At any time t:

- Note that like L, L^* is a step function. At each inflexion point of L^*, say at some time t', compute $I = L^*(t' + \in) - L^*(t')$. If $I > 0$, wake up I interfaces; else, put $(-I)$ interfaces to sleep. $\in$ is some very small number and is simply used to indicate a point in time right after the inflexion point of $L^*(t')$.

The algorithm is run at every inflexion point of $L^*(t)$. Furthermore, as new information about future load becomes available, L^* is updated.

Performance Study

In order to characterize the energy savings possible when using a merge network, we use simulated traffic traces as well as actual traffic traces collected at various interfaces of a LAN switch (gigabit interfaces). The traces were fed into a simulator built to simulate the merge network.

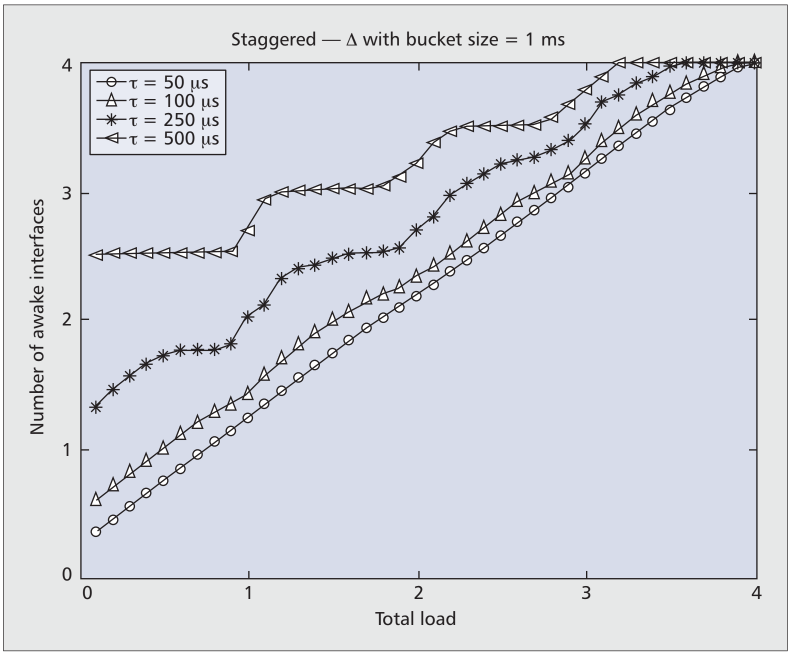


Figure 5. *Performance of a 4 × 4 merge network for varying loads.*

Table 1 summarizes the traffic traces from one of our LAN switches. As is evident, many interfaces on the switch are underutilized. We ran our aggregation algorithm on this data using an 8 × 8 merge network and we monitored the number of interfaces used over time. The bucket size used is 0.5 ms and $\tau = 100$ μs. The main result we obtain is that the maximum number of interfaces used (averaged over time) is 1.2. In other words, by using a merge network, we can power off an average of at least 6.8 interfaces of the switch out of 8 resulting in an energy savings of almost 85 percent.

Since the loading of our own network is so low, we used simulated traffic to examine the behavior of our algorithm over all loads. We assume, 4 incoming links and the aggregate load thus varies from 0 to 4 packets/packet time. Figure 5 plots the energy savings as a function of load over all loads. We ran the simulation for a bucket size of 1 ms and four different values of τ to study the impact of wakeup time on energy savings. As expected, as wakeup time decreases, the amount of sleeping is greater. However, even when the wakeup time is 0.5 ms we still see energy savings for loads up to 3/4 or 75 percent.

Caching Study — We used the LAN traces to understand the impact of merging traffic on cache locality. For our experiments we used a cache size of 2000 entries per interface. We noticed that several traces, such as interfaces P1, P3, P6, and P7, showed no cache misses at all even when using a cache size less than 500 entries. So we concentrated on interfaces P2, P4, P5, and P8, which did show some non-locality in addresses. The results are as follows. When we do not use a merge network at all, the misses are 0.614 percent because the cache needs to be populated. The 4 × 1 merge shows a higher miss rate of 0.77 percent, the 4 × 2 merge has a miss rate of 0.67 percent, and the 4 × 3 merge has a miss rate of 0.89 percent. These miss rates are not significant and demonstrate the feasibility of using the merge in real networks.

Conclusions

This article presents the idea of a *merge network* that enables us to merge traffic from N incoming links prior to feeding them to $K \le N$ switch interfaces. The number of active switch interfaces K can be varied depending on traffic load. This method is shown to linearly scale the number of awake interfaces with load. An important open issue is the changes required to layer 2 protocols to allow using the merge network concept and the effect of merging on TCP due to increased latency. We believe that merge networks can be incorporated directly into current switch designs; and, given the simplicity of the merge network, the added cost will be minimal.

References

[1] J. Chabarek *et al.*, "Power Awareness in Network Design and Routing," *Proc. IEEE INFOCOM '08*, Phoenix, AZ, Apr. 2008.

[2] G. Shen and R. S. Tucker, "Energy-Minimized Design for IP over WDM Networks," *J. Opt. Commun. and Net.*, vol. 1, no. 1, June 2009, pp. 176–86.

[3] L. Peh and V. Sorteiou, "Dynamic Power Management for Power Optimization of Interconnection Networks Using On/Off Links," *11th Symp. High Power Interconnects*, 2003.

[4] M. Gupta and S. Singh, "Energy Conservation with Low Power Modes in Ethernet LAN Environments," *Proc. IEEE INFOCOM (Minisymposium) '07*, Anchorage, AK, May 6-12 2007.

[5] C. Gunaratne *et al.*, "Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR)," *IEEE Trans. Computers*, vol. 57, no. 4, Apr. 2008, pp. 448-61.

[6] S. Nedevschi *et al.*, "Reducing Network Energy Consumption via Sleeping and Rate-Adaptation," *Proc. USENIX NSDI '08*, San Francisco, CA, Apr. 16-18 2008.

[7] G. Ananthanarayanan and R. Katz, "Greening the Switch," *Proc. USENIX HotPower '08*, San Diego, CA, Dec. 7, 2008.

[8] C. Yiu and S. Singh, "Merging Traffic to Save Energy in Enterprise Networks," Portland State Univ., Dept. of Comp. Sci., tech. rep., Oct. 2010.

Biographies

SURESH SINGH [M] (singh@cs.pdx.edu) received his Ph.D. in computer science from the University of Massachusetts at Amherst in 1990 and his B.Tech., also in computer science, from the Indian Institute of Technology in 1984. He is currently a professor in the Department of Computer Science at Portland State University. His research interests include green technologies for Internet devices, energy-efficient computation, wireless networking protocols, information theory, and performance analysis.

CANDY YIU [S] (candy@cs.pdx.edu) is a Ph.D. candidate in the Computer Science Department at Portland State University. She received her M.S. degree in computer science from Portland State University in 2009, an M.Phil. from Hong Kong Baptist University in 2003, and her B.S. from the same university in 2001 in computer science. Her research interests are in wireless communications, protocol design and analysis, low-energy Internet research, and wireless multimedia. She is a student member of the ACM.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

ENERGY EFFICIENCY IN COMMUNICATIONS

On the Design of Green Reconfigurable Router toward Energy Efficient Internet

Chengchen Hu, Xi'an Jiaotong University; SKLNST, BUPT
Chunming Wu and Wei Xiong, Zhejiang University
Binqiang Wang and Jiangxing Wu, National Digital Switching System R&D Center
Ming Jiang, Hangzhou Dianzi University

ABSTRACT

A great deal of research has contributed to the energy efficiency of battery-operated devices in the area of wireless communications, but recently, the energy consumption of the underlying network infrastructure has started to attract more attention as people have become aware of the significant fraction of energy consumed by the Internet of all the energy we consume. We argue that the architecture of routers in constructing the Internet can be improved so as to fully utilize chances to save energy related to transmission equipment. A green reconfigurable router is thus presented for this purpose.

INTRODUCTION

The tremendous success of the Internet makes it a ubiquitous infrastructure nowadays comprising an enormous number of hardware components to deliver a variety of services to end users. In the most recent years, with the aim of building a more sustainable society, research efforts have been made to look into the feasibility and benefits of applying energy-efficient techniques in information and communication technology (ICT) systems.

The concern about energy consumption in today's Internet infrastructure is highlighted in [1]: in 2007, the annual energy consumption of Internet terminal devices and networking and transmission equipment was about 350 billion kWh in the United States (9.4 percent of national electricity consumption) and about 868 billion kWh globally (5.3 percent of global electricity consumption). In addition, the energy demand demonstrates an increasing trend, which calls for novel technologies to be applied in Internet-scale system design to obtain the best trade-off between energy cost and system performance.

In fact, many power management solutions have been available for years and have been well applied to personal computer (PC) systems and different types of battery-operated portable devices, such as dynamic voltage scaling technology in the central processing units (CPUs) of portable devices, energy-saving sleep modes in PC operating systems, and power-aware routing protocols in wireless sensor networks (WSNs). However, the application of these techniques in Internet-scale systems has not been fully considered yet. Some pioneering exploitations of saving energy across the Internet are discussed in [2], which investigated network protocol design with energy-saving considerations and device power consumption optimization techniques against expected performance. The previous studies have confirmed the feasibility and benefits of engineering the next-generation energy-saving Internet infrastructure and have pointed out a set of research directions [2, 3]. In this article, we concentrate on the exploration of power-/energy-saving mechanisms through the design of Internet transmission equipment such as routers. By revisiting the characteristics of Internet behavior and the modular architecture of routers, this article suggests an approach for engineering an energy-efficient Internet from three different perspectives and discusses the imposed technical challenges. To address these challenges and meet the energy-saving requirements, we have proposed a novel conceptual router architectural model, the *green reconfigurable router* (GRecRouter), aiming to contribute to the design and manufacture of energy-efficient routers as the core components of a green Internet.

OPPORTUNITIES

This section presents several intrinsic characteristics of Internet behavior and discusses their implications on opportunities to appl energy-saving techniques.

0163-6804/11/$25.00 © 2011 IEEE

The dynamics of data traffic rate allows us to incorporate energy control mechanisms in the router design (e.g., adapting router processing speed based on the detected traffic rate or queue length). The processors can be decelerated to reduce energy consumption when incoming traffic is at a low data rate.

Low Average Link Utilization and Significant Path Redundancy

Overprovisioning of link bandwidth is common in network planning to harness burst traffic. It is stated that the average utilization of backbone links is less than 30 percent [4]. Besides, massive link redundancies are also built for network failure survival. Although low link utilization and massive link redundancy improve network resilience, they greatly hurt the energy efficiency of the Internet. The links operate at full rate all the time, but they are highly underutilized most of the time. Our first opportunity to save energy is shifting and aggregating traffic from lightly utilized links. In this way, some of the routers (or line cards in a router) can be shut down to reduce power instead of operating all the time.

Variable Internet Traffic Demand

It is now well known that the Internet traffic demand exhibits fluctuations over time due to end-user behavior, temporary link failures, anomalies, and so forth [5]. The dynamics of data traffic rate allows us to incorporate energy control mechanisms in the router design (e.g., adapting router processing speed based on the detected traffic rate or queue length). The processors can be decelerated to reduce energy consumption when incoming traffic is at a low data rate.

Challenges

Although the aforementioned opportunities are promising to minimize the energy consumption in the current Internet infrastructure, they also impose the following questions.

How Do We Properly Aggregate Traffic and Route Packets?

With traffic aggregation and shift in mind, the end-to-end routing paths of individual flows need to be determined, which do not necessarily follow the conventional shortest path first routing paradigm, the Open Shortest Path First (OSPF) protocol. Mathematically, it can be formulated as an optimization problem that maximizes energy savings by identifying as many idle links as possible to be put into sleep mode while guaranteeing the expected network performance. The nature of Internet traffic dynamics over time (e.g., variable traffic demand) imposes additional complexity on finding the solution, which needs to be taken into account.

How Do We Manage Routers under Energy Saving States to Achieve Performance-Energy Trade-off?

Generally, off-the-shelf routers can only work in one of two operational states, on and off, which cannot be flexibly configured or switched to cope with traffic fluctuations. To meet the needs of energy consumption reduction, we suggest that multiple router operational states (e.g., energy saving state) with fast switching ability among them should be supported. Advanced strategies determine when to switch and which state should be adopted to achieve the best trade-off between performance and energy efficiency.

Green Reconfigurable Router

The concept of a green reconfigurable router (i.e., GRecRouter) presented in this article aims to contribute to the creation of the next-generation energy-efficient Internet infrastructure. Through enhancement of the router architectural design, it is expected to reduce both average and peak power consumption during network operation.

The GRecRouter is designed in such a way that its settings (e.g., routing path, clock frequency, supply voltage) are "reconfigurable" based on its awareness of traffic rate fluctuations. In detail, this can be interpreted from three aspects as follows.

- Over a large timescale, it has been known for a long time that Internet traffic exhibits strong daily and weekly patterns, and the behavior remains unchanged over years [5]. Take the enterprise network as an example: the traffic volume in daytime can be approximately 10 times more than that at night , and a similar difference is also observed between workdays and weekends. The time-of-day/time-of-week effect makes the link load vary slowly, but at large magnitude. Keeping this in mind, the GRecRouter first manages the energy consumption of the Internet on a macro timescale by periodically aggregating traffic during lightly loaded periods, and in contrast distributing traffic in heavily loaded periods. The realization of this control mechanism needs modifications of the underlying routing protocols and the routing path selection mechanism.
- On a small timescale, the flow rate varies at a smaller magnitude but more frequently (compared to the time-of-day/time-of-week effect). Accordingly, the GRecRouter also adopts an energy control mechanism operating on a micro timescale, which adaptively tunes the processing rate of the function blocks inside individual routers based on the detected link utilization, traffic rate, or queue length. To minimize energy consumption, fast but more energy-consuming states are suggested to be used under heavy traffic scenarios, while slow but energy-efficient states are suitable for light traffic scenarios.
- With the elaborate design of functional blocks inside the GRecRouter using architectural advances, the router's peak power could also be reduced. Energy-efficient architectural designs for implementing the main router functions in the GRecRouter (routing lookup and packet queuing) are presented later.

To summarize, the design of GRecRouter exhibits many desirable and unique features compared with conventional routers, as shown in Fig. 1. The major features are discussed as follows:

- At the network level, power-aware routing is applied to determine the end-to-end routing paths and forward the packets from the source to the destination with minimal energy consumption. It may change packet routes during lightly loaded periods for traffic aggregation and inform the idle devices to go into sleep mode. It should

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

be noted that with the energy-saving consideration, the expected performance (e.g., quality of service [QoS]) needs to be guaranteed.

- At the node level, rate adaptive processing should be activated inside individual routers. Different function blocks could be flexibly configured to operate in a specific state (e.g., on, off, and low-energy-consumption states with slower clock frequency or/and lower supply voltage) according to the network traffic load.
- At the function block level, each functional block should be designed with energy efficiency in mind so as to reduce the peak power of the router. Many architectural design techniques could be employed, such as caching, clock gating, and processing separation.

The GRecRouter is presented as a conceptual architectural model that is not restricted to any specific deployment. The analysis and discussions on a reference implementation of the GRecRouter are presented in detail in the following sections.

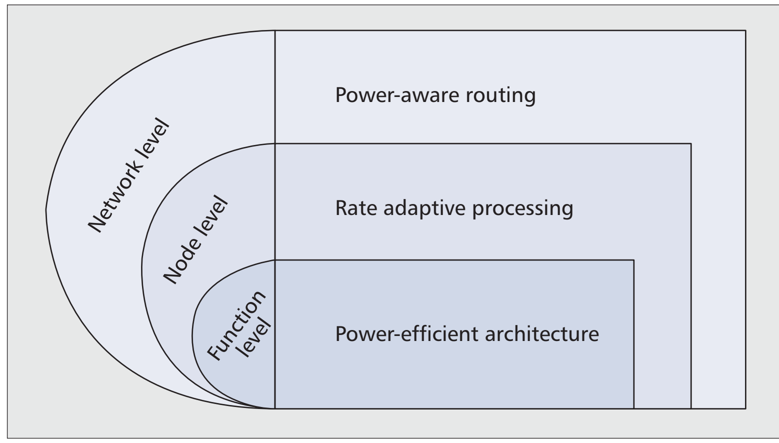


Figure 1. *The GRecRouter controls power dissipation at three levels.*

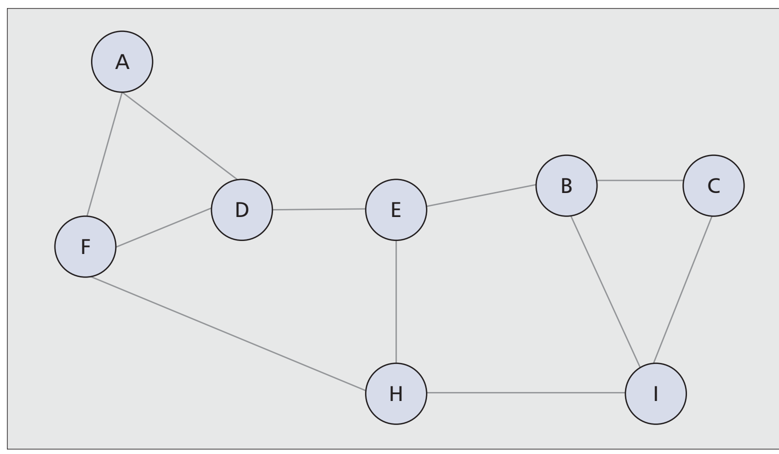


Figure 2. *Topology abstraction of the Abilene network.*

Power-Aware Routing through Green Virtual Networks

As mentioned above, the imbalance of network resource utilization and significant network redundancy offers great potential to reduce Internet energy consumption through adopting a power-aware routing mechanism. In fact, with the modification of some state-of-the-art routing strategies, such as multiprotocol label switching (MPLS) and OSPF, it is technically possible to dynamically shift and aggregate the network traffic based on certain criteria. In recent years, network virtualization technologies [6] have been deemed efficient tools for embedding technological innovations and services in the current Internet against its ossification. In this article, power-aware routing is investigated through the establishment of energy-efficient virtual networks (VNs), that is, green VNs. To save energy, the mapping of green VN requests should use as few nodes and links as possible, at the same time taking the resource into consideration so as to avoid network congestion in the substrate infrastructure.

Figure 2 illustrates a topology abstraction of the Abilene network. We use this topology as an example to present the basic ideas and main issues in power-aware routing through green VNs. Virtual network request mapping is the key problem in green network virtualization. Let us first explain how it works. Suppose the virtual network request or traffic demand among three nodes is illustrated in the right part of Fig. 3, and each link has a unit bandwidth as indicated in left part of Fig. 3 by the labeling on its links. With the green network virtualization rule, (R, S) should be mapped as (A, F, H), which is the shortest path between (A, H). This mapping requires the fewest intermediate nodes; therefore, it would consume the least energy. Similarly, (S, T) is mapped to (H, I, C). For (R, T), there are two shortest paths: (A, F, H, I, C) and (A, D, E, B, C). With the green and energy saving rule, we choose to map (R, T) on (A, F, H, I, C). Although it costs the same resource as (A, D, E, B, C) for link (R, T), all the nodes in this mapping have already been used for (R, S) and (S, T). Thus, this mapping requires no additional physical network nodes.

Due to the fluctuation of traffic rate, the original network service requirements may change, and the changes may require adjustment of the green VN mapping. Suppose now the network is in the state shown in Fig. 4; the original request of (R, S, T) is changed, so node S is no longer required while the traffic demand between R and T still holds. In this case, the map of (R, T) will be changed from the original (A, F, H, I, C) to (A, D, E, B, C). Since route (A, D, E, B, C) is already in working state, such an adjustment on (R, T) would result in the least energy consumption.

Rate Adaptive Processing Inside Routers

Dynamic Voltage and Frequency Scaling

Among the factors determining the dynamic power, voltage and frequency have far-reaching impact. The dynamic power is proportional to the frequency and the square of supply voltage. The frequency also influences the supply voltage: maintaining a higher clock frequency may mean maintaining a higher supply voltage. Therefore,

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

combining the voltage and frequency has a cubic impact on power dissipation. Although voltage and frequency have significant leverage on energy savings, it may degrade system performance due to the decrease of clock frequency. If one could recognize the periods when absolute guarantees or stringent requirements of network performance are not required, the energy could be greatly saved by reducing the voltage and frequency. A power-aware technique to manage the supply voltage and clock frequency, known as dynamic voltage and frequency scaling (DVFS), has been extensively studied in the area of microprocessor (e.g., CPU) design. However, to the authors' best knowledge, little research work has been reported on the application of DVFS in the context of router design.

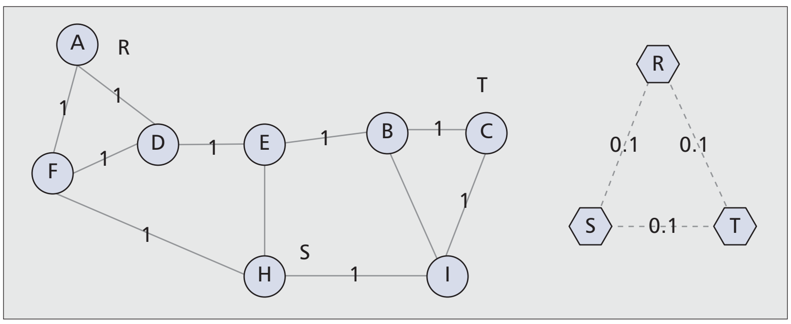


Figure 3. *Example of mapping request.*

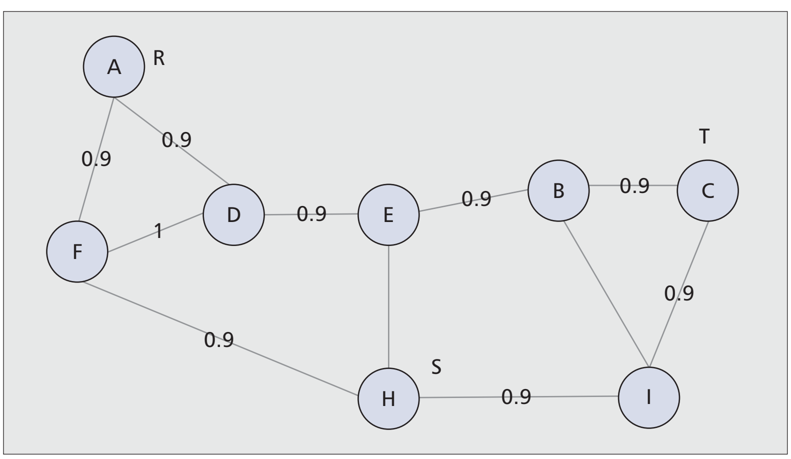


Figure 4. *Example of request adjustment.*

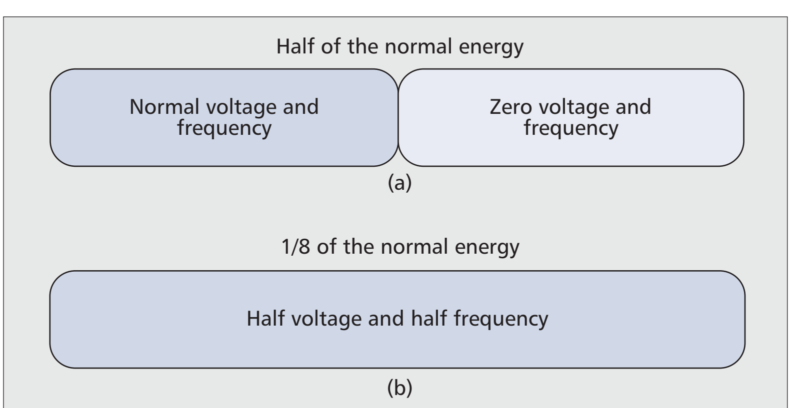


Figure 5. *Benefits of DVFS on energy savings: a) working half of the time and asleep half of the time; b) processing is stretched to the whole time interval.*

Consider a router operation scenario: a router completes the processing of all the packets within a time interval, and energy will be wasted in the remaining idle time periods. A coarse granularity DVFS method is proposed in [7] for Ethernet devices, which attempts to identify the inactive periods of the links and put the associated devices to sleep in such periods. As shown in Fig. 5a, if the processing could be completed by the middle of the interval, half of the energy could be saved in an ideal case. (Suppose zero switching time to sleep state and zero energy consumption in sleep state.) The key question behind this approach is when and for how long to turn off a link. Without detailed knowledge of the incoming traffic pattern, this approach has to compromise with network performance (e.g., packet delay, loss) on energy consumption reduction.

Rather than simply stopping the packet processing in idle periods, more energy is expected to be saved if the packet processing is stretched to a time interval by scaling the frequency and voltage, as illustrated in Fig. 5b, which has the same traffic load as in Fig. 5a. It reduces the clock frequency by half and stretches the processing time to cover the whole time interval. Meanwhile, with halved clock frequency, the voltage could also be scaled down. Suppose that the voltage could also be halved; then the total energy consumed is reduced to 1/8 normal consumption considering the quadratic impact of voltage on power dissipation. It is obvious that more gain will be achieved if the processing is stretched across a larger interval, although resulting in a longer packet delay. Therefore, the energy saving can be optimized subject to the constraint of maximum tolerable delay.

Adaptive Link Rate Interface

Take Ethernet as an example, a 1 Gb/s network interface consumes 4 W more than a 100 Mb/s interface, and a 10 Gb/s interface consumes about 10 to 20 W. In addition, the impact of link utilization on power consumption is considered minimal [8]. Therefore, the interface data rate should be dynamically tailored with flexible configuration according to link utilization. This approach is known as adaptive link rate (ALR), which was proposed in [8] aiming to reduce energy consumption of full-duplex Ethernet networks. The proposed ALR solution only works for Ethernet, and more research is still needed in order to extend ALR to other network interface types. In addition, switching the interface rate with finer-grained steps (instead of only the three steps suggested in [8]) could further enhance the energy consumption efficiency. In this case, two issues need to be addressed for the enhancement of ALR.

First, it can work properly only under the condition that the interfaces of the two end nodes of a link operate at the same data rate. A certain mechanism or protocol is required to negotiate the interface rate between these two nodes. The process of negotiation and switching of different interface rates is expected to be quick so as not to introduce additional packet loss and delay. A two-way handshake procedure could be implemented: a receiver (RX) or transmitter (TX) sends a request message for rate change that contains information on the desired rate; upon receipt of the request, the other node responds with an acknowledgment (ACK) message to agree on the change action or a negative ACK (NACK) to decline the request.

Second, a policy is required to determine when to change the link rate and what the target rate is. The operational states, as well as the interface rate of each state, can be predetermined. As shown in Fig. 6, from left to right, the interface rate increases from idle state *S*0 to the maximum link rate *Sn*. The queue length of RX/TX can be the indicator to trigger the transition of states: an increase of queue length leads to a transition to a state with a faster rate, whereas a decrease of queue length results in a lower rate. The system first operates at the idle state (*S*0 in the figure) before switching to another rate to prevent potential problems caused by glitches [9].

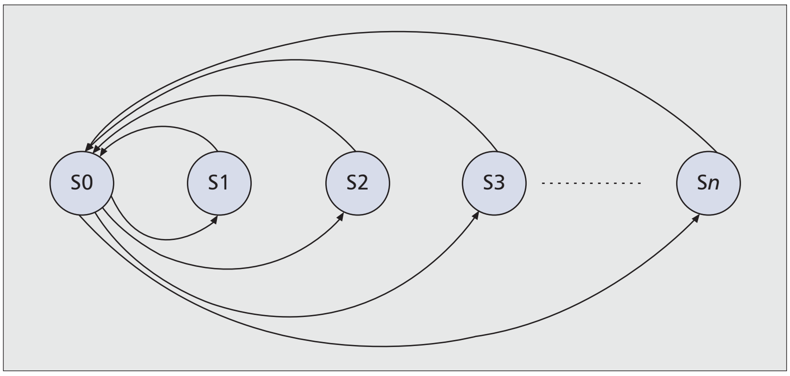


Figure 6. *State machine of link rate transfer.*

Conclusion

Routers are the main equipment used to construct the Internet. By exploring the means to control the energy consumption of routers, the Internet could become energy efficient. The GRecRouter presented in this article seeks opportunities and means to notably reduce the power dissipation at the network, node, and function levels.

Acknowledgment

The authors appreciate the anonymous reviewers for their constructive comments. This work is supported by NSFC (60903182, 61070157, 60773182), 863 Program of China (2008AA01A323, 2009AA01334), and the open project of the State Key Laboratory of Networking and Switching Technology.

References

[1] D. Sarokin, 2007, *Energy Use of Internet*; http://uclue.com/index.php?xq=724.
[2] M. Gupta and S. Singh, "Greening of the Internet," *SIGCOMM '03*, 2003, pp. 19–26.
[3] R. Bolla *et al.*, "Energy-Aware Performance Optimization for Next-Generation Green Network Equipment," *PRESTO '09*, 2009, pp. 49–54.
[4] S. Nedevschi *et al.*, "Reducing Network Energy Consumption via Sleeping and Rate-Adaptation," *NSDI '08*, 2008, pp. 323–36.
[5] C. Fraleigh *et al.*, "Packet-Level Traffic Measurements from the Sprint IP Backbone," *IEEE Network*, vol. 17, no. 6, Nov.–Dec. 2003, pp. 6–16.
[6] N. M. M. K. Chowdhury and R. Boutaba, "Network Virtualization: State of the Art and Research Challenges," *IEEE Commun. Mag.*, vol. 47, no. 7, 2009, pp. 20–26.
[7] M. Gupta and S. Singh, "Using Low-Power Modes for Energy Conservation in Ethernet LANs," *INFOCOM '07*, 2007, pp. 2451–55.
[8] C. Gunaratne *et al.*, "Reducing the Energy Consumption of Ethernet with Adaptive Link Rate (ALR)," *IEEE Trans. Comp.*, vol. 57, no. 4, 2008, pp. 448–61.
[9] Kennedy *et al.*, "Low Power Architecture for High Speed Packet Classification," *ANCS '08*, 2008, pp. 131–40.

Biographies

CHUNMING WU received his B.S., M.S., and Ph.D. degrees in computer science from Zhejiang University, China, in 1989, 1992, and 1995, respectively. He is currently a professor in the College of Computer Science, Zhejiang University, and director of the NGNT laboratory. His research fields include Internet QoS provisioning, reconfigurable network technology, network virtualization, and the architecture of the next-generation Internet.

BINQIANG WANG is Chair Professor and postgraduate and Ph.D. mentor of Information Engineering University, and deputy head of the overall technical committee for the key 863 project "The New Generation of High-Credited Network." His research interest is the architecture of broadband information networks and router switching technology. Currently, he is in charge of the construction of the Flexible Reconfigurable Network. He has published over 150 academic papers and possesses 15 PRC Technological Invention Patents.

CHENGCHEN HU received his B.S. degree from the Department of automation, Northwestern Polytechnical University, Xi'an, China, in 2003; and his Ph.D. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2008. He is currently an associate professor in the Department of Computer Science and Technology of Xi'an Jiaotong University. He is also a guest researcher at SKLNST, BUPT.His research interests include traffic measurement and management, high performance routers, and data center networking.

WEI XIONG received his Bachelor's degree in engineering from Zhejiang University in 2010, and currently is a graduate student in the College of Computer Science and Technology at Zhejiang University and a member of the NGNT laboratory. His research fields include virtual network embedding, reconfigurable network technology, and Internet QoS provisioning.

MING JIANG received B.S. and M.S. degrees in science in 1996 and 2001, respectively, and a Ph.D. degree in computer science in 2004, all from Zhejiang University. He is currently an associate professor at the College of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, Internet QoS provisioning, the differentiated services network, and network congestion control.

JIANGXING WU is an academician of the China National Academy of Engineering and CEO of the China National Digital Switching System Engineering and Technological R&D Center. He also works as Chairman of the Major Dedicated Supervisory Group for China's 863 Hi-Tech Research Program. His major achievements include the first large-capacity digital SPC exchange and the first IN system in China, the first fully distributed mobile switching system in the world, and China's high-speed broadband information demo network.

SERIES EDITORIAL

RADIO COMMUNICATIONS: COMPONENTS, SYSTEMS, AND NETWORKS


Joseph Evans *Zoran Zvonar*

Dear readers:

Continuing our series in 2011, our challenge remains: bring timely, highly relevant contributions of a tutorial nature that can bring new ideas and concepts closer to the Communications Society readership. Radio Communications topics are diverse in nature, spanning components and subsystem enabling technologies, physical and radio access layer approaches, and network techniques. The complexity of modern systems brings all of the disciplines together with a cross-disciplinary emphasis, with enabling solutions being distributed across all aspects of the system. Taking the low power challenge (or maybe calling it the green approach) is an excellent example: savings can come from protocol design, low-complexity physical layer algorithms, and efficient implementations all the way to advances in semiconductor technology. Add on top of this energy harvesting as part of the solution, and one can imagine how important a cross-disciplinary approach is today.

The rate of submissions and the continuous conversations we have with colleagues in the field point to several areas leading the popularity list: LTE topics and future options for wireless systems beyond fourth generation (4G), approaches to cognitive radio, better frequency utilization, and heterogonous systems. It is no surprise that the articles in this issue are focused on those areas.

The first article in the issue, "Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects" by the Nokia Siemens Networks team, provides an overview of one of the key features of LTE-Advanced. The system bandwidth in LTE-Advanced could be approached as contiguous or consisting of several non-contiguous bandwidth segments that can be aggregated. This article provides insight into supported scenarios, an overview of the functionality, as well as interference management schemes.

The second article, "Integrated Solutions for Testing Wireless Communication Systems" by colleagues from Agilent Technologies, outlines the increasing challenges for testing complex wireless communications systems, and uses LTE as a framework to illustrate new approaches and flexible solutions for evolving standards.

The third article, "An Introduction to Millimeter-Wave Mobile Broadband Systems" by authors from Samsung Electronics Technology Laboratory, is an example of new approaches beyond 4G. It looks to the spectrum range above 3 GHz, and demonstrates the feasibility of achieving gigabit per second rates at distances of up to 1 km in an urban mobile environment.

In the final article of the issue, "Spectrally Agile Multicarrier Waveforms for Opportunistic Wireless Access" by Alexander Wyglinski *et al.*, the focus is on spectrally agile wireless access in scenarios where unlicensed users utilize unoccupied licensed bands. Extensions of orthogonal frequencydivision multiplexing techniques to achieve such goals are presented as viable options.

We would like to encourage our readers to be much more proactive in shaping the content of the series — please contact us with your thoughts!

BIOGRAPHIES

JOSEPH B. EVANS [SM] (evans@ittc.ku.edu) is the Deane E. Ackers Distinguished Professor of Electrical Engineering & Computer Science at the University of Kansas. He served as director of the Information and Telecommunication Technology Center from 2008 to 2010 and director of Research Information Technology at the University of Kansas from 2005 to 2008. He served as a program director in the Division of Computer & Network Systems in the Directorate for Computer & Information Science & Engineering at the National Science Foundation from 2003 to 2005. His research interests include mobile and wireless networking, networked computing systems, high-speed networks, and adaptive computing systems. He has been involved in major national high performance networking testbeds and broadband wireless mobile networking efforts, and has published over 110 journal and conference works. He has been a researcher at the Olivetti & Oracle Research Laboratory, Cambridge University Computer Laboratory, USAF Rome Laboratories, and AT&T Bell Laboratories. He has been involved in several startups — as co-founder and member of the board of directors of a network gaming company acquired by Microsoft in 2000, and as a partner in a military software company acquired by General Dynamics in 2010. He received his Ph.D. degree from Princeton University in 1989 and is a member of the ACM.

ZORAN ZVONAR [F] (zoran.zvonar@mediatek.com) is the director of engineering, MediaTek Wireless Technologies, and a MediaTek Fellow. He received his Dipl.Ing. in 1986 and his M.S. degree in 1989 from the Department of Electrical Engineering, University of Belgrade, Serbia, and his Ph.D. degree in electrical engineering from Northeastern University, Boston in 1993. He has been with the Department of Electrical Engineering, University of Belgrade, Yugoslavia, and Woods Hole Oceanographic Institution, Woods Hole, MA. 1994–2008 he had pursued industrial carrier within Analog Devices and has been recipient of the company's highest technical honor of ADI Fellow. Since January 2008 he is with MediaTek focused on the design of algorithms and architectures for cellular standards, with applications to integrated chip-set solutions and real-time software. He is a co-Editor of the Radio Communications Series in *IEEE Communications Magazine*, and has also served as Associate Editor of *IEEE Communications Letters* and Guest Editor of *IEEE Transactions on Vehicular Technology*, the *International Journal of Wireless Information Networks*, and ACM/Baltzer *Wireless Networks*, as well as a member of the editorial boards of a number of professional journals in wireless communications. Also, he was co-editor of *GSM: Evolution Towards Third Generation Systems* (Kluwer), *Wireless Multimedia Networks Technologies* (Kluwer), and *Software Radio Technologies: Selected Reading* (IEEE Press).

Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects

Klaus Ingemann Pedersen, Frank Frederiksen, and Claudio Rosa, Nokia Siemens Networks
Hung Nguyen, Luis Guilherme Uzeda Garcia, and Yuanye Wang, Aalborg University

ABSTRACT

Carrier aggregation is one of the key features for LTE-Advanced. By means of CA, users gain access to a total bandwidth of up to 100 MHz in order to meet the IMT-Advanced requirements. The system bandwidth may be contiguous, or composed of several non-contiguous bandwidth chunks that are aggregated. This article presents a summary of the supported CA scenarios as well as an overview of the CA functionality for LTE-Advanced with special emphasis on the basic concept, control mechanisms, and performance aspects. The discussion includes definitions of the new terms primary cell (PCell) and secondary cell (SCell), mechanisms for activation and deactivation of CCs, and the new cross-CC scheduling functionality for improved control channel optimizations. We also demonstrate how CA can be used as an enabler for simple yet effective frequency domain interference management schemes. In particular, interference management is anticipated to provide significant gains in heterogeneous networks, envisioning intrinsically uncoordinated deployments of home base stations.

INTRODUCTION

The first version of Long Term Evolution (LTE) was completed in March 2009 as part of Third Generation Partnership Program (3GPP) Release 8 (Rel-8) [1]. LTE is based on a flat radio access network architecture without a centralized network component, offering flexible bandwidth options ranging from 1.4 to 20 MHz using orthogonal frequency-division multiple access (OFDMA) in the downlink and single-carrier frequency-division multiple access (SC-FDMA) in the uplink [1]. Multiple-input multiple-output (MIMO) up to order 4 × 4 are supported for the downlink, while only single-layer transmission is supported in the uplink. In March 2008, 3GPP started a new study item in order to further develop LTE toward LTE-Advanced targeting the IMT-Advanced requirements as defined by the International Telecommunication Union (ITU) [2–5]. The LTE-Advanced study item was closed in March 2010. The outcome was a set of new radio features, which are currently being standardized to become part of LTE-Advanced in 3GPP Rel-10.

Carrier aggregation (CA) is one of the main features of LTE-Advanced in Rel-10 for meeting the peak data rate requirements of IMT-Advanced, 1 Gb/s and 500 Mb/s for the downlink and uplink, respectively [6]. This article provides a thorough overview of CA for LTE-Advanced, while elucidating its impact on overall system design and performance. Although we primarily focus on CA for the downlink of frequency-division duplex (FDD) systems, CA is supported in the uplink as well as in time-division duplex (TDD) systems [7].

CA is designed to be backward compatible, meaning that legacy Rel-8 and Rel-9 users should still be able to coexist with LTE-Advanced on at least part of the total bandwidth. Thus, each individual spectrum chunk, denoted component carrier (CC), inherits the core physical layer design and numerology from LTE Rel-8. Nevertheless, the introduction of CA for LTE-Advanced does include new functionalities and modifications to the link layer and radio resource management (RRM) framework. In our description of such modifications for LTE-Advanced, we assume that the corresponding LTE Rel-8 design is known by readers, who may otherwise refer to [1, 8, 9] for additional information.

Additionally, we discuss the potential of CA as an enabler for new frequency domain interference management schemes, providing attractive gains for heterogeneous environments with dense deployment of small base station nodes (e.g., pico or home base stations). For example, a fully distributed interference management concept with a CC resolution, called autonomous component carrier selection (ACCS), has been proposed in [10].

A set of system-level performance results are presented in order to demonstrate the benefits of CA. In particular, we focus on comparing the performance of N separate LTE Rel-8 carriers vs. using CA of N carriers. The performance comparison is presented for a dynamic birth-death traffic model to illustrate how the performance varies with the offered traffic per cell. Performance results for heterogeneous networks with dense deployment of small base station

 0163-6804/11/$25.00 © 2011 IEEE

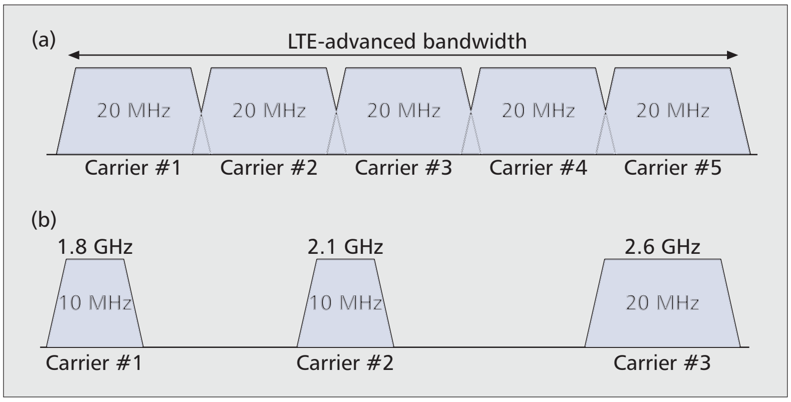


Figure 1. *Example of carrier aggregation scenarios: a) contiguous aggregation of five component carriers with equal bandwidth; b) non-contiguous aggregation of component carriers with different bandwidths.*

nodes are also presented in order to illustrate the potential of the developed ACCS concept.

The rest of the article is organized as follows. We outline the scenarios and basic assumptions for CA configurations. The CA functionality and impact on RRM algorithms is then described. We address interference management on a carrier resolution, followed by presentation of performance results. Finally, we recapitulate the main findings and point out future work.

CA Scenarios and CC Types

The maximum supported bandwidth for LTE-Advanced of 100 MHz can be achieved via CA of 5 CCs of 20 MHz, as illustrated in Fig. 1a. Thus, an LTE-Advanced user supporting such high bandwidths can be served simultaneously on all 5 CCs. The bandwidth of each CC follows the LTE Rel-8 supported bandwidth configurations, meaning 1.4, 3, 5, 10, 15, and 20 MHz. The aggregated CCs may be contiguous, as illustrated in Fig. 1a, or non-contiguous, as depicted in Fig. 1b. Notice also from the example in Fig. 1b that the aggregated CCs can in principle also have different bandwidths. The support for both contiguous and non-contiguous CA of CCs with different bandwidths offers significant flexibility for efficient spectrum utilization, and gradual refarming of frequencies previously used by other systems such as Global System for Mobile Communications (GSM) and code-division multiple access (CDMA). From an implementation and physical layer perspective, contiguous CA is easier, in the sense that it can be realized with a single fast Fourier transform (FFT) and a single radio frequency (RF) unit, while non-contiguous CA in most cases requires multiple RF chains and FFTs. The non-contiguous CA cases have additional implications: the radio network planning phase and the design of the RRM algorithms need to take into account that different CCs will exhibit different path loss and Doppler shifts. For example, Doppler shift influences the ability to gain from frequency domain packet scheduling within a CC [8].

Notice that for LTE Rel-8 with FDD, uplink and downlink carriers are always paired with options for defining the frequency duplex distance and bandwidth through system information signaling. With CA it is also possible to have asymmetric configurations; thus, for example, there are multiple downlink CCs configured for a unit of user equipment (UE) and only one uplink CC. The linking between uplink and downlink configured CCs is signaled to the UE with higher-layer signaling. For each LTE-Advanced user, a CC is defined as its primary cell (PCell) [7]. Different users may not necessarily use the same CC as their PCell. The PCell can be regarded as the anchor carrier for the terminal and is thus used for basic functionalities such as radio link failure monitoring. If more than one CC is configured for a user, the additional CCs are denoted secondary cells (SCells) for the user.

Functionality and Terminology

Protocol Stack

Figure 2 shows an overview of the downlink user plane protocol stack at the base station, as well as the corresponding mapping of the most essential RRM functionalities for CA. Each user has at least one radio bearer, denoted the default radio bearer. The exact mapping of data to the default bearer is up to the operator policy as configured via the traffic flow template (TFT). In addition to the default radio bearer, users may have additional bearers configured. There is one packet data convergence protocol (PDCP) and radio link control (RLC) per radio bearer, including functionalities such as robust header compression (ROHC), security, segmentation, and outer automatic repeat request (ARQ). Thus, the PDCP and RLC are the same as in LTE Rel-8 [1, 8, 9]. The interface between the RLC and the medium access control (MAC) is referred to as *logical channels*. There is one MAC per user, which controls the multiplexing (MUX) of data from all logical channels to the user, and how this data is transmitted on the available CCs. As illustrated in Fig. 2, there is a separate hybrid ARQ (HARQ) entity per CC, which essentially means that transmitted data on CC #X shall also be retransmitted on CC #X in case prior transmission(s) are erroneous. The interface between the MAC and physical layer (PHY) — denoted *transport channels* — is also separate for each CC. The transport blocks sent on different CCs can be transmitted with independent modulation and coding schemes, as well as different MIMO coding schemes. The latter allows data on one CC to be transmitted with open loop transmit diversity, while data on another CC is sent with dual stream closed loop precoding. Thus, there is independent link adaptation per CC to benefit from optimally matching the transmission on different CCs according to the experienced radio conditions (i.e., corresponding to frequency domain link adaptation on a CC resolution). The system also allows using different transmit power settings for the CCs so that in principle they could have different levels of coverage as also discussed in [7].

The LTE Rel-8 control plane protocol stack also applies to LTE-Advanced with multiple CCs, meaning that there is one radio resource control (RRC) per user, independent of the number of CCs. Similarly, idle mode mobility procedures of LTE Rel-8 also apply in a network deploying CA. It is also possible for a network to configure only a subset of CCs for idle mode camping.

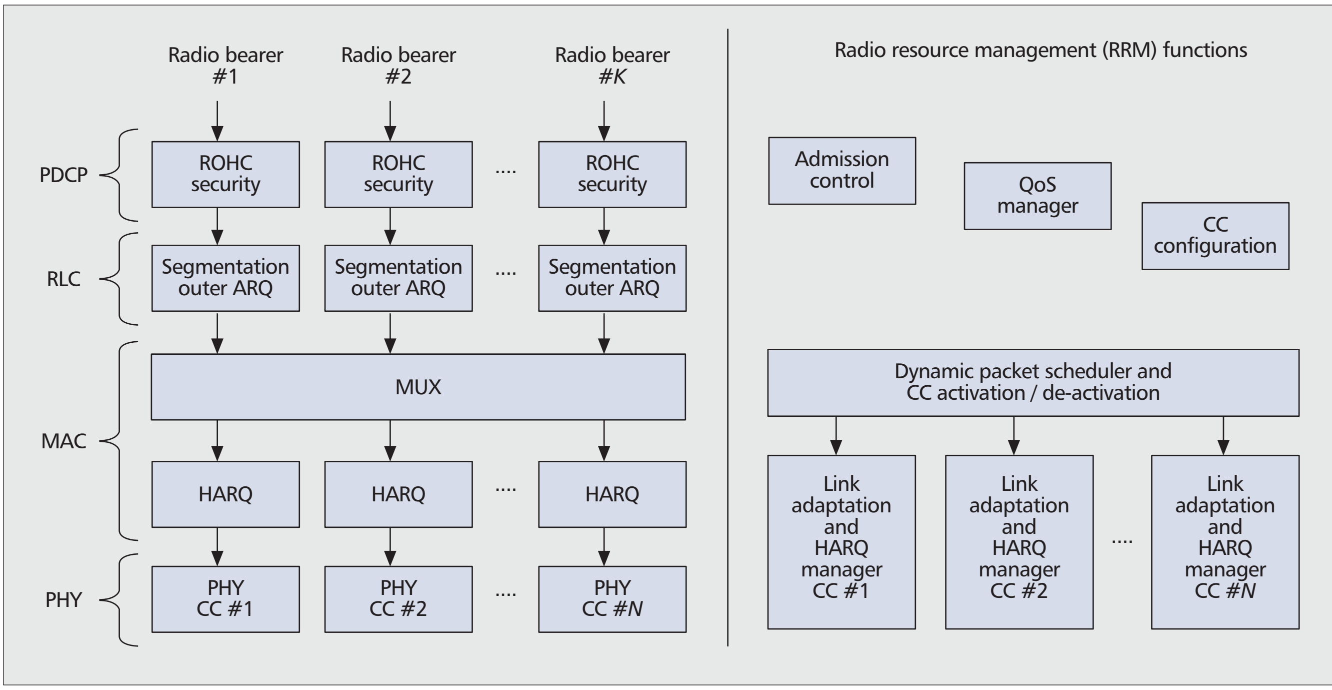


Figure 2. *Overview of the downlink user plane architecture (left side) and the corresponding RRM algorithms (right side).*

RRM Considerations

The RRM framework for LTE-Advanced has many similarities with that of LTE Rel-8 [9]. Admission control is performed at the base station prior to establishment of new radio bearers, and the corresponding quality of service (QoS) parameters are configured. The QoS parameters are the same for LTE Rel-8 and LTE-Advanced, and are thus CC-independent; see more information in [1, 8, 9]. However, a new RRM functionality is introduced with LTE-Advanced, which we refer to as *CC configuration* in the following. The latter functionality configures a CC set for each user. The CC set is the collection of CCs where the user may afterward be scheduled. The CC set is configured to the users with RRC signaling. The CC configuration functionality is an important apparatus for optimizing system performance, as well as limiting the power consumption of users. The latter originates from the fact that the power consumption per user increases with the number of CCs a user has to receive (i.e., increases with bandwidth it needs to process). The overall framework for the CC configuration is illustrated in Fig. 3, where an example of input information is illustrated. For each user, QoS parameters, radio bearer configuration, and terminal capability are useful a priori knowledge for determining the CC set. Legacy Rel-8 users naturally only support one CC, and shall therefore only be allocated on a single CC. For optimal system performance, it is desirable to have approximately equal load on different CCs, so own-cell load information (including load per CC) is needed as input as well to facilitate optimal CC load balancing and configuration [11]. For LTE-Advanced users supporting multiple CCs, QoS parameters such as the QoS class identifier (QCI), guaranteed bit rate (GBR), and aggregated maximum bit rate (AMBR) for non-GBR bearers provide useful information for determining the number of required CCs for the user. As an example, users only having a voice over IP (VoIP) call or a streaming connection with moderate GBR can be assigned a single CC, while still being able to fulfill the users' QoS requirements. For users with best effort traffic, the AMBR can be used to estimate the most sensible CC set size for them. Assigning a single CC to such a user has the advantage that terminal power consumption is kept lower than cases where the user is configured with a CC set larger than one. Second, corresponding control signaling overhead is also reduced by configuring a smaller number of CCs for the user. The exact algorithm for the CC configuration functionality is base station vendor specific, and thus not strictly specified in the standard.

As illustrated in Fig. 2, the layer 2 packet scheduler (PS) is tightly coupled with an additional functionality for more dynamically (de-)activating CCs configured as SCells for different users. This functionality is anticipated as an additional control tool to further optimize the users' power consumption. A user is only schedulable on configured and activated CCs, while it is not schedulable on deactivated CCs. Similarly, a user does not report channel state information (CSI) for deactivated CCs as needed by the base station for radio-channel-aware link adaptation and frequency domain packet scheduling [9]. SCells are activated/de-activated independently via MAC signaling [7]. It is furthermore possible to set a so-called deactivation timer, so an activated SCell automatically gets deactivated without an explicit de-activation message if no traffic has been scheduled on the CC for a given time period. Configured SCells are by default de-activated, so they have to be explicitly activated before being schedulable. However, the PCell

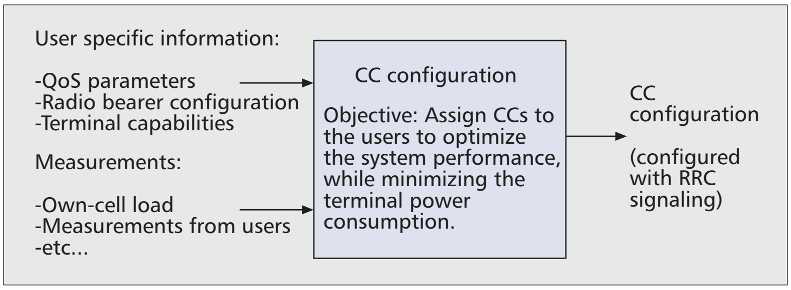


Figure 3. *Overview of CC Configuration functionality, including illustration of possible input parameters.*

for a user is always assumed to be activated and is therefore not subject to any de-activation procedures [7].

The dynamic PS at layer 2 is responsible for scheduling eligible users on their configured and activated CCs. In coherence with the LTE Rel-8 PS framework [9], the smallest frequency domain scheduling resolution within each CC is a physical resource block (PRB) of 12 subcarriers, constituting an equivalent bandwidth of 180 kHz. The PS aims at benefiting from multi-user frequency domain scheduling diversity by primarily allocating PRBs to different users that experience good channel quality (i.e., avoiding scheduling users on PRBs in deep fades). The PS functionality for LTE-Advanced with CA is very similar to the PS for LTE Rel-8, except that the LTE-Advanced PS is allowed to schedule users across multiple CCs. The fact that LTE-Advanced relies on independent transport blocks, link adaptation, and HARQ per CC opens up various implementations of the scheduler. As an example, the scheduling could be done in parallel for the different CCs, including some coordination to ensure fairness and joint control for users scheduled on multiple CCs [11]. As in LTE Rel-8, dynamic scheduling of a user is facilitated via sending a scheduling grant on the control channel (called the physical dedicated control channel, PDCCH), which is time-multiplexed in each transmission time interval (TTI) just before the data channel [9]. One PDCCH is limited to one CC, and the same addressing is used per user independent of the CC where it is scheduled (called the cell radio network temporary identifier, C-RNTI, in 3GPP LTE terminology). However, LTE-Advanced includes enhancements allowing the base station to send a scheduling grant on one CC for scheduling the user on another CC. The latter is referred to as cross-CC scheduling as the scheduling grant and the corresponding data transmission takes place on different CCs. The cross-CC scheduling functionality is incorporated by appending a so-called carrier indicator field (CIF) to the downlink control information (DCI). The DCI is used to indicate the user allocations for uplink and downlink traffic, and the CIF is used to address on which CC the user data is transmitted. When the CIF is appended to the DCI, the payload size increase slightly, and as the radio resources for the transmission of the data is constant, the link performance is slightly worse due to weaker coding. The user configuration and interpretation of the CIF is semi-statically configured on a per-UE basis, and is thus fully backwards compatible with legacy Rel-8 users not having the CIF in the DCI transmitted on the PDCCH. The cross-CC scheduling functionality offers additional system flexibility for further optimizing control and data channel performance across multiple CCs.

In addition to the dynamic Layer-2 packet scheduling, LTE Rel-8 also supports so-called semi-persistent-scheduling (SPS) as a special packet scheduling mode for quasi-deterministic traffic flows such as VoIP to save control channel resources [9]. SPS is also supported for LTE-Advanced with CA, but is limited to be configured on the users PCell only (configured via RRC signaling).

Dynamic Interference Management

For properly planned macro cellular networks, it has typically been found that deployment of LTE (or LTE-Advanced) with plain frequency reuse one is an attractive configuration: simply put, all cells have access to all CCs. However, on heterogeneous networks (HetNet) the interference footprint deviates significantly from that of planned macro cells. This arises from the coexistence of the ordinary macro cell layer with a layer of scattered smaller base station such as micro, pico, and home base stations (HeNB) with closed subscriber groups (CSG). Specifically, dense roll-outs of co-channel CSG HeNBs, popularly known as femtocells, are bound to result in chaotic inter-cell interference if left completely unchecked. It has therefore been found that HetNet cases in many scenarios can benefit from interference management. It then follows naturally, that CA could be employed as a new and promising instrument of inter-cell interference coordination in the frequency domain. The frequency reuse, i.e., CC, configuration yielding the most attractive performance is time-variant and depends on many factors such as the traffic distribution, the relative location of base stations, their mutual interference coupling, etc. Thus, manual configuration of the optimal CC usage pattern becomes nearly impossible.

In an ideal world, each base station node would dynamically select from a finite set which CCs it should deploy. Figure 4 shows an example of a scenario with three available CCs for each base station node. The selected CCs by each node are marked with the dark blue color code, meaning that e.g., the macro eNB is using all three CCs. For the densely deployed indoor HeNBs/pico nodes, each node only uses a subset of available CCs, as this is the best configuration for optimizing the system performance as there is severe interference coupling between those nodes. Notice that by conducting the adaptive frequency reuse on CC resolution, both data and control channels experience benefits as all physical channels are within a single CC.

In this light, a concept called autonomous CC selection (ACCS) has therefore been proposed. The interested reader can find a comprehensive

IEEE Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

description of ACCS in [10], yet its key principles are outlined next. The basic ACCS concept is based on three fundamental premises:

- Each base station node has the right to always have at least one active CC with full cell coverage.
- As the offered traffic increases, additional CCs can be taken into use to increase its capacity.
- However, a base station node is only allowed to take additional CCs into use, provided it does not result in excessive interference to the surrounding cells.

The condition expressed last shall prevent so-called greedy base station nodes from generating disruptive interference levels that severely reduce the performance of surrounding cells. Thus, before a node takes additional CCs into use, it shall estimate the impact on the surrounding cells. The latter evaluation relies on Background Interference Matrices (BIMs) which are built locally by each HeNB based exclusively on *downlink* reference signal received power (RSRP) measurements. Such measurements are processed in a meaningful way and subsequently exchanged among HeNBs. The BIM information essentially predicts the downlink carrier to interference ratios (C/I) experienced whenever two cells (serving and interferer) use the same CC at the same time with equal transmit power spectrum densities. Consequently, by collecting RSRP measurements from the terminals for different cells, each eNB "learns" the interference coupling with neighboring cells in terms of C/I ratios. It is relevant to mention that the collection of various measurements is a by-product of normal system operation and does not entail an extra burden to UEs. Thus, ACCS is essentially a fully distributed and dynamic interference management concept operating in the frequency domain on a CC resolution, based on sensing (measurements) and minimal signaling between base station nodes. A related autonomous carrier selection concept is outlined in [12].

PERFORMANCE OF CA

In order to further illustrate the gain of using CA, extensive system level simulations are conducted for a configuration with 2 × 20 MHz in the downlink. The considered environment is a standard 3-sector macro cellular layout with 500 meter inter-site distance and a 2 × 2 antenna configuration with rank adaptation (also known as macro case #1 environment by 3GPP in Technical Report 36.814). Simulations are conducted for cases where all users are legacy Rel-8 (single CC per user) as well as for cases where all users are LTE-Advanced, and thus are schedulable on multiple CCs. For the cases with Rel-8 users, we use a simple round robin CC load balancing approach, where we aim at having the same number of users allocated on each CC [11]. Frequency domain radio channel aware proportional fair scheduling is applied within each CC. A dynamic birth-death traffic model is considered, where new users arrive in the system according to a homogenous Poisson process (birth process) [3]. The payload for each best effort user equals 4 Mbit, and once this data amount has been successfully delivered to the user the call is terminated (death process). Figure 5 shows the mean experienced, 5 percent-ile (outage), and 95 percent-ile (peak) user data rates versus the average offered traffic per cell. At low offered traffic, it is observed that LTE-Advanced with CA offers significant gains in both mean experienced user data rates and outage performance. A two-fold improvement is exhibited for the LTE-Advanced cases due to using two CCs as compared to the legacy Rel-8 users that are restricted to a single CC. Thus, when there is only a single user in the cell (low offered load), the LTE-Advanced user has access to double bandwidth for this particular example, and hence experience twice as good performance. As the average offered traffic per cell increases to the point where multiple simultaneously schedulable users are present at both CCs for both the Rel-8 and LTE-Advanced cases, the gap between experienced data rate of the two user categories diminishes. This behavior is observed because the user experienced performance for large values of *N* users is approximately the same independently of whether *N* users are multiplexed across 2 CCs, or two groups of N/2 users are multiplexed on each CC. The latter observation links to discussions earlier and Fig. 3, where we recommended that the number of CCs configured per user be done as a function of own-cell load. Thus, for highly loaded cells, one may as well configure a single CC per user to save on terminal power consumption.

In order to further exemplify the possibilities opened by CA, the performance of the proposed autonomous interference management concept for local areas (ACCS) is illustrated in Fig. 6 for a case with three CCs. The three curves shown therein are the result of extensive simulations modeling a dense urban environment with two building blocks separated by a 10 m wide street, totaling 120 apartments. Each block consists of 60 apartments, 20 per floor, assuming a CSG access policy. In our analysis, the probability of having a HeNB deployed and a single active LTE-Advanced user per apartment assumed the values of 25 percent, 50 percent and 75 percent in order to emulate the transition from slightly sparser to extremely dense HeNB deployments. Both HeNBs and UE are dropped uniformly at random indoor locations, while macrocells are not considered here. A full buffer traffic model

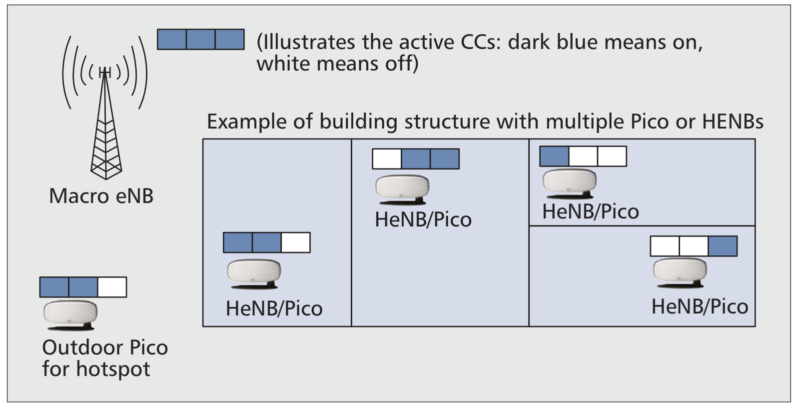


Figure 4. *Simple illustration of autonomous CC selection (ACCS) principle for heterogeneous networks.*

is assumed. Figure 6 shows the relative performance for two different static frequency reuse schemes (labeled R1 and R3) and ACCS. Here R1 refers to plain reuse one (all CCs used by all HeNBs, while R3 corresponds to reuse 3, i.e., each HeNB only uses one of the three available CCs. All mean throughput results were normalized by the maximum theoretical capacity of the system. Hence, a normalized throughput of 100 percent means transmission over the whole bandwidth (all 3 CCs) at the maximum system spectral efficiency.

When compared to universal reuse, the simple yet adaptive nature of ACCS leads to a vastly superior performance in terms of experienced 5 percent-ile (outage) data rates, on par with those offered by sparser and often unpractical preplanned frequency patterns. This trait is especially relevant in very dense deployments, represented by the leftmost points of each curve in Fig. 6. Additionally, it also retains the benefits offered by universal reuse in terms of average data rates, simply because ACCS, as opposed to sparser reuse schemes, does not render cells severely band-limited when that is absolutely not required, for example, in sparser deployments. In fact, it may even surpass the performance of universal reuse, since it allows for a sensible trade-off between bandwidth and signal-to-interference-plus-noise ratio. The latter becomes evident when the rightmost points of each curve are compared.

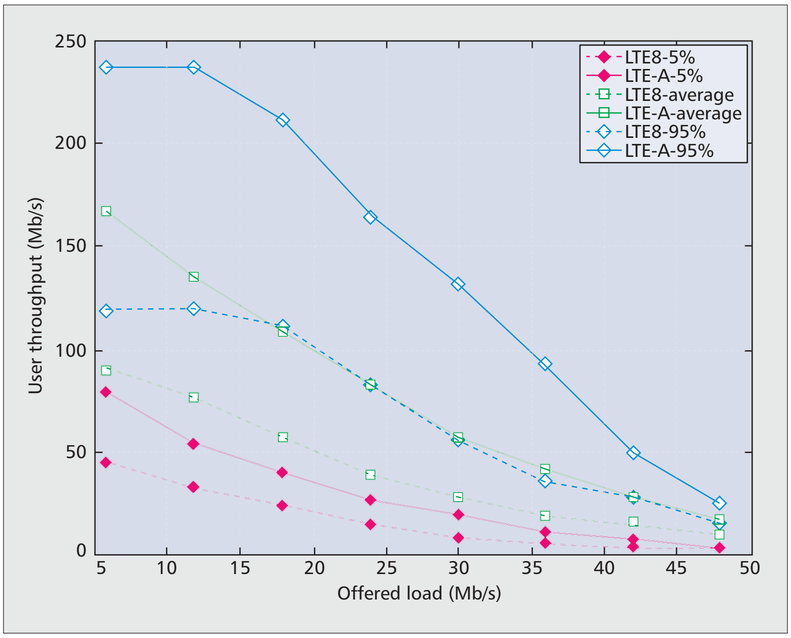


Figure 5. *Experienced user throughput performance vs. the average offered load per cell.*

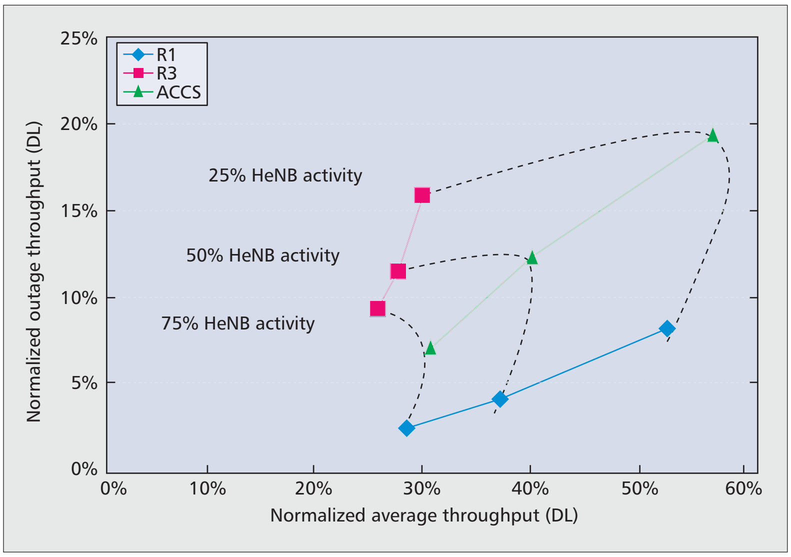


Figure 6. *Relative performance and of different fixed frequency reuse schemes and ACCS for an environment with densely deployed CSG HeNBs.*

Conclusion

In this article we outlined the basic CA concept for LTE-Advanced with both contiguous and non-contiguous aggregation of bandwidths up to 100 MHz. The larger bandwidth obviously results in improved user data rates. But equally important, CA is a powerful feature that enables more flexible and optimal utilization of frequency assets. Especially, non-contiguous CA offers new opportunities for gradually starting to use more and more frequency resources for LTE in different bands that previously were used for e.g., GSM or CDMA without suffering in peak data rates. CA for LTE-Advanced is fully backward compatible, which essentially means that legacy Rel-8 terminals and LTE-Advanced terminals can co-exist. The latter is achieved by relying on MAC level CA with independent Release-8 compliant HARQ and link adaptation per CC. This also implies that CA is transparent from layer 3 and up for the user plane. A flexible layered approach for managing the CCs per LTE-Advanced user is defined; offering configuration of CCs per user via RRC signaling, followed by MAC signaling for activation/ de-activation of CCs configured as SCells. The aforementioned control procedures facilitate efficient power management of terminals, so they are not always mandated to operate at their full bandwidth capability. We have also demonstrated how CA offers attractive opportunities for managing the interference in heterogeneous networks with a mixture of macro cells and various local area smaller base stations (e.g., pico and home base stations). The presented autonomous CC selection concept offers attractive gains for such cases, and can be regarded as a "light cognitive radio" solution, which is facilitated via special use of CA combined with sensing (i.e., based on measurements).

As a last remark, it should be noted that the final standardization of CA for LTE Rel-10 is currently ongoing. In order to meet the IMT-Advanced peak data rate requirements of 1 Gb/s in downlink and 500 Mb/s in uplink, when standardizing the RF requirements 3GPP initially focused on intraband aggregation of carriers with a channel bandwidth larger than or equal to 10 MHz to form an aggregated bandwidth of up to 40 MHz. Currently new bandwidth combinations for interband CA are being agreed in standardization to cover the most interesting cases for operators around the world. To speed up the standardization work, different timescales are set for downlink and uplink, so Rel-10 will only support interband CA in downlink and for a limited number of bandwidth combinations, while full support for non-contiguous CA will come with Rel-11. However, all the related signaling procedures for CA in Rel-10 are standardized so that CA over other bands for both downlink and uplink can be added in later releases by specifying the RF requirements for the corresponding bandwidth combinations.

Acknowledgements

We highly appreciate the joint carrier aggregation work done together with many of our Nokia and Nokia Siemens Networks colleagues, as well as all the received comments and advice during the studies that made the writing of this article possible.

References

[1] H. Holma and A. Toskala, Editors, *LTE for UMTS, OFDMA and SC-FDMA Based Radio Access*, Wiley, 2009.
[2] C. Wijting *et al.*, "Key Technologies for IMT-Advanced Mobile Communication Systems," *IEEE Wireless Commun.*, June 2009, pp. 76–85.
[3] 3GPP Technical Report 36.814, "Further Advancements for E-UTRA Physical Layer Aspects," v. 9.0.0, Mar. 2010; http://www.3gpp.org.
[4] 3GPP Technical Report 36.912, "Feasibility study for Further Advancements for E-UTRA (LTE-Advanced)," v. 9.2.1, Mar. 2010; http://www.3gpp.org
[5] P. E. Mogensen *et al.*, "LTE-Advanced: The Path towards Gigabit/s in Wireless Mobile Communications," *IEEE Proc. Wireless Commun., Vehic. Tech., Info. Theory and Aerospace & Electronic Systems Technology*, May 2009, pp. 147–51.
[6] G. Yuan *et al.*, "Carrier Aggregation for LTE-Advanced Mobile Communication Systems," *IEEE Commun. Mag.*, Feb. 2010, pp. 88–93.
[7] M. Iwamura *et al.*, "Carrier Aggregation Framework in 3GPP LTE-Advanced," *IEEE Commun. Mag.*, Aug. 2010, pp. 60–67.
[8] A. Larmo *et al.*, "The LTE Link-Layer Design," *IEEE Commun. Mag.*, June 2009, pp. 52–59.
[9] K.I. Pedersen *et al.*, "An Overview of Downlink Radio Resource management for UTRAN Long-Term Evolution," *IEEE Commun. Mag.*, July 2009, pp. 86–93.
[10] L. Garcia *et al.*, "Autonomous Component Carrier Selection: Interference Management in Local Area Environments for LTE-Advanced," *IEEE Commun. Mag.*, Sept. 2009, pp. 110–16.
[11] Y. Wang *et al.*, "Carrier Load Balancing and Packet Scheduling for Multi-Carrier Systems," *IEEE Trans. Wireless Commun.*, May 2010, pp. 1780–89.
[12] J. Li *et al.*, "Flexible Carrier Aggregation for Home Base Station in IMT-Advanced System," *Proc. 5th Int'l. Conf. Wireless Commun., Networking and Mobile Computing*, Sept. 2009, pp. 285–88.

Biographies

KLAUS INGEMANN PEDERSEN (klaus.pedersen@nsn.com) received his M.Sc.E.E. and Ph.D. degrees in 1996 and 2000 from Aalborg University, Denmark. He is currently with Nokia Siemens Networks in Aalborg, Denmark, where he is working as a senior wireless network specialist. His current work is related to 3GPP standardization of LTE-Advanced. The latter includes work on carrier aggregation, heterogeneous networks with mixture of different base station types, and system performance.

LUIS GUILHERME UZEDA GARCIA received his Electronics Engineer degree from the Universidade Federal do Rio de Janeiro (UFRJ) in 2003 and his M.Sc. degree from COPPE/UFRJ, Brazil, in 2006. He also worked at Instituto Nokia de Tecnologia (INdT), Brasília, Brazil, as a researcher for nearly four years. He is currently working toward a Ph.D. degree at Aalborg University, Denmark, in close cooperation with Nokia Siemens Networks. His current research interests include radio resource management and concept development for 3GPP LTE-Advanced.

HUNG T. NGUYEN received his M.Sc. degree in Telecommunication from Denmark Technical University and his Ph.D. degree in wireless from Mobile Communication at Aalborg University in 2002 and 2005, respectively. After his Ph.D., he worked on various topics within the Wireless Communication field at Aalborg University as a postdoctoral researcher. Currently he is working on radio resource management topics for LTE and LTE-Advanced systems in cooperation with Nokia Siemens Networks.

YUANYE WANG received his M.Sc.E.E. and Ph.D. degrees from Aalborg University in 2007 and 2010, respectively. He is now working as a post-doc with the Radio Access Technology (RATE) section at Aalborg University, in close cooperation with Nokia Siemens Networks. His current work focuses on enhanced interference management for heterogeneous networks.

FRANK FREDERIKSEN received his M.Sc.E.E. degree from Aalborg University in 1994. Until 2000 he worked for Aalborg University implementing radio channel measurement systems. Since 2000 he has been working for Nokia and Nokia Siemens Networks, where the main focus has been on 3GPP standardization work for HSPA, LTE and LTE-Advanced with main focus on physical and MAC layers.

CLAUDIO ROSA received his M.Sc. E.E. and Ph.D. degrees in 2000 and 2005 from Aalborg University. In 2003 he also received an M.Sc.E.E. degree in telecommunication engineering from Politecnico di Milano, Italy. Since 2005 he has been working with Nokia and Nokia Siemens Networks as a wireless networks specialist with focus on 3GPP standardization work and product-oriented research for LTE and LTE-Advanced. His current work is mainly directed toward the area of carrier aggregation.

All the related signaling procedures for CA in Rel-10 are standardized so that CA over other bands for both downlink and uplink can be added in later releases by specifying the RF requirements for the corresponding bandwidth combinations.

Integrated Solutions for Testing Wireless Communication Systems

Dingqing Lu and Zhengrong Zhou, Agilent Technologies Inc.

ABSTRACT

Wireless communications standards have been evolving rapidly to increase system performance, which poses significant challenges on developing complex test and verification algorithms and schemes early on in a product's life cycle. An integrated test solution answers these challenges through a scalable and reconfigurable integrated test system that is coordinated via integrated core software. This solution not only improves a product's time to market, but is also more efficient from production and economic points of view.

INTRODUCTION

Emerging wireless communication standards have been evolving rapidly since the early second-generation digital cellular systems. With each new generation, there has been a significant increase in spectral efficiency, radio frequency (RF) bandwidth, and peak data rates, which has resulted in higher system capacities and user services. As standards have advanced, the system structure has become more complex, requiring upgrades to new test signals and system performance measurements.

Historically, engineers have coped with this ever changing test challenge by purchasing measurement equipment that supports newer standards, signals, and measurements. In some cases, though, the required standards-based test equipment may not actually be available. With time to market a key issue for engineers designing and testing today's wireless communication systems, this traditional approach of using fixed test waveforms and measurements is no longer optimal. Instead, engineers now require a more dynamic test system. Luckily, a critical trend is emerging in communication systems testing that promises to address this issue head on: the reconfigurable test system. Such a test system not only solves the time-to-market issue, but is also much more efficient from the production and economic viewpoints.

In this article an integrated test solution with reconfiguration capability for testing the evolving wireless communication system is introduced. This solution utilizes integration core software (ICS) to seamlessly integrate all test system instruments, which may include a hardware-based vector signal generator (VSG) and a vector signal analyzer (VSA) with sufficient bandwidth and memory to handle new test waveforms, for automatic testing.

A critical function of the ICS is to generate reconfigurable test waveforms and measurements based on different test standard requirements using an easy-to-use graphical user interface (GUI). The resulting waveforms and measurements can be linked to the test system's instruments to generate test signals and extend measurement capabilities. For receiver performance tests, ICS provides a "golden" transmitter and receiver to act as test references during a new product's development phase. Since no "ideal" physical transmitter or receiver exists, these references serve as critical benchmarks for both transmitter/receiver testing and design troubleshooting.

One of ICS's key benefits is that it extends instrument measurement capabilities by performing pre-programmed measurements (bit error rate [BER], block error rate [BLER], and throughput), as well as more specific measurement standards such as adjacent channel selectivity and blocking. Another advantage of ICS is that it enables the design of customized measurements, a feature that is highly desirable for new protocol or algorithm design. Additionally, it allows complex receiver measurements required by international standards to be derived from existing generic measurements. For example, the reference level sensitivity for a Mobile WiMAX™ receiver can be derived from the frame error rate (FER) measurement, and the reference level sensitivity for Long Term Evolution (LTE) can be derived from the throughput measurement.

The remainder of this article is organized as the follows. The basic structure of the integrated test system is provided and described. An example of LTE base station receiver test is discussed. The test results for the reference sensitivity power level is provided. We then give a summary of the article.

INTEGRATED TEST SYSTEM

Because ICS plays such a critical role in the reconfigurable integrated test system, it is useful to better understand how it interacts with the test system's other instrumentation. The basic configuration of an integrated test system is shown in Fig. 1, and includes ICS, a VSA, and a VSG with arbitrary waveform generation (ARB).

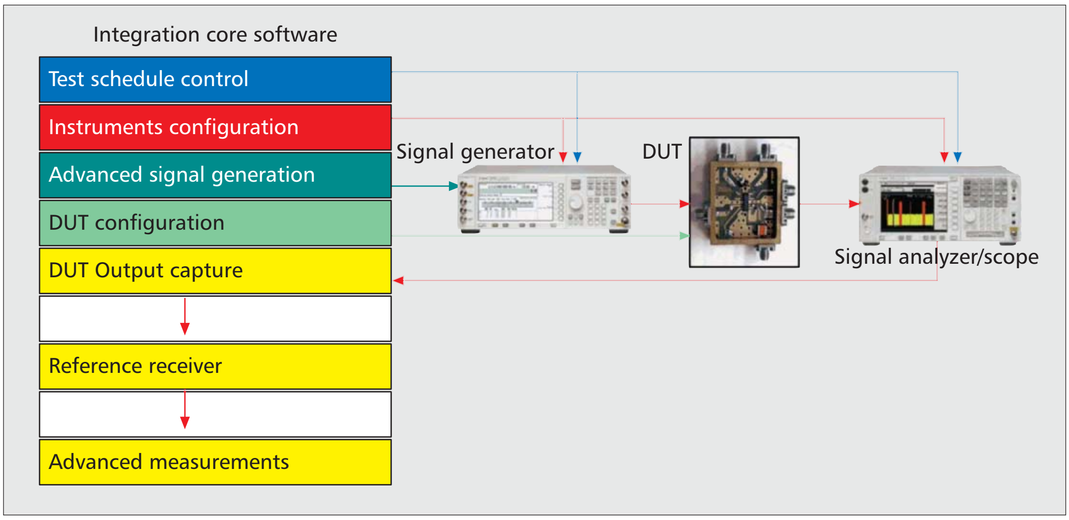


Figure 1. *The basic structure of an integrated test system.*

The ICS integrates all test system instruments together to provide test signals to the DUT, to capture DUT outputs and then synchronized signals. Without integration and synchronization, each instrument would function on its own, making it impossible to perform complex tests.

The test system's key functions are provided and managed by the ICS. These functions include the following.

Instrument and Test Sequence Configuration Manager

The ICS integrates all test system instruments together to provide test signals to the device under test (DUT), to capture DUT outputs and then synchronized signals. Without integration and synchronization, each instrument would function on its own, making it impossible to perform complex tests such as BER and BLER, and hard to perform advanced tests like sensitivity or throughput. More important, all fourth-generation (4G) wireless systems utilize multiple-input multiple-output (MIMO) technology and therefore require synchronized signals from multiple signal generators (including wideband signal generators), and synchronized data captures from multiple signal analyzers (including logic signal analyzers), and multichannel scopes.

To address this issue, the ICS Instruments Configuration Manager ensures that all instruments are set up properly prior to any tests or measurements. The ICS Test Sequence Configuration Manager then invokes the operation of all involved instruments and the DUT in their desired order. This is critical since in some tests (e.g., sensitivity or throughput), multiple measurements must be made under different test conditions and in a specific sequence.

Advanced Waveform Generation

The ICS Advanced Waveform Generation function generates waveforms based on international standards, complex waveforms (mixed-mode/multimodulation waveforms and waveforms using specific framed data or with special modulation data), and highly customized waveforms for commercial, military, and satellite communications. Generated waveforms are automatically downloaded to the VSG via specific instrument control protocols such as LAN, GPIB, and USB to facilitate test setup.

Waveform generation can also be sequenced to support flexible and/or more complex DUT testing, such as the creation of mixed-mode waveforms through combining existing single-mode signals. As an example, for testing an LTE-WiMAX dual-model base station, an existing LTE signal can be combined with a WiMAX signal to obtain dual-mode base station test signals.

DUT Configuration

The ICS's DUT Configuration configures the DUT to its proper test conditions and can also provide field programmable gate array (FPGA) programming capability for applications such as software defined radio (SDR) or cognitive radio. This latter capability is important since FPGAs are broadly used in today's hardware design, and advanced DUT configuration may require programming on-DUT FPGA. In this case, the ability to program FPGA tremendously simplifies the design of both SDR and cognitive radio products.

Acquire and Process DUT Output Data

In an integrated test system, DUT output is typically captured through a VSA by digitizing the DUT output signals and streaming the captured data back to the ICS for further analysis (e.g., advanced measurements like BER, BLER, and throughput). The ICS plays a critical role here since most commercial VSAs provide only transmission test measurements like spectrum, constellation, and error vector magnitude (EVM). For extensive receiver design tests, measurements such as BER, FER, throughput, and sensitivity are often required. These measurements typically go beyond the capability of modern VSAs and can only be performed by ICS.

The advanced measurements required for receiver component testing are facilitated using the ICS's own software receiver. It performs the timing and frequency synchronizations, channel estimation, demodulation, deframing, and decoding in order to troubleshoot and evaluate the performance of a receiver design. The software receiver in ICS can also be used to evaluate and fine tune a transmitter design to ensure it meets critical specifications. It can even easily be modified to test new standards that overlap existing standards. And, because this "golden" reference

LTE PHY models supported by SystemVue	
Available blocks	**Description**
Channel coding/decoding model set	Channel coding/decoding model set for both downlink and uplink channel codec include CRC, convolutional encoding/viterbi decoding, turbo encoding/turbo decoding, scrambler/descrambler, interleaver/de-interleaver, and HARQ models.
Modulation model set	Modulation model set includes mappers/de-mappers for QPSK, 16-QAM, 64-QAM, OFDM, SC-FDMA.
Multiplex models	The multiplex models provide OFDM/SCFDMA symbol multiplexing/demultiplexing, downlink/uplink framing/deframing for the downlink/uplink transceiver.
Receiver models	The receiver models are for constructing both downlink and uplink receivers in which timing/frequency synchronization and channel estimation are implemented.
Measurement models	The measurement models provide basic measurements, waveform, spectrum, constellation, and EVM. Also, receiver measurements include BER, BLER, FER, throughput, and sensitivity.

Table 1. *SystemVue's 3GPP LTE model set supports both FDD and TDD LTE. It includes more than 100 components and 10 test benches.*

covers both RF and baseband domains, it enables RF/digital co-simulation.

Note that to ensure reasonable measurement accuracy, sufficient memory depth of is required for instruments involved in the integrated test system. As an example, the Third Generation Partnership Project (3GPP) LTE base station conformance testing standard [1] requires that throughput be considered at the 95 percent level, corresponding to a 5 percent frame error level. A reasonable accuracy can therefore be achieved by setting the number of subframes to 200, which requires that the memory size be greater than is needed to hold 0.2 s of data.

Extend Measurement Capabilities for Instruments with ICS

Regular test systems with the VSG and VSA using manual processes only provide transmission test measurements such as waveform, spectrum, constellation, and EVM. The proposed integrated test solution not only provides transmission measurements, but also receiver measurements such as BER. In addition to preprogrammed measurements, such as BER, BLER, and throughput, more specific measurements such as adjacent channel selectivity and blocking can also be performed in the integrated test system.

Another advantage of an ICS is that customized measurements can be designed easily, which is desirable for new protocol or algorithm designs. Complex receiver measurements required by international standards can be derived from existing generic measurements. For example, reference level sensitivity for a Mobile WiMAX receiver can be derived from FER measurements, and reference level sensitivity for LTE can be derived from throughput measurements.

"Golden" Reference in ICS

Receiver component testing using instruments always requires a "golden" receiver. ICS can provide a software "golden" receiver to be embedded in the test system for troubleshooting and performance evaluation of receiver design. This software "golden" receiver can also be used to evaluate and fine tune transmitter design to ensure it meets critical specifications. Additionally, the software "golden" receiver can easily be modified or customized to test new standards that overlap existing standards, and it can cover not just baseband but also the RF domain, hence providing the availability of RF-digital co-simulation.

The Integrated Test System in Action

Let us now use the previously described integrated test system to measure the sensitivity of an LTE receiver that has been designed using the

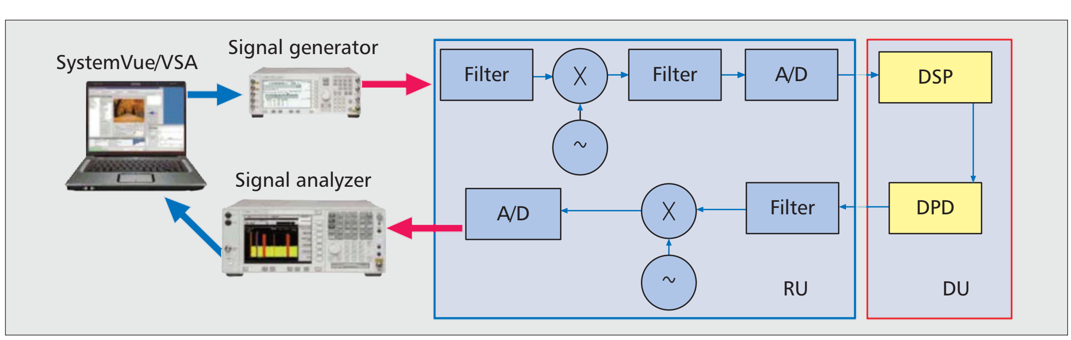


Figure 2. *The test setup for the LTE sensitivity measurement is shown here.*

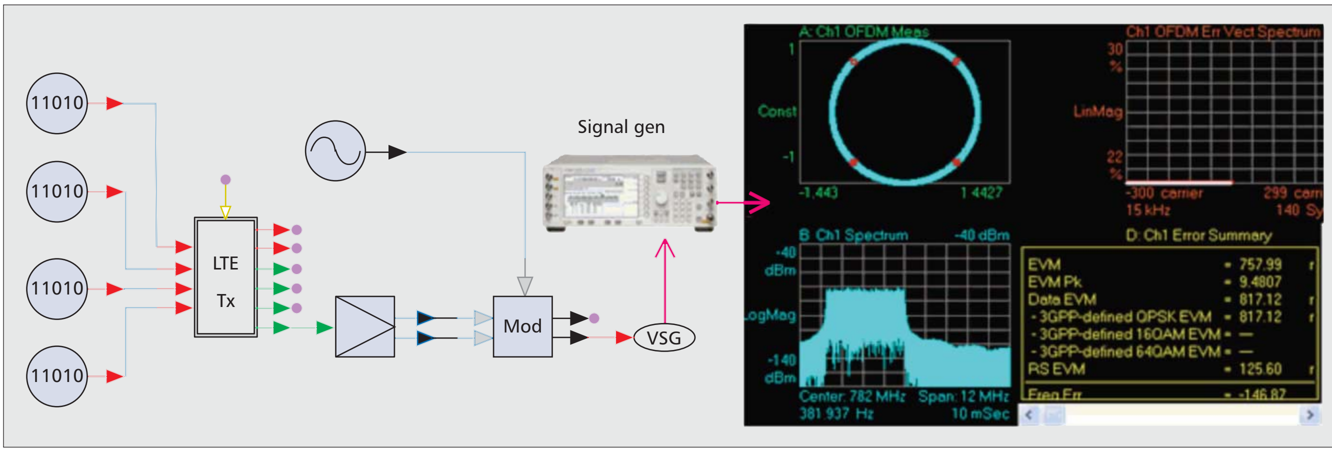


Figure 3. *LTE signal generation.*

latest LTE standard. Note that the LTE uplink receivers for both frequency-division duplexing (FDD) and time-division duplexing (TDD) are tested using this test system.

To cover the LTE specification on the signal generation side, the ARB sampling rate in the VSG must be at least 100 Mb/s for downlink (DL) and 50 Mb/s for uplink (UL). The VSG's maximum frequency must be at least 3 GHz. On the signal capture side, the VSA must have a bandwidth of at least 40 MHz and sufficient memory depth to hold at least 1 s of data (for generating system performance statistics).

Figure 2 shows the test system structure used to characterize LTE receiver performance. Here, SystemVue software from Agilent Technologies is used as the ICS, the Agilent ESG acts as the signal generator, and the Agilent PXA is used to capture DUT output signals. SystemVue provides all the key functionality required in an ICS solution.

During measurement, SystemVue generates the baseband LTE signal and sends it to the signal generator. The RF LTE signal from the signal generator provides the input to the DUT. In this case, the DUT consists of a receiver RF unit (RU), a digital unit (DU) with digital predistortion, and a common public radio interface (CPRI). The output of the DUT waveform is acquired by the signal analyzer and then streamed back to SystemVue, where it is further processed using the embedded golden reference LTE receiver. Following demodulation, deframing, and decoding in the receiver, the received bits are recovered and system throughput is measured.

A key part of the integrated test system and, in particular, an ICS-like SystemVue is that it provides built-in LTE physical layer (PHY) models for LTE design and verification (Table 1). These PHY models follow the 3GPP LTE Release 8 standard [1-4] and are required to test LTE systems. They are intended to provide a baseline to help designers determine the expected nominal or

Reference channel	aA1-1	A1-2	A1-3	A1-4	A1-5
Allocated resource blocks	6	15	25	3	9
DFT-OFDM Symbols per subframe	12	12	12	12	12
Modulation	QPSK	QPSK	QPSK	QPSK	QPSK
Code rate	1/3	1/3	1/3	1/3	1/3
Payload size (bits)	600	1544	2216	256	936
Transport block CRC (bits)	24	24	24	24	24
Code block CRC size (bits)	0	0	0	0	0
Number of code blocks — C	1	1	1	1	1
Coded block size including 12 bits trellis termination (bits)	1884	4716	6732	852	2892
Total number of bits per sub-frame	1728	4320	7200	864	2592
Total symbols per sub-frame	864	2160	3600	432	1296

Table 2. *LTE receiver sensitivity test parameter settings.*

IEEE Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

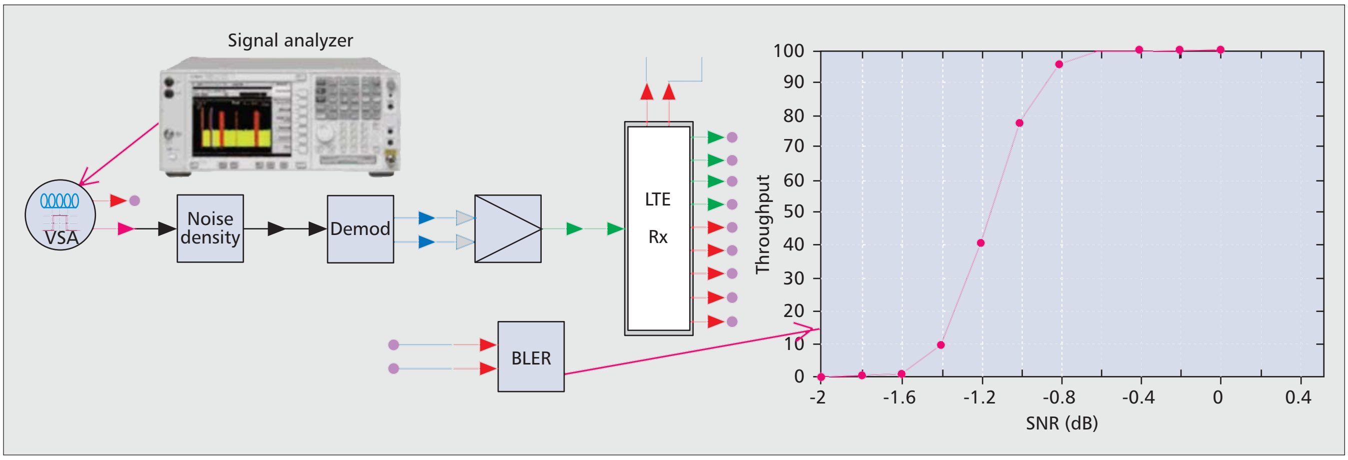


Figure 4. *LTE throughput test results.*

ideal system performance. They can also be used to evaluate degraded system performance due to system impairments that result from factors like non-ideal component performance.

Test Results

According to the 3GPP LTE test specification (TS 36.141), the reference sensitivity power level is defined as the minimum mean power received at the antenna connector at which a throughput requirement of 95 percent is met for a specified reference measurement channel. To set up the LTE receiver sensitivity test using the integrated test system, all system parameters must first be set to align with the LTE test specification (Table 2).

As an example, the fixed reference channel (test case A1-3) for a 10 MHz LTE system is set up. The following test procedure is then performed:

Step 1: LTE FDD or TDD baseband waveforms are generated and downloaded into the ESG signal generator using SystemVue. The RF output of the signal generator is then used to test the DUT. Figure 3 shows both the signal generation structure and the waveform at the signal generator RF output. Note that the LTE waveforms can be customized for any standard update by simply editing the signal generation design.

Step 2: The DUT output signal captured by the PXA signal analyzer is streamed back to SystemVue.

Step 3: SystemVue demodulates and decodes the single-input single-output (SISO) or MIMO signals and provides the receiver performance analysis, including throughput and BLER measurements. Curves for throughput and BLER vs. signal-to-noise ratio (SNR) can then be plotted. The sensitivity can also be measured by sweeping the receiver input power level to meet the 95 percent throughput level. For the fixed reference channel test case, the minimum input power level is less than –101 dBm.

Figure 4 depicts a curve on the throughput vs. SNR plot. Note that when the input power is set to –101 dBm and the SNR is 0 dB, the throughput is 96 percent. This result indicates that the LTE receiver works properly. However, when the power level decreases and the SNR drops to –1 dB, the resulting throughput is less than 95 percent. In this scenario, the LTE receiver would not operate properly; therefore, this operating condition should be avoided.

Conclusion

Today, wireless communication standards are evolving faster than ever before. For the engineer charged with designing wireless communication products, this fast evolution translates into an ever changing test challenge. Dealing with this challenge under the increasing time-to-market pressure demands an integrated, dynamic test system. And since it is critical to perform early and continuous verification of PHY algorithms and prototype designs during a product's development phase, the test system must enable comparisons with "ideal" references. The ICS-based reconfigurable, integrated test system offers the ideal solution to tackling this challenge. It not only provides a "golden" reference, but also enables automatic configuration of test instruments. Just as critical, it presents today's engineers with a viable means of improving their new product's time to market.

References

[1] 3GPP TS 36.141 v8.50, "Base Station Conformance Testing."

[2] 3GPP TS 36.211 v8.9.0, "Physical Channels and Modulation," Dec. 2009.

[3] 3GPP TS 36.101 v8.6.0 "User Equipment (UE) Radio Transmission and Reception," Sept. 2009.

[4] 3GPP TS 36.104 v8.7.0, "Base Station (BS) Radio Transmission and Reception," Sept. 2009.

Biographies

DINGQING LU [M'90, SM'91] (dingquing_lu@agilent.com) has been with Agilent Technologies/Hewlett Packard Company since 1989 and is a scientist with Agilent EEsof EDA. From 1981 to 1986 he was with the University of Sichuan as lecturer and assistant professor. He was a research associate in the Department of Electrical Engineering at the University of California at Los Angeles (UCLA) from 1986 to 1989. He has published 20 papers in IEEE transactions, journals, and conference proceedings, and holds a U.S. patent. His research interests include system modeling, simulation, and measurement techniques.

ZHENGRONG ZHOU has been with Agilent Technologies/Hewlett Packard Company since 1999 and is an R &D engineer with Agilent EEsof EDA. His work at Agilent focuses on test and measurement algorithms, RF calibration algorithms, and EDA software. He has acquired two U.S. patents.

TOPICS IN RADIO COMMUNICATIONS

An Introduction to Millimeter-Wave Mobile Broadband Systems

Zhouyue Pi and Farooq Khan, Samsung Electronics

ABSTRACT

Almost all mobile communication systems today use spectrum in the range of 300 MHz–3 GHz. In this article, we reason why the wireless community should start looking at the 3–300 GHz spectrum for mobile broadband applications. We discuss propagation and device technology challenges associated with this band as well as its unique advantages for mobile communication. We introduce a millimeter-wave mobile broadband (MMB) system as a candidate next-generation mobile communication system. We demonstrate the feasibility for MMB to achieve gigabit-per-second data rates at a distance up to 1 km in an urban mobile environment. A few key concepts in MMB network architecture such as the MMB base station grid, MMB inter- BS backhaul link, and a hybrid MMB + 4G system are described. We also discuss beamforming techniques and the frame structure of the MMB air interface.

INTRODUCTION

Mobile communication has been one of the most successful technology innovations in modern history. The combination of technology breakthroughs and attractive value proposition has made mobile communication an indispensable part of life for 5 billion people. Due to the increasing popularity of smart phones and other mobile data devices such as netbooks and ebook readers, mobile data traffic is experiencing unprecedented growth. Some predictions indicate that mobile data will grow at 108 percent compound annual growth rate (CAGR) [1] with over a thousandfold increase over the next 10 years. In order to meet this exponential growth, improvements in air interface capacity and allocation of new spectrum are of paramount importance.

The current fourth-generation (4G) systems including LTE and Mobile WiMAX already use advanced technologies such as orthogonal frequency-division multiplexing (OFDM), multiple-input multiple-output (MIMO), multi-user diversity, link adaptation, turbo code, and hybrid automatic repeat request (HARQ) in order to achieve spectral efficiencies close to theoretical limits in terms of bits per second per Hertz per cell [2]. With limited room for further spectral efficiency improvement, another possibility to increase capacity per geographic area is to deploy many smaller cells such as femtocells and heterogeneous networks. However, because capacity can only scale linearly with the number of cells, small cells alone will not be able to meet the capacity required to accommodate orders of magnitude increases in mobile data traffic.

As the mobile data demand grows, the sub-3 GHz spectrum is becoming increasingly crowded. On the other hand, a vast amount of spectrum in the 3–300 GHz range remains underutilized. The 3–30 GHz spectrum is generally referred to as the super high frequency (SHF) band, while 30–300 GHz is referred to as the extremely high frequency (EHF) or millimeter-wave band. Since radio waves in the SHF and EHF bands share similar propagation characteristics, we refer to 3–300 GHz spectrum collectively as millimeter-wave bands with wavelengths ranging from 1 to 100 mm.

Millimeter-wave communication systems that can achieve multigigabit data rates at a distance of up to a few kilometers already exist for point-to-point communication. However, the component electronics used in these systems, including power amplifiers, low noise amplifiers, mixers, and antennas, are too big in size and consume too much power to be applicable in mobile communication. The availability of the 60 GHz band as unlicensed spectrum has spurred interest in gigabit-per-second short-range wireless communication. Several industrial standards have been developed, such as WirelessHD technology, ECMA-387, IEEE 802.15.3c, and IEEE 802.11ad. Integrated circuit (IC)-based transceivers are also available for some of these technologies. Much of the engineering efforts have been invested in developing more power-efficient 60 GHz RFICs [3]. Many of these technologies can be transferred to RFIC design for other millimeter-wave bands.

In this article, we explore the 3–300 GHz spectrum and describe a millimeter-wave mobile broadband (MMB) system that utilizes this vast spectrum for mobile communication. We describe the millimeter-wave spectrumand its propagation characteristics. We then discuss the network architecture, followed by the air interface design of the MMB system. After that, we conclude the article with a summary and brief discussion of future work.

0163-6804/11/$25.00

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

The portion of the RF spectrum above 3 GHz has been largely unexploited for commercial wireless applications. More recently there has been some interest in exploring this spectrum for short-range and fixed wireless communications.

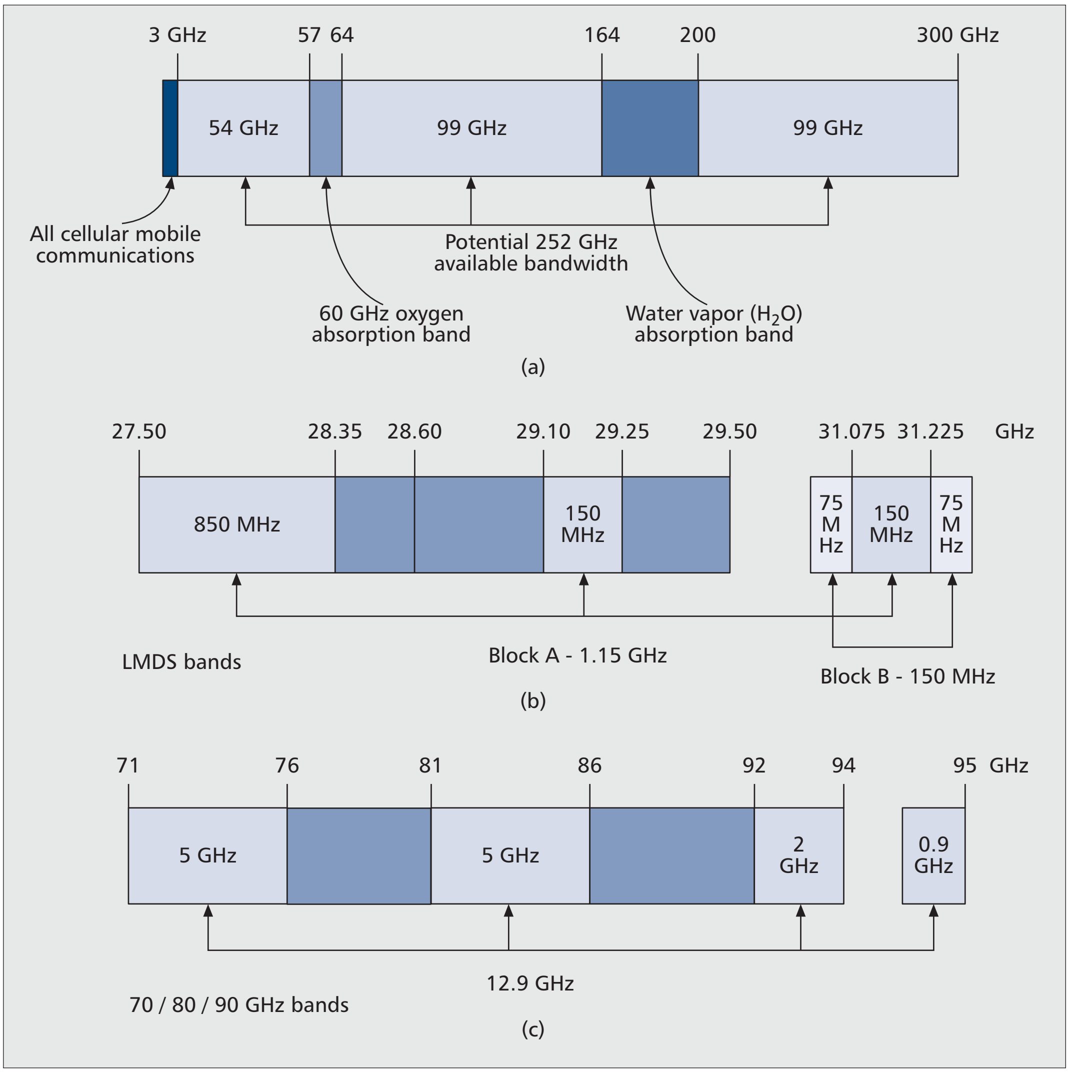


Figure 1. *Millimeter-wave spectrum.*

Millimeter Wave Spectrum

Unleashing the 3–300 GHz Spectrum

Almost all commercial radio communications including AM/FM radio, high-definition TV, cellular, satellite communication, GPS, and Wi-Fi have been contained in a narrow band of the RF spectrum in 300 MHz–3 GHz. This band is generally referred to as the *sweet spot* due to its favorable propagation characteristics for commercial wireless applications. The portion of the RF spectrum above 3 GHz, however, has been largely unexploited for commercial wireless applications. More recently there has been some interest in exploring this spectrum for short-range and fixed wireless communications. For example, unlicensed use of ultra-wideband (UWB) in the range of 3.1–10.6 GHz frequencies has been proposed to enable high data rate connectivity in personal area networks. The use of the 57–64 GHz oxygen absorption band is also being promoted to provide multigigabit data rates for short-range connectivity and wireless local area networks. Additionally, local multipoint distribution service (LMDS) operating on frequencies from 28 to 30 GHz was conceived as a broadband, fixed wireless, point-to-multipoint technology for utilization in the last mile.

Within the 3–300 GHz spectrum, up to 252 GHz can potentially be suitable for mobile broadband as depicted in Fig. 1a. Millimeter waves are absorbed by oxygen and water vapor in the atmosphere. The frequencies in the 57–64 GHz oxygen absorption band can experience attenuation of about 15 dB/km as the oxygen molecule (O_2) absorbs electromagnetic energy at around 60 GHz. The absorption rate by water vapor (H_2O) depends on the amount of water vapor and can be up to tens of dBs in the range of 164–200 GHz [4]. We exclude these bands for mobile broadband applications as the transmission range in these bands will be limited. With a reasonable assumption that 40 percent of the remaining spectrum can be made available over time, millimeter-wave mobile broadband (MMB) opens the door for a possible 100 GHz new spectrum for mobile communication — more than 200 times the spectrum currently allocated for this purpose below 3 GHz.

LMDS and 70/80/90 GHz Bands

LMDS was standardized by the IEEE 802 LAN/MAN Standards Committee through the efforts of the IEEE 802.16.1 Task Group ("Air Interface for Fixed Broadband Wireless Access

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Systems" for 10–66 GHz). LMDS uses a cellular infrastructure, with multiple base stations supporting point-to-multipoint communication to small customer transceivers. The Federal Communications Commission (FCC) auctioned two LMDS licenses per market (basic trading areas). The A license includes a total of 1.15 GHz bandwidth, and consists of the 27.5–28.35 GHz, 29.1–29.25 GHz, and 31.075–31.225 GHz bands. The B license is 150 MHz wide, covering the 31.0–31.075 GHz and 31.225–31.3 GHz bands as depicted in Fig. 1b.

On October 16, 2003 the FCC announced that the 71–76 GHz, 81–86 GHz, and 92–95 GHz frequency bands collectively referred as E-band had become available to ultra-high-speed data communication including point-to-point wireless local area networks, mobile backhaul, and broadband Internet access. A total of 12.9 GHz bandwidth is available in the E-band as shown in Fig. 1c, with a narrow 100 MHz exclusion band at 94.0–94.1 GHz. Highly directional "pencilbeam" signal characteristics in E-band permit systems in these bands to be engineered in close proximity to one another without causing interference. Therefore, the FCC and regulators in other countries have introduced "light licensing" schemes for managing this band. These innovative licenses retain the benefits of full interference protection and can be applied for in minutes over the Internet at costs of a few tens of dollars per year.

We note that regulations would need to be changed with provisioning to support mobility and higher transmit powers to enable mobile broadband communications in LMDS, 70/80/90 GHz, and possibly other millimeter-wave bands.

Millimeter-Wave Propagation

Free-Space Propagation

Transmission loss of millimeter wave is accounted for principally by free space loss. A general misconception among wireless engineers is that free-space propagation loss depends on frequency, so higher frequencies propagate less well than lower frequencies. The reason for this misconception is the underlying assumption often used in radio engineering textbooks that the path loss is calculated at a specific frequency between two isotropic antennas or $\lambda/2$ dipoles, whose effective aperture area increases with the wavelength (decreases with carrier frequency). An antenna with a larger aperture has larger gain than a smaller one as it captures more energy from a passing radio wave. However, with shorter wavelengths more antennas can be packed into the same area. For the same antenna aperture areas, shorter wavelengths (higher frequencies) should not have any inherent disadvantage compared to longer wavelengths (lower frequencies) in terms of free space loss [5]. In addition, large numbers of antennas enable transmitter and receiver beamforming with high gains. For example, a beam at 80 GHz will have about 30 dB more gain (narrower beam) than a beam at 2.4 GHz if the antenna areas are kept constant.

Penetration and Other Losses

For 3–300 GHz frequencies, atmosphere gaseous losses and precipitation attenuation are typically less than a few dB per kilometer [4], excluding the oxygen and water absorption bands. The loss due to reflection and diffraction depends greatly on the material and the surface. Although reflection and diffraction reduce the range of millimeter-wave, it also facilitates non-line-of-sight (NLOS) communication.

		Attenuation (dB)		
Material	Thickness (cm)	< 3 GHz [6, 8]	40 GHz [7]	60 GHz [6]
Drywall	2.5	5.4	–	6.0
Office whiteboard	1.9	0.5	–	9.6
Clear glass	0.3/0.4	6.4	2.5	3.6
Mesh glass	0.3	7.7	–	10.2
Chipwood	1.6	–	.6	–
Wood	0.7	5.4	3.5	–
Plasterboard	1.5	–	2.9	–
Mortar	10	–	160	–
Brick wall	10	–	t178	–
Concrete	10	17.7	175	–

Table 1. *Attenuations for different materials.*

While signals at lower frequencies can penetrate more easily through buildings, millimeter-wave signals do not penetrate most solid materials very well. In Table 1, we provide attenuation values for common materials [6, 7]. High levels of attenuation for certain building materials (e.g., brick and concrete) may keep millimeter waves transmitted from outdoor base stations confined to streets and other outdoor structures, although some signals might reach inside the buildings through glass windows and wood doors. The indoor coverage in this case can be provided by other means such as indoor millimeter-wave femtocell or Wi-Fi solutions. It should be noted that next-generation Wi-Fi technology using 60 GHz millimeter waves is already being developed in IEEE 802.11ad [9].

Foliage losses for millimeter waves are significant and can be a limiting impairment for propagation in some cases. An empirical formula has been developed in [6] to calculate the propagation through foliage. In Fig. 2a, we plot penetration losses for foliage depth of 5, 10, 20, and 40 m. We note, for example, that at 80 GHz frequency and 10 m foliage penetration, the loss can be about 23.5 dB, which is about 15 dB higher than the loss at 3 GHz frequency.

Millimeter-wave transmissions can experience significant attenuations in the presence of heavy rain. Raindrops are roughly the same size as the radio wavelengths (millimeters) and therefore cause scattering of the radio signal. The attenuation (dB per kilometer) can be calculated from rain rates (millimeters per hour) [10], and the curves are plotted in Fig. 2b. For example, light

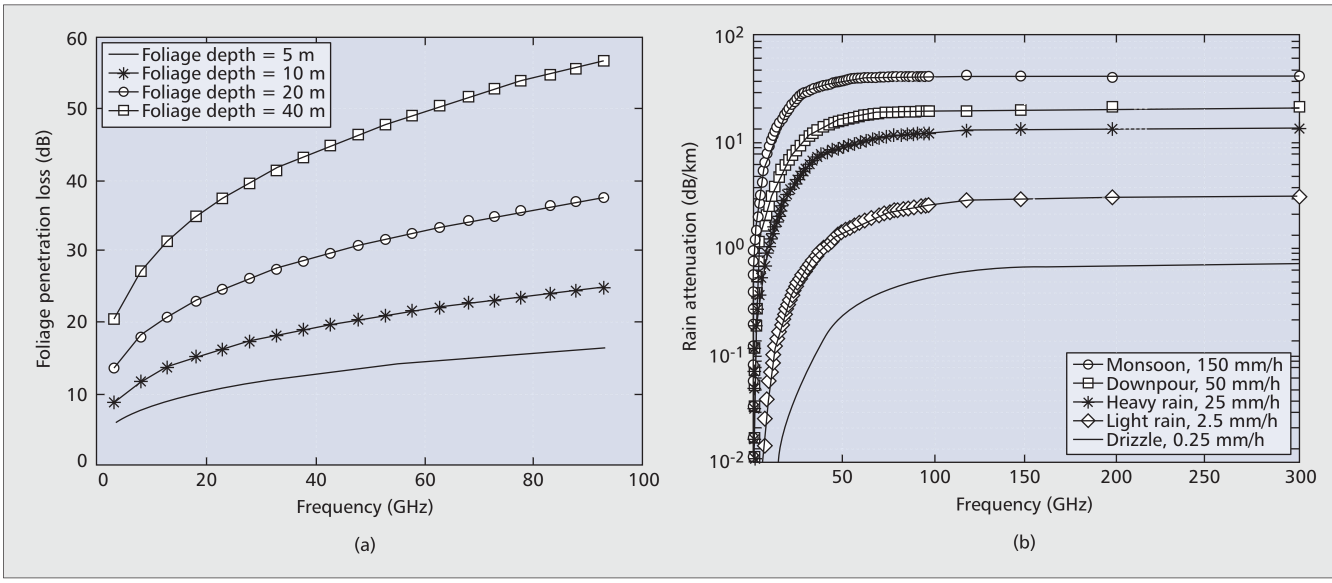


Figure 2. *Millimeter-wave propagation characteristics: a) foliage penetration loss; b) rain attenuation.*

rain at a rate of 2.5 mm/h yields just over 1 dB/km attenuation, while severe rain such as a monsoon at a rate of 150 mm/h can jeopardize communications with up to tens of dB loss per kilometer at millimeter-wave frequencies. Fortunately, the most intense rain tends to fall in selected countries of the world, and happen in short bursts and small clusters. A mechanism such as supporting emergency communications over cellular bands when millimeter-wave communications are disrupted by heavy rains should be considered as part of the MMB system design.

Doppler and Multipath

The Doppler of a wireless channel depends on the carrier frequency and mobility. Assuming a rich scattering environment and omnidirectional antennas, the maximum Doppler shift for carrier frequency of 3–60 GHz with mobility of 3–350 km/h ranges from 10 Hz to 20 kHz. The Doppler shift values of incoming waves on different angles at the receiver are different, resulting in a phenomenon called Doppler spread. In the case of MMB, the narrow beams at the transmitter and receiver will significantly reduce angular spread of the incoming waves, which in turn reduces the Doppler spread. In addition, as the incoming waves are concentrated in a certain direction, there will be a non-zero bias in the Doppler spectrum, which will be largely compensated by the automatic frequency control (AFC) loop in the receiver. Therefore, the time-domain variation of an MMB channel is likely to be much less than that observed by omnidirectional antennas in a rich scattering environment.

With narrow transmitter and receiver beams, the multipath components of millimeter waves are limited. Studies show that the root mean square (RMS) of the power delay profile (PDP) of a millimeter-wave channel in an urban environment is 1–10 ns, and the coherent bandwidth of the channel is around 10–100 MHz [11]. However, it is noted that the transmitter and receiver antenna gains used in these studies are higher than those used in MMB. Therefore, it is possible that in an MMB system a longer path can be observed and the coherence bandwidth is smaller than those reported in these studies.

MMB Network Architecture

A Standalone MMB Network

An MMB network consists of multiple MMB base stations that cover a geographic area. In order to ensure good coverage, MMB base stations need to be deployed with higher density than macrocellular base stations. In general, roughly the same site-to-site distance as microcell or picocell deployment in an urban environment is recommended. An example MMB network is shown in Fig. 3.

The transmission and/or reception in an MMB system are based on narrow beams, which suppress the interference from neighboring MMB base stations and extend the range of an MMB link. This allows significant overlap of coverage among neighboring base stations. Unlike cellular systems that partition the geographic area into cells with each cell served by one or a few base stations, the MMB base stations form a grid with a large number of nodes to which an MMB mobile station can attach. For example, with a site-to-site distance of 500 m and a range of 1 km for an MMB link, an MMB mobile station can access up to 14 MMB base stations on the grid, as shown in Fig. 3a. The MMB base station grid eliminates the problem of poor link quality at the cell edge that is inherent in cellular system and enables high-quality equal grade of service (EGOS) regardless of the location of a mobile.

With the high density of MMB base stations, the cost to connect every MMB base station via a wired network can be significant. One solution to mitigate the cost (and expedite the deployment) is to allow some MMB base stations to connect to the backhaul via other MMB base stations. Due to large beamforming gains, the MMB inter-BS backhaul link can be deployed in the same frequency as the MMB access link —

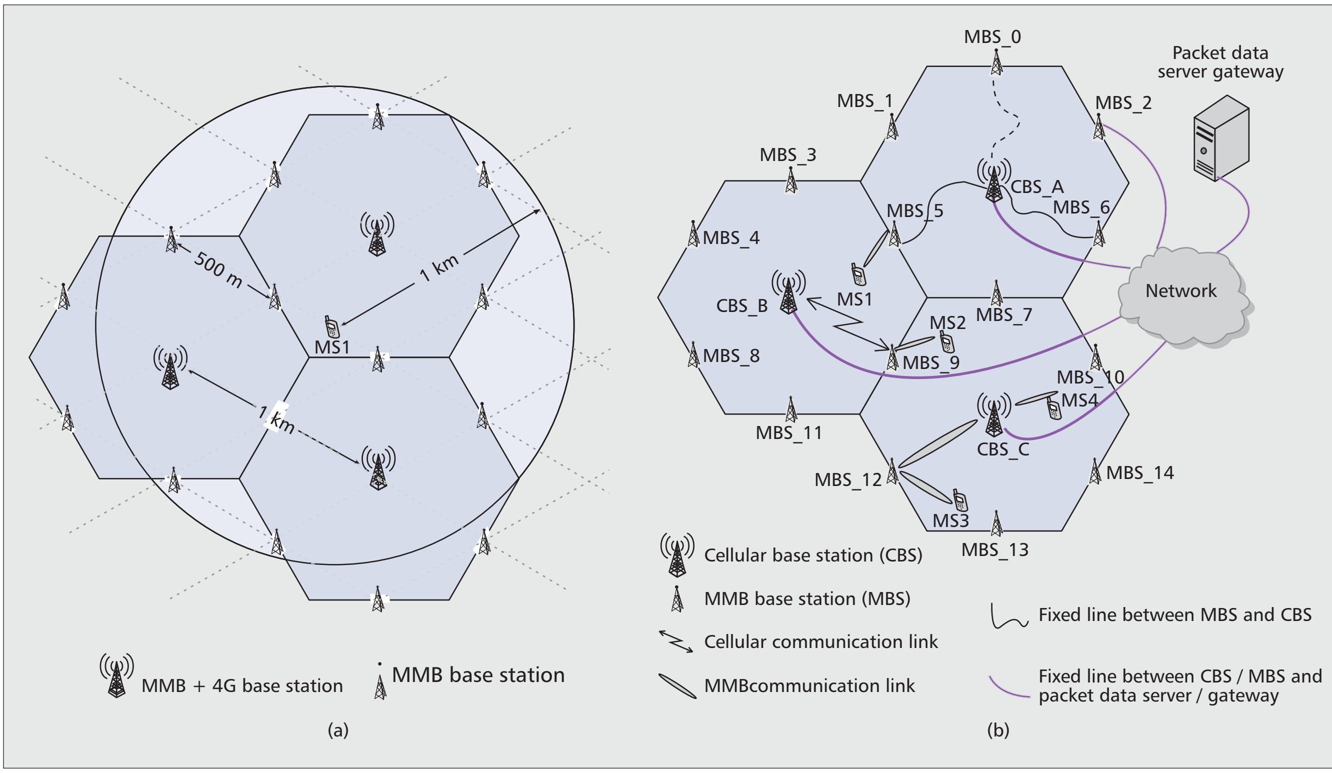


Figure 3. *MMB network: a) architecture; b) hybrid MMG + 4G system.*

the downlink to and uplink from an MMB mobile station — without causing much interference. This greatly increases the deployment flexibility of MMB and allows MMB to achieve higher-density deployment than femtocells or heterogeneous networks deployed in sub-3 GHz spectrum.

Another challenge with millimeter-wave is the low efficiency of RF devices such as power amplifiers and multi-antenna arrays with current technology. A solution to avoid multi-antenna arrays at the MMB base station is to use fixed beams or sectors with horn antennas. Horn antennas can provide similar gains and beam widths as sector antennas in current cellular systems in a cost-effective manner [12]. The mobile station receiver still needs to use a multi-antenna array to form a beamforming pattern toward the base station. As the mobile station moves around, beamforming weights can be adjusted so that the beam is always pointing toward the base station.

Hybrid MMB + 4G System

In the early deployment of MMB, there may be coverage holes in areas where the MMB base station density is low. However, it is expected that 4G systems will have good coverage and reliability when MMB systems start to deploy. A hybrid MMB + 4G system can improve coverage and ensure seamless user experience in mobile applications. In a hybrid MMB + 4G system, system information, control channel, and feedback are transmitted in the 4G system, making the entire millimeter-wave spectrum available for data communication. One example of a hybrid MMB + 4G system is shown in Fig. 3b. Compared with millimeter waves, the radio waves at < 3 GHz frequencies can better penetrate obstacles and are less sensitive to non-line-of-sight (NLOS) communication link or other impairments such as absorption by foliage, rain, and other particles in the air. Therefore, it is advantageous to transmit important control channels and signals via cellular radio frequencies, while utilizing the millimeter waves for high data rate communication.

MMB Air Interface Design

Beamforming

Beamforming is a signal processing technique used for directional signal transmission or reception. Spatial selectivity/directionality is achieved by using adaptive transmit/receive beam patterns. When transmitting, a beamformer controls the phase and relative amplitude of the signal at each transmitter antenna to create a pattern of constructive and destructive interference in the wavefront. When receiving, signals from different receiver antennas are combined in such a way that the expected pattern of radiation is preferentially observed.

Beamforming is a key enabling technology of MMB. For MMB transceivers, the small size ($\lambda/2$ dipoles) and separation (also around $\lambda/2$) of millimeter-wave antennas allow a large number of antennas and thus achieve high beamforming gain in a relative small area (e.g., tens of antennas per square centimeter area at 80 GHz carrier frequency). Additionally, with a large number of antennas and high-gain (and thus narrow) beams, antenna technologies such as spatial-division multiple access (SDMA) can be implemented readily.

Beamforming can be achieved in digital base-

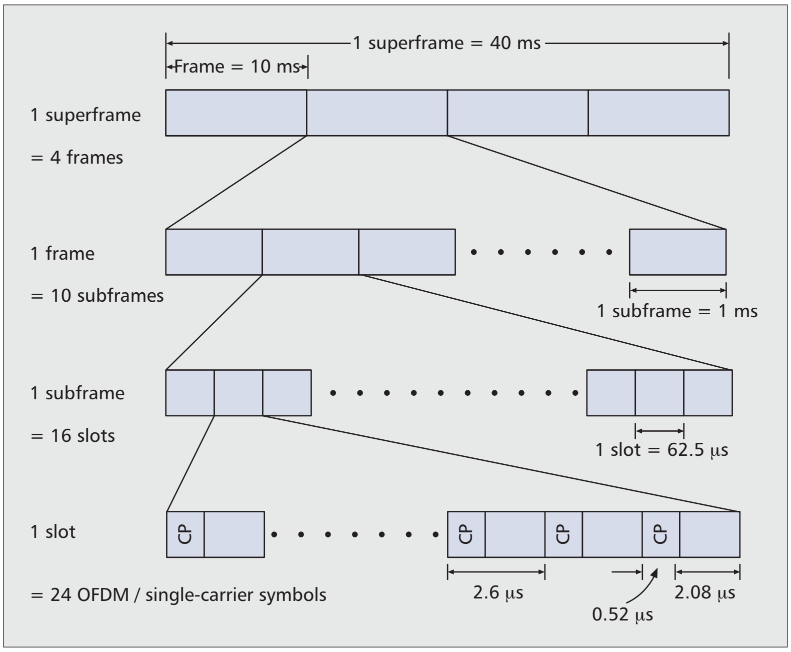


Figure 4. *MMB frame structure (1/4 CP).*

band, analog baseband, or RF front-end. With digital beamforming and multiple RF chains, it is possible to transmit multiple streams of data simultaneously, thus enabling SDMA or MIMO operation. However, the cost of implementing one RF chain per antenna can be prohibitive, especially given the large number of antennas in MMB. With analog baseband beamforming or RF beamforming, one or a few RF chains can be used. In that case, the number of data streams that can be transmitted is limited by the number of RF chains. These approaches require fewer RF components and are typically chosen for low-cost/low-power solutions.

Transmit beamforming is generally more challenging, requiring either antenna weights feedback from the receiver or antenna calibrations. Moreover, due to low efficiency of millimeter-wave power amplifiers with the current technology, battery power consumption is another issue for mobile station transmitter beamforming. To reduce the cost and complexity of mobile stations, a phased approach where initial deployments are hybrid MMB + 4G systems with downlink-only transmission in the millimeter-wave band can be considered. This removes the requirement for mobile stations to transmit in the millimeter-wave band.

Frame Structure

OFDM and single-carrier FDM were chosen to be the multiplexing schemes of 4G systems due to a variety of reasons (e.g., flexibility in support multiple bandwidths, simpler equalizer and MIMO receiver, and ability to support efficient multiple access, etc.). In MMB, we also use OFDM and single-carrier waveform for largely the same reasons.

One configuration of MMB frame structure is shown in Fig. 4. The basic transmission time interval (TTI) of MMB is a slot, whose duration is 62.5 us. In order to facilitate hybrid MMB + 4G operation, the durations of subframe, frame, and superframe are chosen to be 1 ms, 10 ms, and 40 ms, the same as those of LTE systems.

The OFDM/single-carrier numerology is carefully chosen according to a number of engineering considerations. For example, the sampling rate is chosen to be a multiple of 30.72 MHz, a popular frequency at which clocks with reasonable accuracy are readily available at low cost. The cyclic prefix (CP) is chosen to be 520 ns, which gives sufficient margin in accommodating the longest path, different deployment scenarios, and the potential increase of delay spread in the case of small antenna arrays (e.g., smart phones with small form factors) or wider beams (e.g., control channel transmissions). The subcarrier spacing is chosen to be 480 kHz, small enough to stay within the coherent bandwidth of most multipath channels expected in MMB. The corresponding OFDM symbol length (without CP) is 2.08 μs, resulting in 20 percent CP overhead. The subcarrier spacing is also wide enough to keep the size of fast/inverse fast Fourtier transform (FFT/IFFT) small (2048 points for 1 GHz system bandwidth) and accommodate inaccuracies of low-cost clocks. For example, with a carrier frequency of 28 GHz and a clock with 20 ppm accuracy, the clock drift is at most 560 kHz, less than 2 times the subcarrier spacing. This enables simple design of synchronization and system acquisition.

Additionally, MMB also supports transmission with single-carrier waveform. Single-carrier waveform has lower peak-to-average-power ratio (PAPR) than OFDM. As the solid-state devices today only have a limited amount of output power rating (< 1 W) in 60–100 GHz frequencies, it is beneficial to use single-carrier waveform to maximize the output power so that MMB can achieve the longest range possible. A lower PAPR also allows the receiver to use a low-resolution analog-to-digital converter (ADC). For single-carrier transmissions with binary phase shift keying (BPSK) or quaternary PSK (QPSK), an ADC with 2–4 bits would suffice, which greatly reduces the power consumption of the MMB receiver.

Link Budget

The key factors that determine the downlink link budget of an MMB system are the base station transmission power, transmitter and receiver beamforming gains, and path loss.

Table 2 shows the link budget for four different MMB systems. A 20 dB margin is assumed to account for cable loss and losses due to penetration, reflection, or diffraction. A noise figure of 10 dB and an implementation loss of 5 dB are assumed at the receiver. As shown in Table 2, with 35 dBm transmission power, 1 GHz system bandwidth, 28 GHz carrier frequency, and realistic assumptions of transmitter and receiver antenna gains (case 1), more than 2 Gb/s can be achieved at 1 km distance.

Conclusion

Millimeter-wave spectrum with frequencies in the range of 3–300 GHz can potentially provide the bandwidth required for mobile broadband applications for the next few decades and beyond.

In this article, we have analyzed the suitability of different millimeter-wave frequencies for mobile communication. We have discussed the propagation characteristics of millimeter waves, including the propagation and penetration losses, Doppler, and multipath. Due to the narrow beam width of MMB transmissions, the interference among MMB base stations is a lot smaller than traditional cellular systems, and the coverage of neighboring base stations significantly overlap. As a result, the MMB base stations form a grid that can provide communication with good link quality regardless of the mobile station's location within the coverage of the grid. The inter-BS backhaul link can be used to mitigate the cost of backhauling (and to expedite deployment). It is also possible to operate a hybrid MMB + 4G system such that existing 4G systems can be leveraged for reliable system information broadcast, packet data control, and feedback of MMB systems.

In order to operate in an urban mobile environment while keeping a low overhead, we chose the MMB subcarrier spacing to be 480 kHz and the CP to be 520 ns. We also designed the frame structure to facilitate hybrid MMB + 4G operation. In the link budget analysis, we show that a 2 Gb/s data rate is achievable at 1 km distance with millimeter waves in an urban mobile environment.

MMB link budget analysis	Case 1	Case 2	Case 3	Case 4
TX power (dBm)	35.00	35.00	25.00	25.00
TX antenna gain (dBi)	30.00	30.00	30.00	30.00
Carrier frequency (GHz)	28.00	72.00	28.00	72.00
Distance (km)	1.00	1.00	0.50	0.50
Propagation loss (dB)	121.34	129.55	115.32	123.53
Other losses	20.00	20.00	20.00	20.00
RX antenna gain (dBi)	15.00	15.00	15.00	15.00
Received power (dBm)	–61.34	–69.55	–65.32	–73.53
Bandwidth (GHz)	1.00	1.00	1.00	1.00
Thermal PSD (dBm/Hz)	–174.00	–174.00	–174.00	–174.00
Noise figure (dB)	10.00	10.00	10.00	10.00
Thermal noise (dBm)	–74.00	–74.00	–74.00	–74.00
SNR (dB)	12.66	4.45	8.68	0.47
Implementation loss (dB)	5.00	5.00	5.00	5.00
Data rate (Gb/s)	2.77	0.91	1.74	0.4

Table 2. *MMB link budget.*

Acknowledgment

The authors would like to thank their colleagues at Samsung for valuable discussions and feedback, and the reviewers for their constructive comments.

References

[1] White_paper_c11-481360, "Cisco Visual Networking Index: Forecast and Methodology," June 2010.
[2] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance*, Cambridge Univ. Press, 2009.
[3] C. H. Doann, "Millimeterwave CMOS Design," *IEEE J. Solid-States Circuits*, Jan. 2005.
[4] FCC, "Millimeter wave Propagation: Spectrum Management Implications," Office of Eng. and Tech., Bulletin no. 70, July, 1997.
[5] D. M. Pozar, *Microwave Engineering*, 3rd ed., Wiley, 2005.
[6] C. R. Anderson and T. S. Rappaport, "In-Building Wideband Partition Loss Measurements at 2.5 and 60 GHz," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, May 2004.
[7] A. Alejos, M. G. Sanchez, and I. Cuinas, "Measurement and Analysis of Propagation Mechanisms at 40 GHz: Viability of Site Shielding Forced By Obstacles," *IEEE Trans. Vehic. Tech.*, vol. 57, no. 6, Nov. 2008.
[8] K. C. Allen *et al.*, "Building Penetration Loss Measurements at 900 MHz, 11.4 GHz, and 28.8 GHz," NTIA rep. 94-306, May 1994.
[9] E. Perahia *et al.*, "IEEE 802.11ad: Defining the Next Generation Multi-Gb/s Wi-Fi," *2010 7th IEEE Consumer Commun. and Net. Conf.)*, 9–12 Jan. 2010, pp. 1–5.
[10] ITU-R P.838-3, "Specific Model for Rain for Use in Prediction Methods," 2005.
[11] H. Xu *et al.*, "Measurements and Models for 38-GHz Point-to-Multipoint Radiowave Propagation," *IEEE JSAC*, vol. 18, no. 3, Mar. 2000.
[12] C. A. Balanis, *Antenna Theory: Analysis and Design*, 3rd ed., Wiley, 2005.

Biographies

ZHOUYUE PI (zpi@sta.samsung.com) is a director at the Samsung R&D Center in Dallas, Texas, where he leads 4G standardization and development efforts. Before joining Samsung, he worked at Nokia Research Center in Dallas and San Diego, California, on 3G standardization and modem development. He has published more than 15 research papers and more than 80 patents and patent applications. He received his B.E. degree in automation from Tsinghua University (with honors), his M.S. degree in electrical engineering from Ohio State University, and his M.B.A. degree from Cornell University (with distinction).

FAOOQ KHARN (fkhan@sta.samsung.com) is a senior director at the Samsung R&D Center in Dallas, Texas, where he manages research in the areas of wireless communications, RF systems, multimedia, computing, and smart energy. Previously, he held research positions with Bell Laboratories in New Jersey and Ericsson Research in Sweden. He has authored more than 35 research papers and holds over 50 U.S. patents. He also authored a book, *LTE for 4G Mobile Broadband — Air Interface Technologies and Performance*. He holds an M.S. degree in electrical engineering from Ecole Supérieure d'Electricité, Paris, France, and a Ph.D. degree in computer science from Université de Versailles, France.

TOPICS IN RADIO COMMUNICATIONS

Spectrally Agile Multicarrier Waveforms for Opportunistic Wireless Access

Hanna Bogucka, Alexander M. Wyglinski, Srikanth Pagadarai, and Adrian Kliks

ABSTRACT

Multicarrier modulation has been employed in numerous modern wireless communications standards due to its ability in providing high data rates while simultaneously counteracting the effects of intersymbol interference due to multipath fading channels. In the recent years, multicarrier modulation is being investigated as a possible candidate data transmission scheme for achieving spectrally agile wireless access in scenarios where unlicensed users temporarily "borrow" unoccupied licensed frequency bands while minimizing interference with spectrally adjacent signals, i.e., opportunistic wireless access. Specifically, the *divide-and-conquer* data transmission approach employed by multicarrier modulation makes it an attractive option for realizing wireless communication systems that do not require a single continuous transmission frequency band. In this article, we present several multicarrier transmission solutions designed for efficiently achieving opportunistic wireless access. In particular, we shall investigate two types of spectrally-agile multicarrier modulation schemes, namely, non-contiguous orthogonal frequency-division multiplexing and non-contiguous non-orthogonal frequency-division multiplexing. The viability of these two techniques employed within an opportunistic wireless access scenario is assessed using actual spectrum measurement data, and a comparative study in terms of out-of-band interference mitigation as well as implementation complexity is provided.

INTRODUCTION

Modern society's demand for access to wireless spectral bandwidth has been rapidly increasing over the past several decades due to the deployment of new services, as well as the steadily growing number of wireless users. Simultaneously, measurement studies have shown that much of the licensed spectrum is underutilized across the temporal, frequency, and spatial domains [1]. Thus, to continue providing sufficient spectral bandwidth for satisfying current and future wireless access needs, both spectrum policy makers and communication technologists have proposed an innovative approach to access the wireless spectrum through *opportunistic wireless access*. In contrast to traditional approaches for accessing spectrum, where incumbent licensed transmissions possess exclusive rights across dedicated frequency bands, opportunistic wireless access allows for unlicensed wireless users to temporarily "borrow" unoccupied licensed frequency bands [2]. However, these unlicensed (i.e., *secondary*) devices must still guarantee interferencefree wireless access to incumbent licensed (i.e., *primary*) signals.

Given that the wireless spectrum occupancy for a specific band may simultaneously vary over time, frequency, and/or space, it is necessary to devise spectrally agile transmission schemes for secondary wireless transceivers such that they are capable of utilizing unoccupied frequency bands while respecting the rights of the primary wireless users under real-time operating conditions. In particular, it is essential that all out-of-band (OOB) radiation generated by the secondary wireless device is mitigated in order to prevent interference with primary wireless signals located in the frequency vicinity. Moreover, even if the secondary wireless devices are capable of transmitting across licensed spectrum causing minimal interference, it is possible that the size of the contiguous unoccupied frequency bands available to the secondary wireless device is insufficient, despite the fact that the aggregate bandwidth of the unoccupied frequencies might be adequate. Thus, secondary wireless transceivers require a level of *spectral agility* in order to operate in the presence of primary signals, especially with respect to mitigating interference resulting from OOB radiation, as well as simultaneously transmitting across several unoccupied frequency bands that are fragmented across the wireless spectrum whose aggregate bandwidth satisfies the secondary transmission requirements.

Many conventional signal waveforms do not possess the level of spectral agility needed to enable secondary wireless access within a dynamic spectrum access networks, whether this is due to a non-negligible amount of OOB radiation or the inability to simultaneously transmit across several non-contiguous bands. As a result, several researchers have proposed a spectrally agile modulation technique based on *multicarrier modulation* that deactivates subcarriers that would otherwise interfere with primary wireless transmissions. This technique is called *noncontiguous multicarrier modulation* [3], and it possesses the ability to efficiently use fragmented spectrum opportunities as well as perform spectrum shaping in order to suppress interference that may affect nearby primary wire-

0163-6804/11/$25.00 © 2011 IEEE

less transmissions. In this article, we present an overview of the state of the art in spectrally agile non-contiguous multicarrier modulation techniques that can be employed within an opportunistic wireless access network, as well as study two proposed implementations, non-contiguous orthogonal frequency-division multiplexing (NC-OFDM) [4] and non-orthogonal frequency-division multiplexing (NOFDM) [5], along with its non-contiguous variant (NC-NOFDM). Performance issues associated with these two techniques, including interference due to OOB radiation and implementation complexity, are carefully examined and compared with each other. The goal of this article is to provide the reader with insights on the emerging transmission techniques that will ultimately enable opportunistic wireless access and help to accommodate the growing demand for wireless spectrum.

The rest of this article is organized as follows. A brief taxonomy of multicarrier technologies is presented, as well as the concept of spectral agility within the context of opportunistic wireless access networks employing these technologies. Two spectrally agile multicarrier transmission approaches are presented. These include the fast Fourier transform (FFT)-based NC-OFDM approach and the filter bank-based NC-NOFDM approach. The performance of these two approaches in terms of OOB interference mitigation and implementation complexity is studied. This performance assessment is conducted using both computer simulations as well as actual real-world wireless spectrum measurements. Finally, practical considerations for spectrally agile multicarrier transceivers are discussed, and several concluding remarks are outlined.

Multicarrier Transmission Techniques for Opportunistic Radio

Taxonomy of Multicarrier Transmission Approaches

Multicarrier modulation is a form of FDM, where data is transmitted across several narrowband streams located at different carrier frequencies. As opposed to conventional FDM systems, where the narrowband subcarrier signals are separated by guard bands in the frequency domain, multicarrier modulation allows for the potential overlapping of adjacent subcarriers under a certain set of operating conditions, thus making this form of data transmission spectrally efficient. The parallelization of data symbols across several simultaneous subcarriers due to the "divide-and-conquer" nature of multicarrier modulation yields relatively long symbol durations when compared with the encountered duration of a time-dispersive channel impulse response. As a result, communication systems employing multicarrier modulation can efficiently handle the effects of intersymbol interference due to multipath propagation.

There are several implementations of multicarrier modulation that are based on how data is multiplexed and modulated onto several parallel subcarriers. In general, these implementations can be categorized into two families of multicarrier modulation:

Discrete Fourier transform-based multicarrier modulation: Employs discrete Fourier transform (DFT) basis functions in order to modulate subcarriers to different center frequencies. This form of multicarrier modulation can be efficiently implemented using an FFT algorithm, for example, radix-2 FFT, which possesses $M \log_2(M)$ complexity for M subcarriers, when M is the integer power of 2. Several popular examples employed in numerous commercial networking standards include OFDM and discrete multitone (DMT) modulation.

Filterbank multicarrier modulation: Employs bandpass filters at both the transmitter and the receiver to filter modulated subcarriers prior to combining them and separating them, respectively. Several examples include:

- *Complex exponential-modulated filterbanks* and *cosine-modulated filterbanks* modulate a prototype lowpass filter to different center frequencies using complex exponentials and cosines, respectively.
- *Transmultiplexers* are filterbanks used in multirate signal processing, and can be considered to be the functional dual of subband coders.
- *Perfect reconstruction filterbanks* are designed to eliminate crosstalk (intercarrier interference) under ideal channel conditions.
- *Oversampled filterbanks* employ a sampling factor higher than the total number of subcarriers.
- *Modified DFT filterbanks* delay either the real or imaginary components of each subcarrier signal with respect to each other to minimize crosstalk.

There have been several attempts to generalize the structure of a multicarrier modulation transceiver [6–8]. Consequently, a class of data transmission signals can be defined that is based on a generalized multicarrier (GMC) waveform characterized by two requirements:

- The number of subcarriers is higher than one.
- All transmit waveforms constitute an overcomplete basis set.[1]

As shown in Fig. 1a, both the GMC transmitter and receiver can readily be implemented by means of an FFT and an inverse FFT (IFFT) pair combined with polyphase filter banks[2] (or alternatively, via short-time Fourier transforms or Gabor transforms), which implement the subcarrier filtering (employing specific digital prototype filters at the transmitter and at the receiver, denoted as $\gamma_m[k]$ and $g_m[k]$ in Fig. 1a) and multicarrier modulation. In the time-frequency (T-F) representation of a GMC signal presented in Fig. 1b, we can observe T-F blocks of size $N_t \times M$, where N_t and M denote the number of pulses in time and frequency domain, respectively. In this figure, it can be readily observed that depending on the relation between the number of subcarriers M, the pulse-shape duration in samples L, and upsampling factor N, the information-bearing subcarrier pulses are overlapping in the T-F plane to a smaller or larger extent.

The GMC representation can be used to define a wide range of multicarrier waveform classes based on several characteristics, including whether the subcarriers are filtered or non-filtered, orthogonal or non-orthogonal, and precod-

As opposed to conventional FDM systems, where the narrowband subcarrier signals are separated by guard bands in the frequency domain, multicarrier modulation allows for the potential overlapping of adjacent subcarriers under a certain set of operating conditions, making this form of data transmission spectrally efficient.

[1] *Representation of any signal can be accomplished using several basis waveforms from this set multiplied by the analysis coefficients. Furthermore, the reconstruction of these coefficients (usually the transmit data) is feasible without any loss of information.*

[2] *A polyphase filter bank splits an input signal into a given number* M *of equidistant subband signals and implements filtering of these subband signals.*

Multicarrier modulation possesses the required level of spectral agility in order to transmit data from unlicensed transmitters across several fragmented frequency bands simultaneously even in the presence of licensed signals, thus resulting in an increase in spectrum utilization.

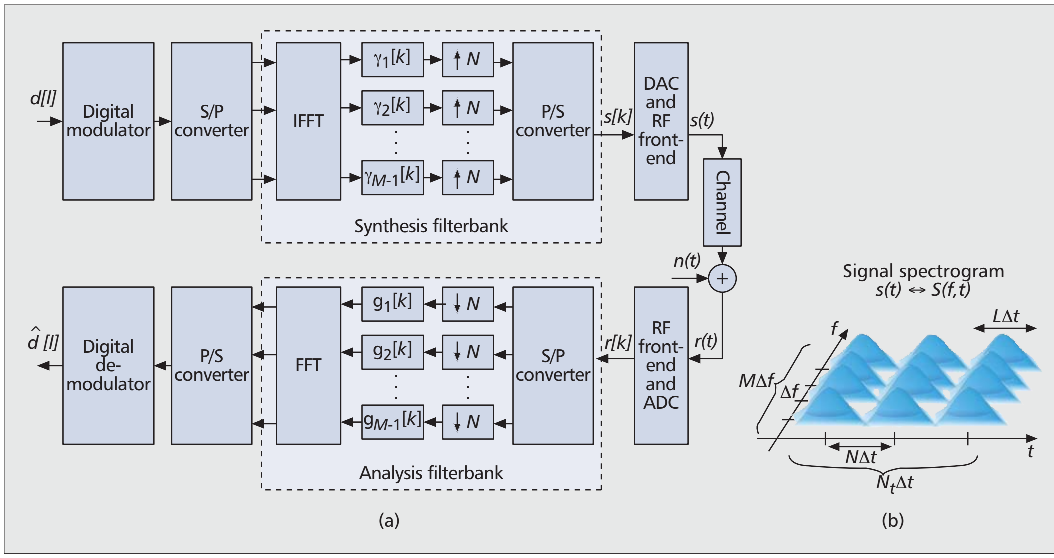


Figure 1. *a) Signal processing and representation in GMC transmission link; b) the time-frequency (T-F) representation of the transmit signal* s(t). Δf *— subcarriers distance,* Δt *— sampling period.* L, N, M, N_t *denote: the duration of the pulse-shaping filter impulse response* $\gamma[k]$ *in samples (equal to the duration of the dual-filter impulse response* g[k] *at the receiver), the oversampling rate, the number of subcarriers, and the number of pulses in time, respectively.*

ed or non-precoded. An example of multicarrier waveform classification is shown in Fig. 2. In what follows, we focus on the waveforms that possess the necessary level of spectral agility for performing opportunistic spectrum access, mainly in cognitive radio, where the protection of the primary users' transmission is essential. Within the GMC framework, we will consider several candidate multicarrier technologies, possibly applying non-contiguous and non-orthogonal subcarriers.[3]

One characteristic that will be discussed further in this article is the division of GMC waveforms into two subcategories based on the orthogonality of their subcarrier waveforms. One of the subcategories possess subcarrier waveforms satisfying an orthogonality property (e.g., OFDM), while the other subcategory does not satisfy this property, an example of which is NOFDM. It should be noted that the subcarriers in the latter subcategory do not need to satisfy the orthogonality property, and thus do not possess the same transmission restrictions imposed on orthogonal subcarrier waveforms with respect to subcarrier pulses distances in either the time or frequency domains. Relaxation of these restrictions can increase the spectral efficiency of NOFDM. However, subband filtering is necessary in order to limit the interference between the subcarriers. An example technique that employs both the subcarriers orthogonality and their filtering is filtered multitone (FMT) modulation, and its version that uses oversampled filterbanks (FMT-OS).

The taxonomy discussed in this section only addresses the various forms of modulation that can be employed by multicarrier modulation transceivers. Accompanying technologies, such as channel-coding, spectrum-spreading, or multi-antenna technologies, have been omitted from this overview. Nevertheless, the impact and importance of these additional techniques on the performance of multicarrier communication systems is the subject of numerous research investigations from around the world.

Multicarrier-Based Opportunistic Wireless Access

Multicarrier modulation possesses the required level of spectral agility in order to transmit data from unlicensed transmitters across several fragmented frequency bands simultaneously even in the presence of licensed signals, thus resulting in an increase in spectrum utilization [9]. In particular, subcarriers located in the frequency vicinity of unoccupied wireless spectrum can be used for transmitting data while those subcarriers that could potentially interfere with licensed signals are deactivated (i.e., turned off). However, simply deactivating subcarriers located in the vicinity of licensed signals may not be sufficient in minimizing the impact of OOB interference, especially for low interference tolerance levels such as –70 dB and below. On the other hand, there exist several published techniques in the literature that are designed to counteract the effects of OOB interference.

Another approach for minimizing the amount of OOB interference generated via non-contiguous multicarrier data transmission is to employ multi-rate filter banks, whose subband spectra can be designed to be highly spectrally selective. Moreover, the subband spectral selectivity limits the amount of intercarrier interference (ICI) considerably, which is especially advantageous in mobile communications that suffer from Doppler effects and frequency selectivity of the channels.[4] Conversely, in terms of implementation complexity, OFDM-based transceivers possess an obvious advantage over multi-rate filter banks approaches due to the usage of radix-2 IFFT/FFT pairs in the subcarrier modulation/demodulation stage of the former. Note that the proper design of a filterbank-based multicarrier transceiver may apply the same radix-2 IFFT/FFT accompanied by the polyphase filterbank, and may present relatively low complexity, although higher than that of an

[3] *For this reason, orthogonally precoded multicarrier schemes considered in [6] and other papers on orthogonal GMC or CDMA-GMC transmission schemes will not be considered.*

OFDM transceiver. This point will be elaborated later, where the design trade-offs that exist between the selection of various non-contiguous multicarrier transmission schemes based on interference and complexity are discussed in greater detail.

Apart from achieving a required level of OOB interference within a given spectrum mask, an unlicensed transmitter performing opportunistic wireless access must be capable of tailoring its spectral characteristics dynamically to respect the rights of the dynamically changing incumbent licensed transmissions. For non-contiguous OFDM-based opportunistic wireless transceivers, the dynamic spectral-mask constraints can be respected via one of several techniques published in the open literature, such as by varying the number of guard subcarriers and/or cancellation subcarriers, as well as by filtering the composite OFDM transmit signal. If filtering is performed, the filters will need to be tunable since the available spectrum holes and the interference requirements may vary in time. As for filter bank-based multicarrier transmission techniques, such as non-orthogonal frequency division multiplexing (NOFDM), tunable digital filters are needed on a per-subcarrier basis to achieve the required spectral selectivity and OOB interference levels.

One challenge encountered by all multicarrier transmission approaches is the possibility of exhibiting large envelope variations in the time domain, which is often characterized by a high peak-to-average power ratio (PAPR). When high PAPR occurs, the digital-to-analog (D/A) converter and power amplifier (PA) of the transmitter would require a large dynamic range to avoid amplitude clipping, thus increasing both power consumption and component cost. If amplitude clipping does occur, the resulting transmission spectrum broadens and produces OOB interference regardless of whether the initial spectral waveform has been properly shaped at the transmitter for low OOB interference. Furthermore, as shown in [11], filter-bank-based multicarrier transmission schemes may suffer even more since its PAPR is higher than that of an OFDM signal with the same number of subcarriers. Thus, apart from the spectrum shaping, efficient PAPR reduction methods must be employed in order to counteract this serious issue. In the next section, this issue will be addressed in greater detail with respect to the considered spectrally-agile multicarrier waveforms.

Overall, non-contiguous multicarrier modulation techniques have been recognized as a suitable candidate for opportunistic wireless access due to their efficient spectrum utilization and achieving high data rates while being sufficiently spectrally agile for transmitting across several fragmented frequency bands. In fact, this form of data transmission approach is well-suited for future wireless communication systems and even cognitive radio systems [12]. In the following section, we will investigate two non-contiguous multicarrier data transmission approaches that could be employed in an opportunistic wireless access environment, namely: NC-OFDM and NC-NOFDM.

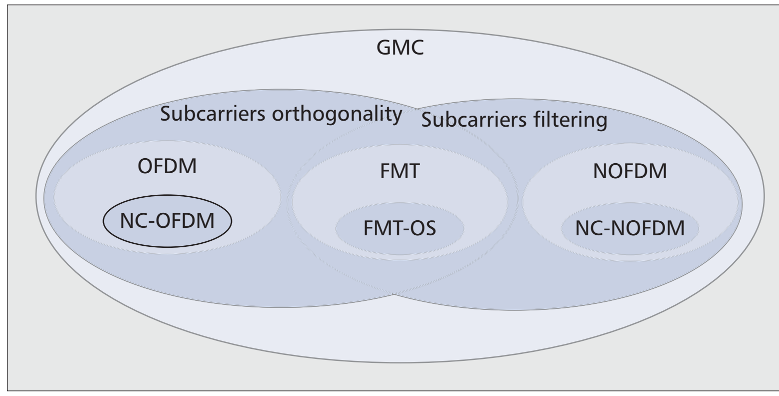


Figure 2. *The taxonomy of multicarrier technologies.*

Waveforms Case Studies for Spectral Agility

Given the variety of GMC representations that could be employed for achieving opportunistic wireless access, we will focus our attention on two fundamentally different implementations, with one implementation being based on a sole DFT approach (i.e., NC-OFDM), while the other is based on a filter bank approach (i.e., NC-NOFDM). For a detailed mathematical presentation of these two multicarrier data transmission approaches, the interested reader is encouraged to refer to the authors' other publication on this topic [13].

Non-Contiguous Orthogonal Frequency-Division Multiplexing Waveforms

For OFDM-based data transmissions within an opportunistic wireless access environment, there exists the possibility of having unoccupied licensed frequency bands that are not wide enough to accommodate a single contiguous multicarrier signal given a specific overall data rate. Consequently, to provide high data rates while avoiding interference with incumbent licensed transmissions, a variant of OFDM called non-contiguous OFDM (NC-OFDM) has been proposed in the open literature [4], where the implementation achieves high data rates via the collective usage of non-contiguous blocks of OFDM subcarriers. Simultaneously, the NC-OFDM transceiver avoids interference with incumbent licensed users by deactivating subcarriers within their frequency vicinity.[5] Thus, NC-OFDM is a viable transmission technology for cognitive radio transceivers and opportunistic wireless access networks.

When implementing an OFDM transceiver, the IFFT/FFT algorithm is employed to make modulation and demodulation highly efficient in terms of hardware and computational complexity. In fact, it has been shown that a radix-2 FFT/IFFT pair possesses $\mathcal{O}(M \log_2(M))$ complexity for $M = 2^p$ subcarriers (where p is a positive integer), thus making it an attractive option for multicarrier-based communication systems. Moreover, the hardware implementation of an NC-OFDM transceiver can be further streamlined by performing *FFT pruning* [4] on the FFT/IFFT pair, especially if the baseband transceiver modules are constructed on a programmable or software-defined communication system.

[4] *OFDM subcarriers lose orthogonality when operating in hostile wireless transmission environments. Moreover, imperfections occurring between the transmitter and receiver during synchronization as well as frequency offset compensation may result in serious ICI issues.*

[5] *Note that single-user transmissions in OFDMA (or its frequency-hopping equivalent) can be treated as a special case of NCOFDM when the primary users employ subcarriers from the same orthogonal basis-set. In the case of cognitive radio transmission, the primary user transmission may be based on any technology (e.g., single- or multicarrier with different subcarrier spacing quantities). Thus, it is crucial to limit the OOB and related interference to this user.*

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

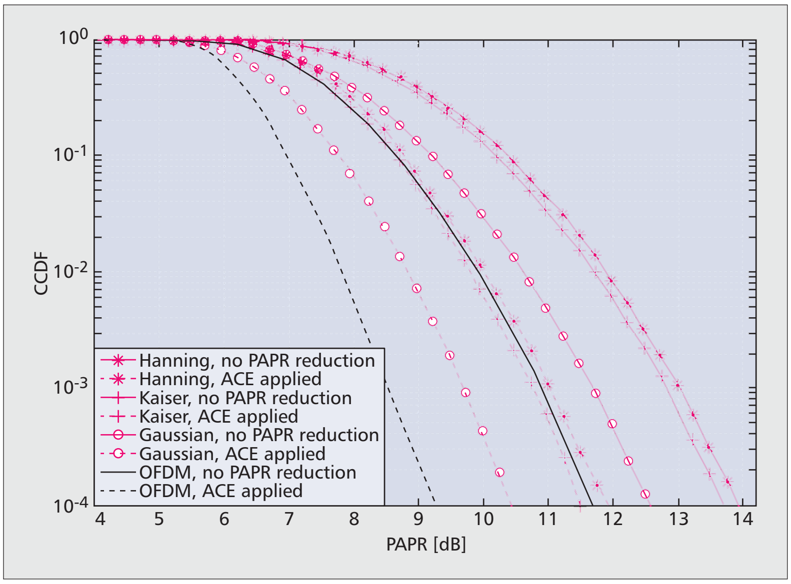


Figure 3. *Complementary cumulative distribution function of the PAPR in QPSK-based (NC-)OFDM and (NC-)NOFDM with various pulse shapes (the shape-factor* $\beta = 5$ *used in Kaiser window) with and without the application of ACE — the PAPR reduction method;* $M = 64$, $N = 48$, $L = 72$.

An important design issue associated with the use of OFDM for spectrum pooling-based wireless systems is the potentially large amount of OOB interference levels, which can negatively affect neighboring wireless signals. Consequently, several techniques have been proposed in the literature that are designed to significantly suppress these sidelobes in order to make coexistence between primary and secondary users feasible. The suppression achieved by applying the earliest proposed techniques, such as *windowing* and *insertion of guard carriers*, is not commensurate with the loss of other system resources, such as prolonged symbol duration in the case of windowing and wasted transmission bandwidth in the case of guard band insertion. Subsequently proposed techniques, such as *insertion of cancellation carriers*, *subcarrier weighting*, and *reserved tones*, involve complex optimization methods, making their real-time implementation expensive when the number of subcarriers is large and when higher order modulation schemes are involved [10]. Finally, techniques such as *constellation expansion* and *modulated filter banks* applied to contiguous groupings of active subcarriers are relatively low-complexity approaches to counteract the effects of OOB interference.

Non-contiguous Non-orthogonal Frequency Division Multiplexing Waveforms

As previously mentioned, NOFDM transmission offers high spectral efficiency at the expense of the higher computation complexity relative to OFDM-based multicarrier modulation [7]. When applying NOFDM, some weak overlapping between neighboring waveforms in both the time and frequency domains (called *self-interference*) is permitted in order to reduce the distances between these pulses, thus making the T-F grid denser. As a result, an increase in the density of the T-F grid of pulses will result in a higher spectral efficiency. Moreover, the level of self-interference can be controlled via the proper design of the transmitter-receiver pulse pairs. In other words, when addressing the issue of interference avoidance or mitigation, an NOFDM transceiver can be designed such that an advantageous pulse shape can be selected that would enable non-contiguous transmission across the frequency domain. Specifically, NOFDM composite waveforms can be designed such that the OOB interference resulting from the waveform sidelobe levels are very low. Moreover, since the distances between neighboring waveforms both in the time and frequency domains can be lower than in the OFDM case, it is evident that the allowable narrow spectrum gaps (especially in the interleaved spectrum) can be utilized more efficiently.

Another important issue that is related to the OOB interference is the potential for high PAPR occurring across the time-domain signal samples. In a case of NOFDM transmission, the problem is even more significant since the filtering, i.e., pulse shaping, realized by the polyphase filters can instantaneously increase the amplitude of the signal samples. To help mitigate the occurrence of high PAPR, a PAPR reduction method based on the so-called active constellation expansion (ACE) approach has been proposed for NOFDM systems [11]. It has been shown that the ACE method properly modified and improved for NOFDM can reduce PAPR as efficiently as in the case of regular ACE applied for regular OFDM.

In Fig. 3, a complementary cumulative distribution function (CCDF) of PAPR is presented for the OFDM and for a number of example NOFDM transmission schemes employing QPSK modulation at $M = 64$ (possibly non-contiguous) subcarriers. In the considered NOFDM cases, various pulse-shapes have been applied with pulse duration in samples $L = 72$ and the oversampling rate $N = 48$. Note that in such cases subsequent pulses overlap each other in both the time and frequency domain. In this figure, the results are presented for respective transmit signals in the case when no PAPR reduction method has been applied and in the case, when the ACE method has been applied modified appropriately for NOFDM signaling [11]. One can see that the PAPR-reduction results for NOFDM are similar to the ones obtained for OFDM (PAPR is reduced by approximately 2.5 dB), despite much higher initial PAPR inherent for NOFDM signals resulting from overcritical sampling and relatively long pulse duration ($L > N$). Finally, similar conclusions can be drawn for OFDM and NOFDM transmitters using the same actual number (M) of non-contiguous subcarriers.

Regarding the non-contiguous variant of NOFDM transmission, called NC-NOFDM, its operation is similar to that of NC-OFDM in the sense that subcarriers located in the spectral vicinity of the incumbent licensed signals can be "turned off" in order to prevent interference. Furthermore, due to the high level of spectral selectivity of the subcarriers, no guard or cancellation subcarriers for the pulses are required. As a result, it is possible for an NC-NOFDM transceiver to achieve higher resource utilization in terms of spectral efficiency since fewer subcarriers are needed to mitigate the effects of high OOB interference. This is illustrated in Fig. 4, where the NC-NOFDM waveform employs a larger number of subcarriers and possesses higher spectral selectivity relative to the NC-OFDM waveform. On the other hand, this efficiency in resource utilization comes at the cost of implementation complexity. In the following section, the issues of interference and complexity will be discussed in further detail.

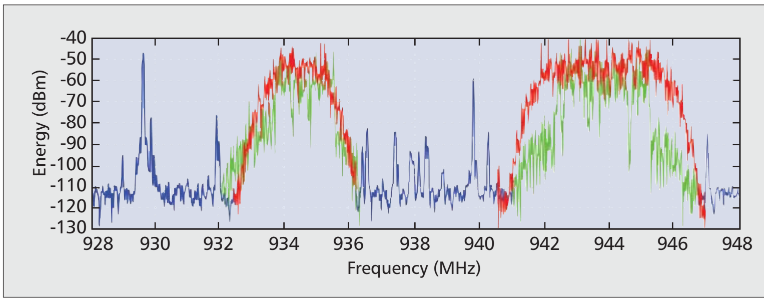


Figure 4. *An illustration of the non-contiguous OFDM (shown in green) and non-contiguous NOFDM (shown in red) transmissions in the paging band (928–948 MHz). Paging transmission is shown in blue.*

Performance and Complexity Evaluation

Interference Mitigation

The process of quantifying the OOB interference experienced by both licensed and unlicensed transmissions located within the same frequency vicinity was based on 125 spectrum measurement sweeps of the paging band (928–948 MHz) taken in Worcester, MA, USA on July 27, 2008 at location N42°17.829′, W071°50.3607′. The frequency resolution for these time sweeps is 20 kHz. These real-world spectrum measurements were then employed in an emulation environment, where 25, 000 binary phase shift keying (BPSK) symbols where simulated, and the number of subcarriers transmitting these symbols were based on the primary user spectrum occupancy characteristics observed. For the considered NC-NOFDM case, Gaussian window was applied for shaping the subband pulses. (Note relatively low CCDF of PAPR for this pulse shape that can be seen in Fig. 3.) The duration of the pulse-shape in samples L was equal to the oversampling rate N in the time domain ($L = N = 1024$). Otsu's algorithm was employed to classify the channels as either "occupied" or "available"[6] [14], and data was inserted in the available channels such that the interference affecting neighboring licensed transmissions was below a set power-level [15]. The Parks-McClellan optimal equiripple finite impulse response (FIR) receiver filter was employed at the receiver front-end to eliminate the interference from the neighboring licensed or unlicensed signals. The following assumptions were made in performing our simulation-based analysis in order to understand the interference characteristics under certain basic conditions:

- The primary user spectrum occupancy profile observed by the secondary user transmitter and the intended secondary user receiver is the same. That is, the secondary transmitter and the receiver are closely located.
- The primary user spectrum occupancy profile does not change between the scanning time and the transmission time of the secondary user transmitter.

Interference as a Function of the Receiver Filter Order — Figure 5a shows the plot of the interference experienced by the licensed and unlicensed transmissions as a function of the receiver filter order used at the receiver end. The y-axis shown is the percentage of the interference power as seen by the receiver filter with respect to the power of the desired signal spectrum in the absence of an interfering signal. In other words, it is the ratio of the interference power seen by licensed/unlicensed transmissions to the total power used in signal transmission times 100. The difference in the interference powers "seen" by the licensed (paging) transmission and the unlicensed transmission is due to the fact that the amount of signal power that enters the passband of the receiver filter that can potentially cause interference to the OFDM signal is considerably smaller compared to the amount of OFDM signal power that interferes with the narrowband paging signal. As expected, with an increase in the receiver filter order, the passband can be made steeper and hence a lower amount of interference power is drawn by the receiver filter. This plot highlights the constraints on the receiver filter in order to minimize the amount of interference in an NC-OFDM scenario. In contrast, since the NC-NOFDM pulse contains smoother edges compared to an NC-OFDM signal with a rectangular pulse, the NC-NOFDM transceiver will achieve better OOB interference mitigation performance relative to an NC-OFDM transceiver.

Interference as a Function of the Spectrum-Hole-Detection Threshold — Relaxing the first assumption introduced at the start of this section, allows us to examine and understand the impact of the variation in Otsu's threshold on the interference characteristics. In generating Fig. 5b, a Parks-McClellan FIR receive filter of order 25 was employed. The spectrum-hole-detection threshold varied according to Otsu's algorithm such that the interference resulting from the sidelobes of the OFDM and NOFDM signal at the frequency location of the licensed transmission is above the optimum threshold. Thus, the x-axis in Fig. 5b represents the difference between the optimum value of the threshold (Otsu's threshold) and the actual threshold used in inserting the subcarriers in the spectrum holes. Clearly, when the threshold difference is well below the optimum value (the thresholding difference equals 0 dBm), the interference caused is extremely small since a fewer number of data subcarriers are inserted in order to have the sidelobes that are of extremely small power at the frequency location of the licensed transmission. As the threshold increases beyond the optimum value, the interference power quickly rises to extremely high values. Thus, the (licensed) paging signal is affected more when compared to the unlicensed transmission. Another way of interpreting the above result is by the

[6] *Otsu's method is used to calculate the optimum threshold (the maximum inter-class variance) separating two classes of data (e.g., pixels of a foreground and a background in image processing). Here, we use it to separate the paging bands from the spectrum holes.*

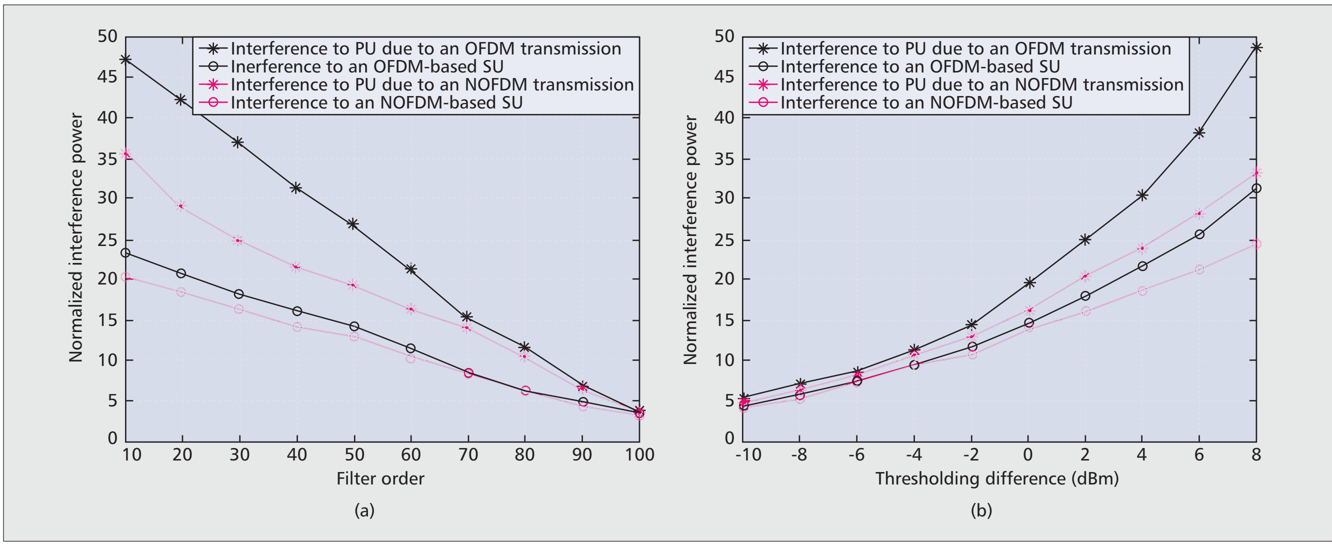


Figure 5. *Normalized interference caused to the primary and secondary user signals as a function of a) the Parks-McClellan FIR receiver filter order; b) the spectrum-hole-detection thresholding difference between the transmitter and the receiver.*

notion that the licensed transmission spectrum occupancy profile seen by the unlicensed transmitter and receiver are no longer the same, i.e., the transmitter and the receiver are sufficiently separated so that the log-normal shadowing affects the received signal. As a result of subcarriers selectivity, the NCNOFDM transceiver achieves better interference mitigation performance relative to NC-OFDM system for every value of the thresholding difference.

Implementation Complexity

Clearly, while considering the non-continuous OFDM or NOFDM techniques, the question concerning the computational complexity and the feasibility for practical implementation arises. The benefits of applying the non-continuous subcarriers are noticeable when the number of subcarriers is sufficiently large, and when we apply the pruning mechanisms in the IFFT/FFT implementation. The price for high spectral efficiency and for the OOB interference mitigation capabilities of these techniques is either the increase in the overall transmitter-receiver complexity or the decrease of spectral efficiency. This is due to the following reasons:

- In the NOFDM case, additional filtering operations are required when compared with the OFDM case
- In both NC-OFDM and NC-NOFDM systems, the subcarriers located in the frequency vicinity of the PU together with several guard subcarriers must be excluded from the modulation process.

In Fig. 6, the number of required operations employed at the transmitter (resp. modulator) and at the receiver (resp. demodulator) per sampling period versus the number of subcarriers is presented for the worst-case scenario, i.e., when no simplifications in the IFFT/FFT implementation is possible. It can be observed that for the NOFDM modulator/demodulator, the polyphase decomposition of the pulse-shaping filter has been applied, thus resulting in an IFFT/FFT implementation with a polyphase filterbank. Moreover, it can be observed that the NOFDM-system transceiver is more demanding in terms of computational complexity due to the application of filterbanks. Although the complexity of the NOFDM modulator/demodulator appears to be marginally higher when compared with OFDM, it should be noted that the complexity of several associated NOFDM signal processing algorithms, such as channel equalization or interference cancellation, can be significantly more complex relative to the same algorithms employed by an OFDM-based system. Additionally, the complexity increase associated with these NOFDM signal processing modules also depend on the employed pulse shaping filters and other GMC parameters, i.e. M, N, and L. Finally, as mentioned before, the problem of PAPR is more severe in NOFDM and PAPR reduction techniques are usually more complex (again to the degree depending on the adopted parameters).

Note that in practice, various techniques can be used to reduce the number of operations needed for modulation and demodulation of NC-OFDM and NC-NOFDM signals, with one example being the pruning of the IFFF/FFT operations. Moreover, several of the algorithms employed at the receiver working in conjunction with the demodulator can be simplified due to some of the features associated with NOFDM waveforms, e.g. frequency-offset correction. The interested reader is encouraged to refer to the complexity analysis derived for GMC transmission links found in [7].

Conclusion

Multicarrier modulation is well suited for opportunistic wireless access and cognitive radio applications due to its spectral agility, transmission parameter flexibility, efficient spectrum utilization, aggregation and spectrum shaping capabilities. The challenges posed by employing non-contiguous multicarrier-modulation-based data transmission systems include high out-of-band interference and implementation complexity, which become especially apparent when dealing with fragmented bands of unoccupied wireless spectrum. As described in this article, there does not exist a single non-contiguous multicarrier solution that possesses both low out-of-band interference and low implementation complexity, but rather these objectives form the core of a trade-off analysis that is dependent on the deployment scenario for these systems and their desired performance. Future work in this area would definitely include the development of a quantitative cost-benefit or complexity-performance analysis that could be used to assess different orthogonal or non-orthogonal spectrally agile waveforms for specific wireless applications. Such an analysis could also be employed as part of the decision-making process in cognitive radio systems.

REFERENCES

[1] FCC, "Spectrum Policy Task Force Report," ET Docket No. 02-135, 2002.

[2] FCC, "Unlicensed Operation in the TV Broadcast Bands," ET Docket No. 04-186, 2004.

[3] A.M. Wyglinski. "Effects of Bit Allocation on Non-contiguous Multicarrier-based Cognitive Radio Transceivers," *Proc. IEEE VTC-Fall '06*, Sept. 2006.

[4] R. Rajbanshi, A. M. Wyglinski, and G. J. Minden, "An Efficient Implementation of NC-OFDM Transceivers for Cognitive Radios," *Proc. Int'l Conf. Cognitive Radio Oriented Wireless Networks and Commun.*, June 2006.

[5] W. Kozek and A. F. Molisch, "Nonorthogonal Pulse-shapes for Multicarrier Communications in Doubly Dispersive Channels," *IEEE Trans. Commun.*, vol. 16, no. 8, 1998, pp. 1579–89.

[6] A. Scaglione, G. B. Giannakis, and S. Barbarossa, "Redundant Filterbank Precoders and Equalizers Part I: Unification and Optimal Designs," *IEEE Trans. Sig. Proc.*, vol. 47, no. 7, July 1999, pp. 1988–2005.

[7] S. Stefanatos and A. Polydoros, "Gabor-based Waveform Generation for Parametrically Flexible, Multi-Standard Transmitters," *Proc. Euro. Sig. Processing Conf.*, Sept. 3–7, 2007, Poznan, Poland, pp. 871–75.

[8] H. Bogucka *et al.*, "GMCR Transceivers for air Interfaces in Single Mode Operation," Deliv. D3.1, IST-027960 Euro. 6th FP Project URANUS, Jan. 2007.

[9] T. A. Weiss and F. K. Jondral, "Spectrum Pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency?," *IEEE Commun. Mag.*, vol. 43, Mar. 2004, pp. S8–14.

[10] S. Brandes, I. Cosovic, and M. Schnell, "Reduction of Out-of-Band Radiation in OFDM Systems by Insertion of Cancellation Carriers," *IEEE Commun. Letters*, vol. 10, no. 6, 2006, pp. 420–22.

[11] A. Kliks and H. Bogucka, "Improving Effectiveness of the Active Constellation Extension Method for PAPR Reduction in Generalized Multicarrier Signals," *Wireless Pers. Commun., An International Journal*, 2010.

[12] H. A. Mahmoud, T. Yucek, and H. Arslan, "OFDM for Cognitive Radio: Merits and Challenges," *IEEE Wireless Commun.*, Apr. 2009, vol. 16, no. 2, pp. 6–14.

[13] S. Pagadarai *et al.*, "On Non-Contiguous Multicarrier Waveforms for Spectrally Opportunistic Cognitive Radio Systems" *Proc. 5th IEEE Int'l. Waveform Diversity and Design Conf.*, 8–13 Aug. 2010.

[14] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Trans. Sys., Man., Cyber.*, Jan. 1979, vol. 9, no. 1, pp. 62–66.

[15] S. Pagadarai and A. M. Wyglinski, "A Quantitative Assessment of Wireless Spectrum Measurements for Dynamic Spectrum Access," *Proc. Int'l. Conf. Cognitive Radio Oriented Wireless Networks and Commun.*, June 2009.

BIOGRAPHIES

HANNA BOGUCKA (hbogucka@et.put.poznan.pl) received M.Sc. and Ph.D. degrees in telecommunications from Poznan University of Technology (PUT), Poland, in 1988 and 1995, respectively. Since 1988 she has been employed at PUT, currently in the Chair of Wireless Communications as a professor and deputy dean for research on the Faculty of Electronics and Telecommunication. She has research interests in the area of wireless communications, and flexible, adaptive, and cognitive radio systems. She is the author of more than 90 papers and three handbooks (in Polish) in the area of radio communications.

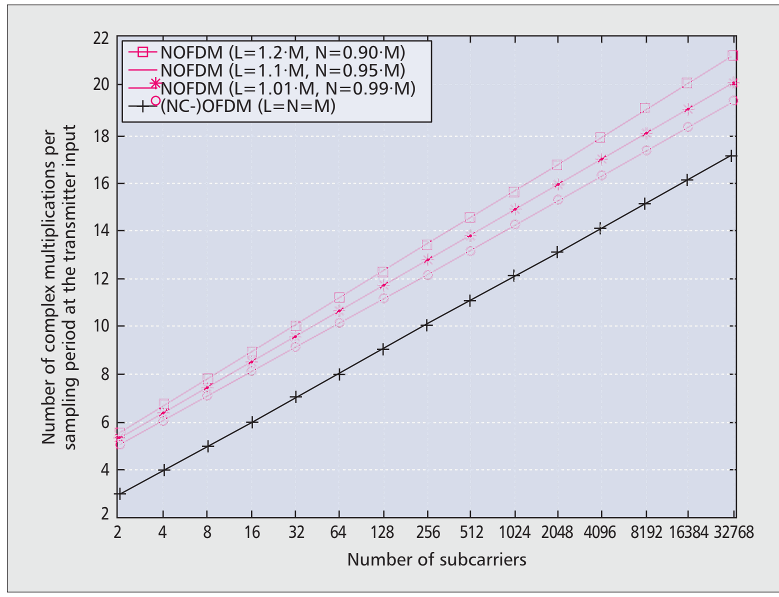


Figure 6. *Complexity of NC-NOFDM and NC-OFDM transmissions as a function of the number of subcarriers.*

ALEXANDER M. WYGLINSKI [SM] is an assistant professor of electrical and computer engineering at Worcester Polytechnic Institute (WPI), director of the WPI Limerick (Ireland) Project Center, and director of the Wireless Innovation Laboratory (WI Lab). He received his Ph.D. degree from McGill University in 2005, his M.S.(Eng.) degree from Queens University at Kingston in 2000, and his B.Eng. degree from McGill University in 1999, all in electrical engineering. He is very actively involved in the wireless communications research community, especially in the fields of cognitive radio systems and dynamic spectrum access networks. He currently serves on the editorial boards of both *IEEE Communications Magazine* and *IEEE Communications Surveys and Tutorials*, has co-edited/co-written the first-ever textbook on cognitive radio, *Cognitive Radio Communications and Networks: Principles and Practice* (Academic Press, December 2009), has been a guest co-editor for *IEEE Communications Magazine* of two feature topics on cognitive radio (May 2007, April 2008), and has served on the organizing committees of numerous international conferences including IEEE GLOBECOM 2011, IEEE DySPAN 2011, IEEE MILCOM 2009, IEEE VTC 2007-Fall, and CrownCom 2007. He is a senior member of the IEEE Signal Processing Society, IEEE Vehicular Technology Society, IEEE Women in Engineering, and Sigma Xi. His current research interests are in the areas of wireless communications, wireless networks, cognitive radios, software-defined radios, transceiver optimization algorithms, dynamic spectrum access networks, spectrum sensing techniques, machine learning techniques for communication systems, and signal processing techniques for digital communications.

SRIKANTH PAGADARAI is a Ph.D. candidate in the Department of Electrical and Computer Engineering at WPI. He received his B.S. degree from Jawaharlal Nehru Technological University, Hyderabad, India, in 2005, and his M.S. degree from the University of Kansas, Lawrence, in 2007. His M.S. thesis focused on sidelobe suppression techniques for NC-OFDM systems. His current doctoral research focuses on the quantification, characterization, and exploitation of wireless spectrum, spectrally agile OFDM-based transceiver systems, and adaptive signal processing techniques for the extraction of wireless signatures from interference-prone propagation environments.

ADRIAN KLIKS received his M.Sc. degree in telecommunication from Poznan University of Technology in 2005. Since 2005 he has been employed at the Institute of Electronics and Telecommunications, and since 2007 as the senior researcher at the Chair of Wireless Communication in the Faculty of Electronics and Telecommunication. Since 2005 he has been a Ph.D. student, and starting from 1 March 2009 he has been working as an assistant. His research interests cover the wide spectrum of wireless communications. In particular he is interested in multicarrier (both orthogonal and non-orthogonal) systems, in the area of software defined, adaptive, and cognitive radios, in hardware implementation on DSP/FPGA, and in radio planning.

SERIES EDITORIAL

THE LATEST IN CONSUMER COMMUNICATIONS, STRAIGHT FROM THE IEEE CONSUMER COMMUNICATIONS AND NETWORKING CONFERENCE 2011


Mario Kolberg


Madjid Merabti


Stan Moyer

Every year in January, researchers working on consumer communications and networking come together for the IEEE Consumer Communications and Networking conference in Las Vegas. Major themes of the conference are Emerging Consumer Technologies and Applications, Peer-to-Peer Networking and Content Distribution, Security and Content Protection, Multimedia and Entertainment Networking and Services, Smart Spaces and Personal Area Networks, and Wireless Consumer Communications and Networking. Each of these themes forms a separate track in the conference program. In addition, the conference is host to a series of Special Sessions and Workshops as well as prototype demonstrations. Some of the latter were selected to also be shown at the Consumer Electronics Show, the world's largest consumer electronics trade show, which is hosted in Las Vegas around the same time as CCNC.

This issue of the Consumer Communications Series features selected articles from the CCNC conference. These have been nominated as candidates for the Best Paper Awards at the conference. The first article in this series, "An Adaptive User Interface Based on Spatiotemporal Structure Learning" by Hosub Lee and his co-authors from the Samsung Advanced Institute of Technology, was the winner of the award. Their article describes a novel user interface prototype for the Android smartphone that recommends a number of applications to best match the user's context. Context is the combined state from time, location, weather, emotion, and activities. Using a probabilistic learning and inference algorithm, the best three applications are recommended.

The second article, "Synchronized Multimedia Streaming on the iPhone Platform with Network Coding" by Peter Vingelmann and his co authors from Budapest University of Technology and Arlborg University, won the Best Student Paper Award. This article proposes a solution for streaming content from a single source to multiple receivers using a novel network coding approach to efficiently and reliably deliver multimedia content to the destinations in a synchronized manner.

The remaining three articles were selected as best in their respective track and were runners up to the awards. These articles are "Real-Time Probing of Available Bandwidth in Home Networks" by Archi Delphinanto *et al.*, who propose a novel probing technique for heterogeneous IP-based home networks. This approach, which does not require a priori knowledge on the link layer topology, is sufficiently accurate to decide on admission of high-throughput media streams such as IPTV. The next article, "Using Home Routers as Proxies for Energy-Efficient BitTorrent Downloads to Mobile Phones" by Imre Kelenyi *et al.*, analyzes the implications of hosting a BitTorrent proxy on home routers. The final article in this series, "Takeover TV: Facilitating the Social Negotiation of Television Content in Public Spaces" by Greg T. Elliott *et al.*, provides a new approach to social TV. This article introduces a novel system that lets users influence and interact with movies and TV shown in public spaces such as bars.

All of these articles are reprints from the CCNC conference proceedings. If any of the articles in this series are of interest to you, we strongly urge you to consider participating in the next running of the IEEE Consumer Communications and Networking Conference (CCNC) that will be held next January in Las Vegas in conjunction with the Consumer Electronics Show, the largest CE show in the world. See http://www.ieee-ccnc.org for details.

BIOGRAPHIES

MARIO KOLBERG [SM] (mko@cs.stir.ac.uk) is a lecturer within the Institute of Computing Science and Mathematics at the University of Stirling. His research interests include peer-to-peer overlay networks, home automation, and IP telephony. He was leading a project funded by Panasonic (USA) investigating efficiency gains in structured peer-to-peer overlays. He was the academic supervisor in a Knowledge Transfer Partnership focusing on developing a peer-to-peer overlay for mobile handsets. He is working in the ESRC project Interlife where he is working on using peer-to-peer networks with 3D virtual worlds in an educational context. He is also involved in the MATCH project, focusing on integrating different network technologies for care in the home. He is on the editorial Board of the Springer journal *Peer-to-Peer Networking and Applications* and has a longstanding involvement with the IEEE CCNC conference series. He served as its TPC Chair for the January 2011 running. He is TPC co-chair of the 5th International Conference on Internet Multimedia Systems Architecture and Applications (IMSAA-11) to be held in December 2011 in Bangalore, India. He has published more than 50 papers in leading journals and conferences. He is a member of a number of international conferences program committees on networking and communications. He holds a Ph.D. from the University of Strathclyde, United Kingdom.

STAN MOYER [SM] (stanm@research.telcordia.com) is executive director and strategic research program manager in the Applied Research area of Telcordia Technologies, where he has worked since 1990. Currently, he is the product manager for the Telcordia Mobile Messaging and Application Solution. In the past he has led research and business development activities related to digital content services and home networking. Prior to that he worked on ATM switch hardware, broadband network architectures and protocols, middleware, Internet network and application security, Internet QoS, and voice over IP. He is currently president of the OSGi™ Alliance. He is a member of the Board and Treasurer for the IEEE Communications Society, vice-chair of the IEEE CCNC steering committee, and a member of the IEEE Technical Activities Board Finance Committee.

MADJID MERABTI [M] (M.Merabti@ljmu.ac.uk) is a professor of networked systems and director of the School of Computing and Mathematical Sciences at Liverpool John Moores University, United Kingdom. He holds a Ph.D. from Lancaster University, United Kingdom. He has over 20 years' experience in conducting research and teaching in the areas of computer networks (fixed and wireless), mobile computing, and computer network security. He is widely published, with over 150 publications in these areas, and leads the Distributed Multimedia Systems and Security Research Group. He is principal investigator in a number of current projects: Mobile Networks Security and Privacy Architectures and Protocols, Secure Component Composition in Ubiquitous Personal Networks, Networked Appliances, Mobile and Ad Hoc Computing Environments, Sensor Networks, and computer games technology. He was Guest Editor for the Special Issue on Research Developments in *Consumer Communications and Networking of Multimedia Tools and Applications: An International Journal* (Kluwer, September 2005). He is a member of the Steering Committee for IEEE CCNC. He has acted as TPC chair for a number of international conferences, including the 5th IEEE Workshop on Networked Appliances, Liverpool, October 2002. He is a member of a number of international conferences program committees on networking, security, and computer entertainment.

An Adaptive User Interface Based on Spatiotemporal Structure Learning

Hosub Lee, Young Sang Choi, and Yeo-Jin Kim, Samsung Electronics Co., Ltd.

ABSTRACT

We developed a user interface prototype for the Android smartphone, which recommends a number of applications to best match the user's context. To consider the user's context of use, we utilized 5 prototypical variables; time, location, weather, emotion, and activities. The developed system derives the best three recommended applications based on the results of supervised machine learning from such data sets. To consider the history of past context information, in addition to the current one, we developed a novel and effective probabilistic learning and inference algorithm named "Spatiotemporal Structure Learning." By extending Naïve Bayesian Classifier, the spatiotemporal structure learning can create a probability model which represents relationship between time-series contextual variables. We implemented a prototype system which shows the current context and the inferred recommendation of applications. For the prototype system, we developed an Android widget application for the user interface and a Java-based server application which learns structure from training data and provides inference results in real time. To gather training data and evaluate the proposed system, we conducted a pilot study which showed 69 percent accuracy in predicting the user's application usage. The prototype demonstrated the feasibility of an adaptive user interface applied to a state of the art smartphone. We also believe that the suggested spatiotemporal structure learning can be applied to number of application areas including healthcare or energy problems.

INTRODUCTION

Compared to the relatively monotonous user interface and fixed number of application software of feature phones, smartphones allow users to download various application software through market environments such as Apple's app store and Google's Android Market [1]. Although the numerous applications have potential to meet different user's personal needs and wants, sometimes it is difficult to browse and search applications which the user wants to use at a specific time and context.

One approach to tackle this difficulty of application choice is designing user-friendly menu architecture, enabling users to find the most necessary menu or functions more easily [2]. Nevertheless, this approach is mainly intended for not smartphones but general mobile phones which have almost fixed applications and user interfaces. The other strategy is to use folders to categorize similar applications such as games, media, and so on [3]. However, categories have limitations in the convenience of searching applications because it only reduces the pool of choice step by step. Another possible strategy is to change the order of applications based on the recentness or frequency of use of them [4]. Although this simple method of re-ordering by frequently or last visited applications can be helpful in some situations, it cannot be the solution to the user's choice on each different application needs. For example, a user needs to use a weather forecast application when the weather changed to cloudy although her most frequent application is e-mail.

In this article, we report a prototype user interface for an Android smartphone, Samsung Galaxy S [5] which adapts to the user's history of application usages and the context-relevant information such as time, location, weather, emotion, and the user's own activities. To make the user interface adaptive, we applied machine learning approach named spatiotemporal structure learning which considers chronological changes in various contextual variables. Spatiotemporal structure learning is a supervised machine learning algorithm which can formulate a probability model among time-series context data and make inferences about the most suitable applications for the current context. That is, it can automatically learn to recognize historical usage patterns according to the context of the user, and infer the most appropriate applications for the user's current situation.

Adaptive user interfaces for smartphones benefit users by conveniently locating services that the users need and want every time they use the device. We expect the presented technology can provide differentiating values to smartphone manufacturers facing situations in which smartphones become heavily dependent on a few open software platforms and hardware components become commoditized.

 0163-6804/11/$25.00 © 2011 IEEE

Figure 1. *Conceptual diagram of adaptive user interface.*

Related Work

Researchers have studied various methods to facilitate browsing and accessing specific applications on mobile devices.

D. Billsus *et al.* designed the adaptive personalization engine for mobile users [6]. They surveyed existing technologies and concluded that the combination of similarity-based methods, Bayesian methods, and collaborative methods is the best way to extract personalized information, which will be added to predefined user interface of the mobile phone. To enhance personalization effect, however, they took explicit editorial input from the user and used the similarity-based method only to present the personalized user interface.

Y. Fukuzawa *et al.* proposed an automatic menu customization system that ranked menu functions so as to make both frequently and rarely used functions easy to access [7]. They defined the features of each function on a mobile phone via extracting keywords from the manufacturer's manual, and designed the method that ranked the functions based on user operation history by using Ranking SVM (Support Vector Machine).

D. S. Weld *et al.* subdivided the personalization technology into customization and adaptation [8]. According to their definition, customization means a change guided by explicit user requests; on the other hand, adaptation means interface-initiated change in response to a routine user behavior. They surveyed a decade's research on the personalization and concluded that both customization and adaptation are required to satisfy user's need for personalization. Furthermore, they mentioned that user interfaces should automatically adapt to the capabilities of the device at hand, to network connectivity, and to the user's activities, location, and context. However, if users wish to provide customization guidance, user interfaces should provide ways to control adaptation process to the user at any level.

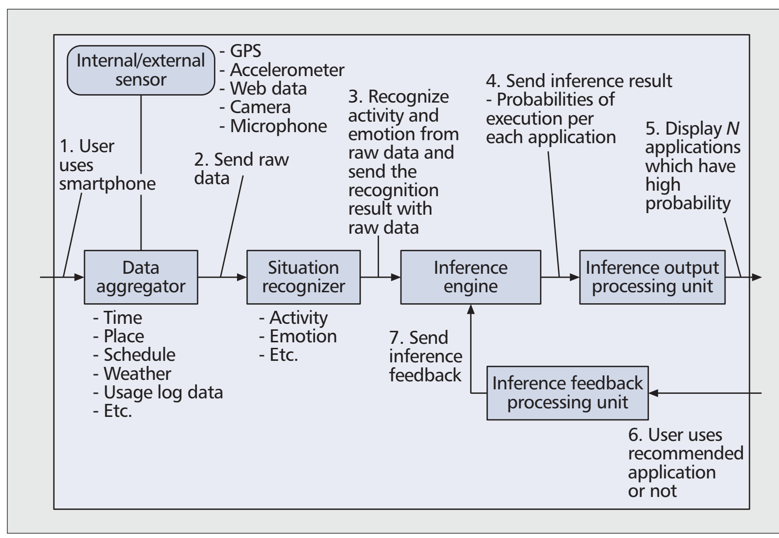


Figure 2. *Architecture of adaptive user interface. From raw sensor readings, activity and emotion of user are extracted and other raw data such as time, location, and weather are gathered. Then, all of them sent to the Inference Engine for inferring recommendable applications.*

The Adaptive User Interface

Overview

The proposed adaptive user interface can recommend applications of a smartphone which have highest chance to be executed by the user. The probabilities of execution of applications are calculated by the spatiotemporal structure learning algorithm which is explained in the subsection "The Interference Engine." The system learns the application choices based on five contextual variables with the algorithm. Therefore, even if the user interfaces of all users' smartphones are initially identical, the user interfaces are gradually adapted to each user as users use them. Figure 1 shows high-level concept of the adaptive user interface.

The proposed adaptive user interface mainly

IEEE Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

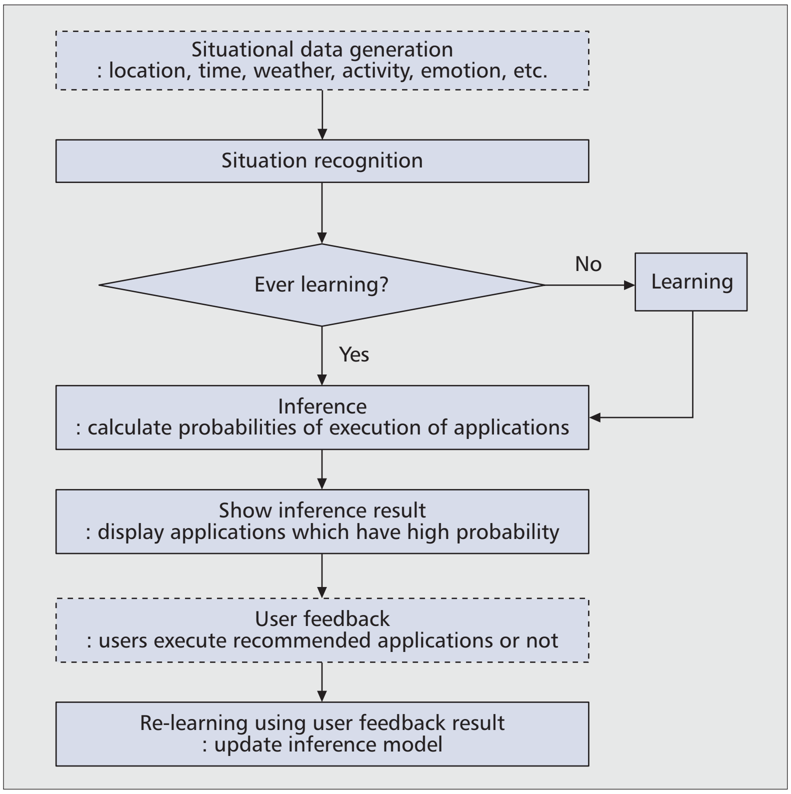


Figure 3. *Flow chart of adaptive user interface. This flow chart explains the overall process from the gathering of sensor data to (re-)learning and inference process. The part shown in dotted line means the interaction performed outside the system.*

consists of the 'Situation Recognition' process and the 'Learning & Inference' process. The overall architecture including these two main processes is explained in the following sub-section.

System Architecture

A block diagram of the adaptive user interface is shown in Fig. 2. Data Aggregator sub-module collects various usage data from internal and external sensors: the internal sensors collect various sensor data or information like the present weather gathered through running software in smartphones; the external sensor assembles data from outside of the device using camera or microphone. Then, Data Aggregator sub-module delivers these data to Situation Recognizer sub-module.

By analyzing 3-axis accelerometer data, the Situation Recognizer sub-module classifies the current activity of the user into four classes; walk, stop, running, and riding in a car. Additionally, the Situation Recognizer sub-module extracts emotional state of the user. In the current implementation, we asked users to manually change the emotional state and we will apply emotion recognition technology via facial image and voice analysis [9] in the future research. Then, the Situation Recognizer sub-module delivers recognized activities and emotion of user with other self-explanatory raw data such as time, location, and weather to the Inference Engine.

The Inference Engine infers applications that the user most need at the moment from the context data delivered by the Situation Recognizer sub-module. Next, the Inference Engine sends the list of recommended applications to the Inference Output Processing Unit sub-module. Then, the Inference Output Processing Unit sub-module displays these applications to user. Finally, the Inference Feedback Processing Unit sub-module examines whether the user executes recommended applications or not. To improve the accuracy of inferences, the Inference Feedback Processing Unit sub-module sends back this result to the Inference Engine. That is, the Inference Engine can continuously (re-)learn chronologically delivered context data from both the Situation Recognizer and the Inference Feedback Processing Unit sub-module. The overall process of adaptive user interface is depicted as Fig. 3.

Details of learning and inference process of the Inference Engine are explained in the following sub-section.

The Inference Engine

The main function of the Inference Engine is to receive the context data and provide three identifiers of applications which have the highest conditional probabilities of execution under the current context of the user. To infer applications appropriate to the user's preference and situation, the Inference Engine builds a personalized inference model by learning the context data of a user. The inference model is a type of a naive Bayes probabilistic model with strong condition-

Time Order	Time	Location	Weather	Emotion	Activity	Application	Probability
t (Present)	**Morning**	**Home**	**Cloudy**	**Neutral**	**Stop**	**News**	**78%**
						Music	14%
						Web	8%
t-1	Morning	Office	Cloudy	Negative	Walk	...	...
t-2	Afternoon	Office	Sunny	Positive	Stop	...	...
t-3 (Past)	Afternoon	Office	Sunny	Positive	Walk	...	...

Table 1. *Example of conditional probability table. Shaded cells show possible spatiotemporal node (Morning-Home-Cloudy-Neutral-Stop) and application (news) which has the highest execution probability.*

al independence assumptions [10]. We used five types of context data; time, location, weather, emotion, and activities, which are constantly collected in time order.

Figure 4 shows three functional steps of the Inference Engine: data processing step, learning step and execution step. To generate the user's inference model, the Inference Engine learns the pattern of application usage by discovering the relational structure among the collected context data of user.

First, at the data processing step, the data acquisition module collects and pre-processes context data. Preprocessing includes conversion process from raw data to CSV (Comma-Separated Value) file format. As the preprocessed data is transferred, the expansion module generates spatiotemporal data with the spatiotemporal order N which refers to the number of recent contexts including the current one. Given the spatiotemporal order 3, for example, the expansion module arranges 15 contextual variables related to five sensors and three time points (t, t-1, t-2). And, the expansion module generates total 575 spatiotemporal nodes by combination ${}_{15}C_1$, ${}_{15}C_2$, and ${}_{15}C_3$ among 15 contextual variables.

Then, the expansion module makes an initial inference model in which the spatiotemporal nodes are linked to identifiers of applications and each node has a conditional probabilistic table. The conditional probability table shows usage probabilities of executable applications for each spatiotemporal node which means one of the possible combinations of contextual variables of the user. Example of the conditional probability table is shown in Table 1.

The initial inference model is similar to Naïve Bayesian Classifiers, but the spatiotemporal data in our model are different from them in that

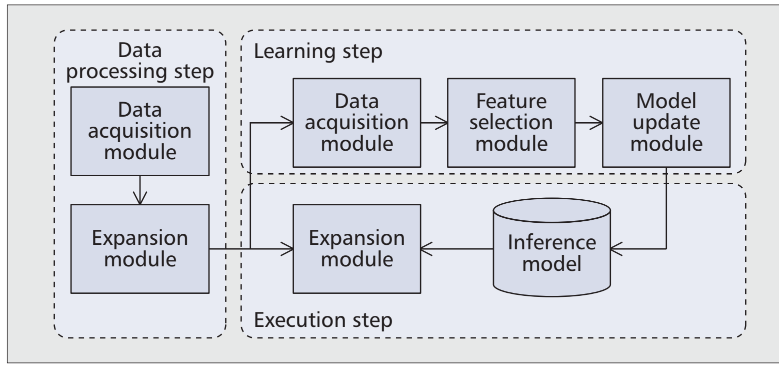


Figure 4. *Functional flows of inference engine. There are three major procedures in inference engine: data processing step, learning step, and execution step.*

each spatiotemporal node includes a time (order) property. In other words, spatiotemporal node can consider changes in sensor data over time. Then, we can analyze each data such as location, weather, emotion, and activities in both space and time domain. As a result, we get an initial inference model spatiotemporally considering context data through the expansion module.

At the learning step, the learning module learns all conditional probability tables of the initial inference model with the spatiotemporal data based on the Bayes theory [11]. Next, the feature selection module selects the meaningful relationships that have higher conditional probabilities than a given the threshold. As a result, a new model structure only containing these important relationships is generated. Finally, the model update module replaces the existing model with the new one.

Then, when specific input data are entered,

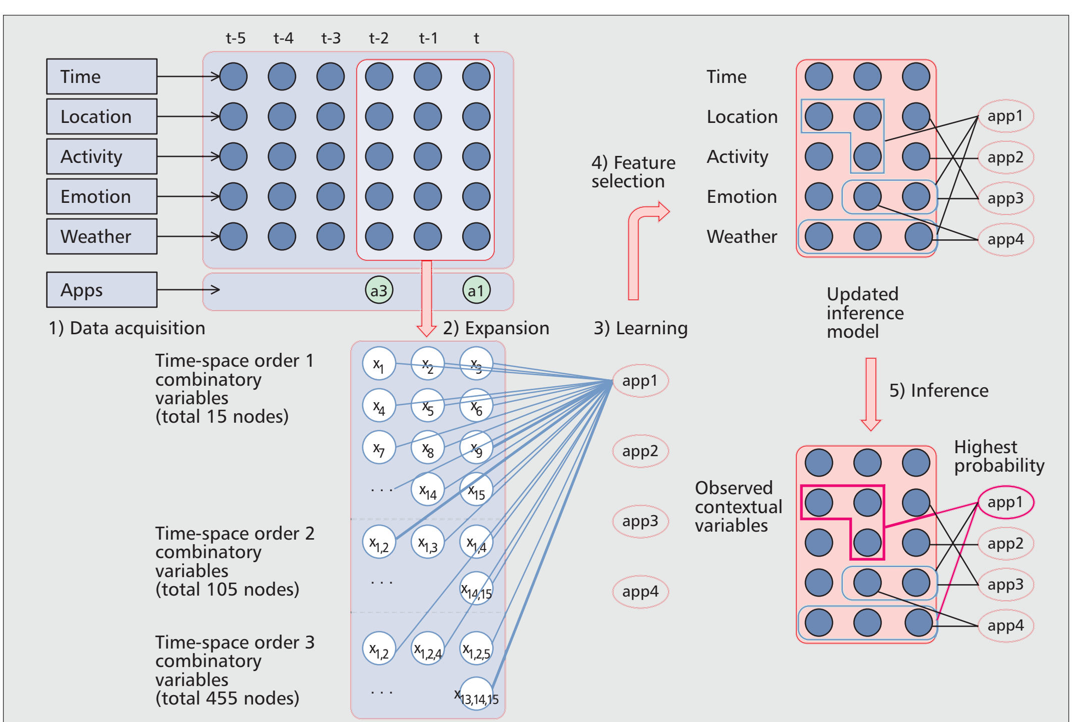


Figure 5. *Visualized process of inference engine.*

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

Item	Language/SDK	Used Library	Target Device
Client	Java/Android 2.1	Android API Level 7, Google Weather API	Samsung Galaxy S
Server	Java/JDK 6	Weka 3.6.2	Dual-core CPU PC

Table 2. *Implementation Detail. Used sensors or modules in the Samsung Galaxy S for recognizing user contexts: 3-axis accelerometer (activity), GPS (location), 5.0 Megapixel AF camera and microphone (emotion).*

Figure 6. *Adaptive user iinterface loaded on Samsung Galaxy S. Contextual variables: location, emotion, activities, weather, and time; three recommended applications. If there is no application to execute, user can press the More button to search all applications installed on the smartphone.*

inference module calculates each probability of execution of each application based on the updated inference model.

Figure 5 shows summary of the overall process performed by the Inference Engine. Brief descriptions are as follows.

1 Contextual data are gathered by five sensors in time order. (Data Acquisition)
2 If we specify time order is 3, 15 context data are gathered and it will be expanded to 575 spatiotemporal context nodes. Then, by linking these nodes to applications, initial inference model is built. (Data Expansion)
3 Next, the Inference Engine learns conditional probabilistic table which means relationships between each pair of spatiotemporal node and application. (Learning)
4 Among these relations, the Inference Engine can selectively choose some meaningful relations by threshold-based filtering. By this process, we can get updated inference model. (Feature Selection)
5 When specific input data are entered, the Inference Engine calculates the probability of execution of each application based on the updated inference model. Finally, the Inference Engine infers some applications, which have high probabilities as a recommendation result. (Inference)

IMPLEMENTATION

We designed and developed a research prototype of the adaptive user interface as client-server architecture. We implemented a client user interface on a Samsung Galaxy S which has various built-in sensors; on the other hand, we implemented the Inference Engine as a Java-based server program using Weka library [12]. Weka is a popular library of Java routines for various machine learning tasks and we used it to build Situation Recognizer and Inference Engine on the server. Also, client and server communicate each other via TCP/IP protocol over Wi-Fi connection. Table 2 shows implementation detail of client and server.

Client acts as the main user interface and communicates with the server. Client sends request for inference to the server when any contextual change occurs to a user while monitoring the user. Then, server performs two principal functions as follows.

1 "Recognize" the current activity of the user by analyzing 3-axis accelerometer data in real time
2 "Infer" recommended applications by the request of client

When client receives the inference result from server, the user interface showing recommended applications is dynamically reformulated. Figure 6 shows the adaptive user interface implemented on the Samsung Galaxy S.

In this article, we proposed a client-server architecture in which our learning and inference technology resides on the server that interacts with the widget application on the smartphone. As processing power and memory of mobile devices rapidly increase, we plan to integrate functions of the server into the client side.

USER STUDY

When the user turns or touches on the smartphone, the Inference Engine infers the most probable applications that the user is going to use in the current time. In this process, the sensor data, which are periodically collected, are transformed to the spatiotemporal data through the data acquisition module and the expansion module. Given the spatiotemporal data, the inference module calculates the conditional probabilities of the available applications and recommends three applications with the highest probabilities.

Time	Location	Mood	Weather	Physical Activity	Business	iPhone App Categories
Morning	Home	N	Clear	Stop	Private	Mail
Morning	Home	N	Clear	Stop	Private	Weather
Morning	Home	N	Clear	Stop	Private	Social
Morning	Home	N	Clear	Stop	Private	Mail
Morning	Home	N	Clear	Stop	Private	News
Morning	Home	N	Clear	Stop	Private	News
Morning	Company	N	Clear	Stop	Private	Mail
Morning	Company	N	Clear	Stop	Private	e-Book
Morning	Company	N	Clear	Stop	Private	News
Morning	Company	N	Clear	Stop	Biz	Web
Afternoon	Company	N	Clear	Stop	Private	Mail
Afternoon	Company	N	Clear	Stop	Private	Social
Afternoon	Company	N	Clear	Stop	Private	Call
Afternoon	Company	N	Clear	Stop	Private	AppEtc
Afternoon	Company	N	Clear	Stop	Private	iPod
Afternoon	Company	N	Clear	Stop	Private	Util
Afternoon	Company	N	Clear	Stop	Private	Education
Afternoon	Company	N	Clear	Stop	Private	Mail
Afternoon	Company	N	Clear	Stop	Biz	Education
Afternoon	Company	N	Clear	Stop	Private	News
Evening	LocEtc	P	Clear	Walk	Private	Mail
Evening	LocEtc	P	Clear	Walk	Private	e-Book
Evening	LocEtc	P	Clear	Walk	Private	iPod

Table 3. *Training data of one participant. The participant is a man about forty. He is married and works at a research facility. He is an experienced user of smartphone and these data are gathered through an iPhone.*

With the built inference structure, we validated the accuracy of application recommendation with the training data. The resulting accuracy was 69 percent on average. We defined accuracy as a case in which the user actually used an application among the three recommended applications.

Two subjects from our research group members (one male in his 40s and one female in her 30s) were recruited for the pilot study. They collected their usage of applications and contexts such as time, location, weather, emotion, and activities for two days. A total of 163 usage data (83 cases for the male and 80 cases for the female) were gathered via paper-based questionnaire and utilized as the training data for the Inference Engine. Gathered context data are organized as follows.

- Time (3): morning, afternoon, evening
- Location (3): home, company, etc.
- Activity (4): stop, walk, run, riding in vehicle
- Emotion (3): positive, negative, neutral
- Weather (3): sunny, cloudy, rainy

The list of executed applications per each combination of these context data is accumulated in a chronological order. Then, the Inference Engine would be learned using this training data. A sample of the training data is shown in Table 3.

With the built inference structure, we validated the accuracy of application recommendation with the training data. The resulting accuracy was 69 percent on average. We defined accuracy as a case in which the user actually used an application among the three recommended applications. We expect the accuracy will be

We plan to gather real-world context data from participants and apply it to our adaptive user interface. We believe that the proposed technology can be applied to other domains such as healthcare or energy.

enhanced as more data are collected in on-going research.

Conclusion and Future Work

We proposed the architecture for an adaptive user interfaces and implemented a prototype for smartphones. To implement the adaptive user interface, we developed a novel machine learning and inference algorithm named spatiotemporal structure learning. This algorithm was designed based on the Naïve Bayesian Classifier and it can create dependency structure between various time series data, which represents the contexts of the user. We tested this algorithm with training data collected by two participants. By learning and adapting to application usage history of participants, the prototype user interface can effectively recommend applications with acceptable accuracy.

We are revising our system as a client-only architecture in which user interface and machine learning module are integrated into one place. Furthermore, we are improving the accuracy of the inference through parameter learning using more training data from more than 10 participants. Finally, we plan to gather real-world context data from participants and apply it to our adaptive user interface. We believe that the proposed technology can be applied to other domains such as healthcare or energy.

Acknowledgement

The authors would like to thank all those who provided the training data we used, and Byung Chull Bae and Yun Gyung Cheong for useful discussions.

References

[1] Google's Android Market, http://www.android.com/market/.

[2] J. Beck, S. H. Han, and J. Park, "Presenting a Submenu Window for Menu Search on a Cellular Phone," *Int'l. J. Human-Computer Interaction*, vol. 20, issue 3, July 2006, pp. 233–45

[3] iPhone Folders: A New Way to Organize Your Apps, http://www.apple.com/iphone/features/folders.html

[4] A. Sears and B. Shneiderman, "Split Menus: Effectively Using Selection Frequency to Organize Menus," *ACM Trans. Computer-Human Interaction*, vol. 1, no. 1, 1994, pp. 27–51.

[5] Samsung Galaxy S, http://galaxys.samsungmobile.com/.

[6] D. Billsus et al., "Adaptive Interfaces for Ubiquitous Web Access," *Commun. ACM*, vol. 45, issue 5, May 2002.

[7] Y. Fukazawa et al., "Automatic Mobile Menu Customization based on User Operation History," *Proc. 11th Int'l. Conf. Human-Computer Interaction with Mobile Devices and Services, Industrial Case Studies*, no. 50, 2009.

[8] D. S. Weld et al., "Automatically Personalizing User Interfaces," *Proc. 18th Int'l. Joint Conf. Artificial Intelligence*, 2003, pp. 1613–19.

[9] C. Busso et al., "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information," *Proc. 6th Int'l. Conf. Multimodal Interfaces*, Poster session 1, 2004, pp. 205–11.

[10] D. Lowd and P. Domingos, "Naive Bayes Models for Probability Estimation," *Proc. 22nd Int'l. Conf. Machine Learning*, 2005, pp. 529–36.

[11] J. Hartigan, *Bayes Theory*, Springer-Verlag, 1983.

[12] Weka 3, http://www.cs.waikato.ac.nz/~ml/weka/.

Biographies

Hosub Lee (horus.lee@samsung.com) is a research staff of intelligent computing lab at Samsung Electronics. His research interests include ambient intelligence, affective computing, and human computer interaction (HCI). Lee received a B.S. in computer science in 2005 from Korea University, Korea and M.S. in computer science and engineering in 2007 from Seoul National University, Korea.

Young Sang Choi (macho@samsung.com) is a senior researcher of Samsung Advanced Institute of Technology (SAIT). He is researching machine intelligence and affective computing for future information devices. He received his Ph.D. in Industrial and Systems Engineering at the Georgia Institute of Technology in 2009, studying on human robot interaction in healthcare applications. He received his M.S. and B.S. in Industrial Engineering from Seoul National University in 1999, 1997 respectively.

Yeo-Jin Kim (yeoj.kim@samsung.com) is a research staff of the Intelligent Computing Lab. at Samsung Electronics. Her research interests concentrate in machine learning, artificial intelligence, and interactive storytelling. She received a B.S degree in Computer Science in 2002 and a M.S. degree in Computer Science in 2005 both from Yonsei University, Korea. From 2004–2005, she was a research student at the Department of ECE, University of Illinois at Urbana-Champaign.

Synchronized Multimedia Streaming on the iPhone Platform with Network Coding

Péter Vingelmann, Budapest University of Technology and Economics and Aalborg University
Frank H. P. Fitzek, Morten Videbæk Pedersen, and Janus Heide, Aalborg University
Hassan Charaf, Budapest University of Technology and Economics

ABSTRACT

This work presents the implementation of synchronized multimedia streaming for the Apple iPhone platform. The idea is to stream multimedia content from a single source to multiple receivers with direct or multihop connections to the source. First we look into existing solutions for video streaming on the iPhone that use point-to-point architectures. After acknowledging their limitations, we propose a solution based on network coding to efficiently and reliably deliver the multimedia content to many devices in a synchronized manner. Then we introduce an application that implements this technique on the iPhone. We also present our testbed, which consists of 16 iPod Touch devices to showcase the capabilities of our application.

INTRODUCTION

Multimedia content distribution has received a lot of attention lately in the mobile world. New ways to convey multimedia content to mobile devices are discussed after the recent failure of DVB-H and DVBM. Besides the bare technology there is also the question how mobile users are looking at multimedia content. So far the main architecture was designed such that the overlay network with its highly centralized architecture is providing the content, and the mobile users are consuming it. But more and more users are starting to generate and collect their own content that they would like to distribute among each other in local area networks.

Apart from this discussion in this work we investigate the possibility of sending multimedia content from one device to many devices in closer proximity. In our previous work we have shown that it is possible to share photos and audio files among mobile devices even across different platforms [1]. Therefore, in this work we address mobile video as the next logical step.

Synchronized video playback can be used among friends to show their latest videos to each other. Exchanging music videos is especially interesting at social events if everybody can play them at the same time. Another fascinating application is for home entertainment: we can deploy a simple server that broadcasts a live video stream (e.g. a sporting event) that is accessible on every mobile device in the household.

In this article we not only present a mobile application supporting the described use cases, we also advocate the use of network coding in order to address the channel characteristics of wireless networks and the limited energy of mobile devices.

SHORTCOMINGS OF EXISTING SOLUTIONS

There are several applications that can stream multimedia content to the iPhone, for example AirVideo and TVersity. Basically these applications run a webserver to which the iPhone media player can connect. A TCP connection is established and the player issues standard HTTP range requests, then the webserver sends raw file data with HTTP headers in response. This approach has the clear drawback that with an increasing number of receivers the bandwidth of a given cell or access point will become the bottleneck due to the use of unicast connections.

In order to prove this, first we used a single iPod Touch to connect to an AirVideo server running on an iMac to play a video that was previously transcoded to a suitable format for the iPhone platform (Xvid and AAC codecs with an overall data rate of 500 kb/s). The video playback was fine, so this approach is sufficient for a single device playing a single media file. As the next step, we tried the same experiment with five iPod Touches as receivers, and we often experienced stuttering in the video playback. We also monitored the network traffic with Wireshark running in promiscuous mode.

The server sending rates in the two experiments are shown in Fig. 5 as captured by Wire-

0163-6804/11/$25.00 © 2011 IEEE

shark. We observed that the overall throughput reached 16 Mb/s in a few seconds, and it remained this high for several hundred seconds. We can conclude that the video playback on the iPod Touches was not satisfactory due to the insufficient incoming data. Thus if we connect to the same server with multiple devices at the same time, we can quickly saturate the wireless network. This is an inherent drawback of unicast connections.

If we intend to efficiently deliver the same content to several devices, multicast transmissions provide a favorable solution. In [2] it was shown that multicast video streaming is feasible on the iPhone platform with network coding. The limitation of this work was the lack of synchronized playback and the reliance on jailbroken components on the iPhone. Moreover, the architecture design was limited to point-to-multipoint communication, whereas network coding implies the possibility of recoding packets at the intermediate nodes. This feature is particularly useful if the source and the receiver do not have a direct link, that is we have a multihop network.

Scenario and Solution

In our scenario a server *S* wants to reliably transmit the same media file to several nearby receivers, $t_1, t_2, \ldots, t_N$, via a wireless link. This basic scenario is depicted in Fig. 1. As mentioned earlier, the traditional point-to-point data distribution paradigms (i.e., unicast transmissions) provide poor utilization of the available network resources for one-to-many services. Since all receivers are interested in the same content, we can efficiently utilize the wireless channel with broadcast transmissions.

Under ideal channel conditions all broadcast packets are delivered to all nodes simultaneously. In real-life wireless networks packet losses frequently occur [4]; thus, some sort of retransmission is necessary to ensure reliability (i.e., to correct packet losses at the receivers). A simple solution would be that the individual nodes request all missing packets from the original source. This would imply that every lost packet is transmitted again, and if packet losses are uncorrelated, most retransmissions will not be useful to many receivers since they have received those packets in the first place. To put it differently, it is likely that a single retransmission will only benefit a single receiver.

A shrewd way to maximize the impact of each retransmission is by using network coding [3, 4]. Researchers have shown that network coding can provide several advantages: improved throughput, robustness, security, and lower complexity in communication networks [5].

Network Coding

Network coding differs from channel or source coding, because it is not limited to end-to-end communication, but allows on-the-fly recoding of information whenever needed. Another important fact is that network coding breaks with the store-and-forward policy of existing communication systems. It has been widely accepted that in packet-based communication networks, all packets that enter a node will also leave the node in

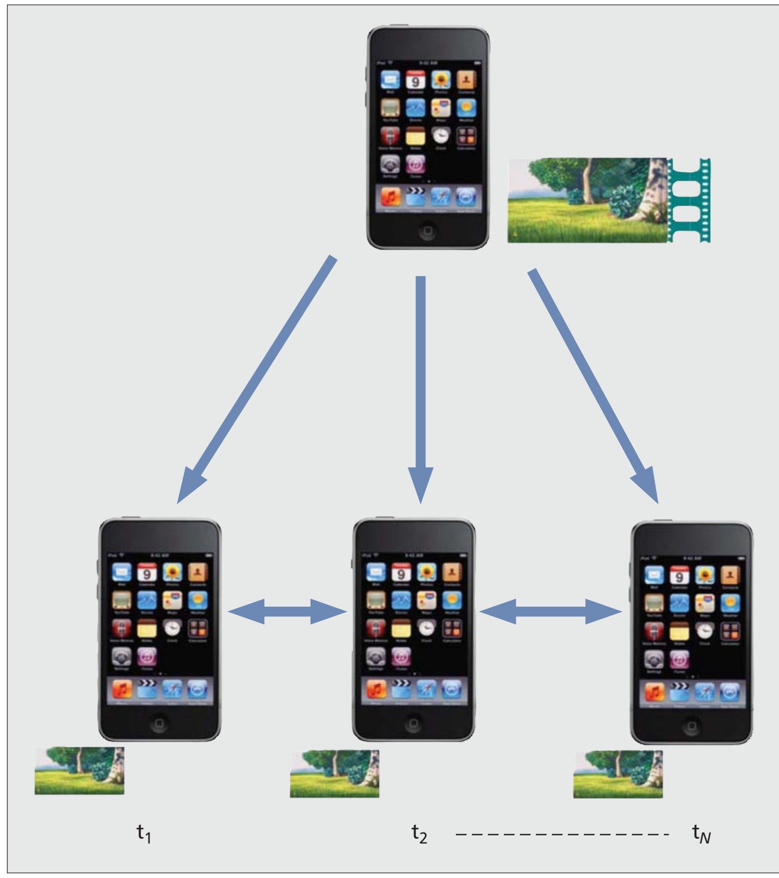


Figure 1. *A server* S *transmitting data to* N *receivers* $t_1, t_2, \ldots, t_N$.

one way or another (packet drops due to buffer overflow are neglected for ease of illustration). In contrast, a network coding-enabled communication node is able to recode incoming data to tailor it to the needs of the outgoing channels. Network coding was introduced by Ahlswede *et al.* in [3] for fixed networks, and it was adapted by other researchers for wireless and mobile networks [6].

Figure 2 gives a basic overview of the operations performed in a network coding system. If we intend to encode a large file, it should be split into several chunks, also called generations, each consisting of *g* packets [4]. Otherwise, the computational complexity of the encoding and decoding operations would be prohibitively high.

The top component in Fig. 2 is the encoder that generates and transmits linear combinations of the original data packets in the current generation. Addition and multiplication are performed over a Galois field; therefore, a linear combination of several packets will have the same size as a single packet. Note that any number of encoded packets can be generated for a single generation. The middle layer in this system is the wireless channel, where packet erasures may occur depending on the channel conditions. The network nodes receive a series of encoded packets that are passed to the decoder (the bottom component in the figure), which will be able to reconstruct the original data packets after receiving at least *g* linearly independent packets.

Recoding is an additional operation of net-

IEEE Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

An obvious benefit of using network coding is that a network node is no longer required to gather all data packets one by one; instead, it only has to receive enough linearly independent encoded packets. This is important for our scenario.

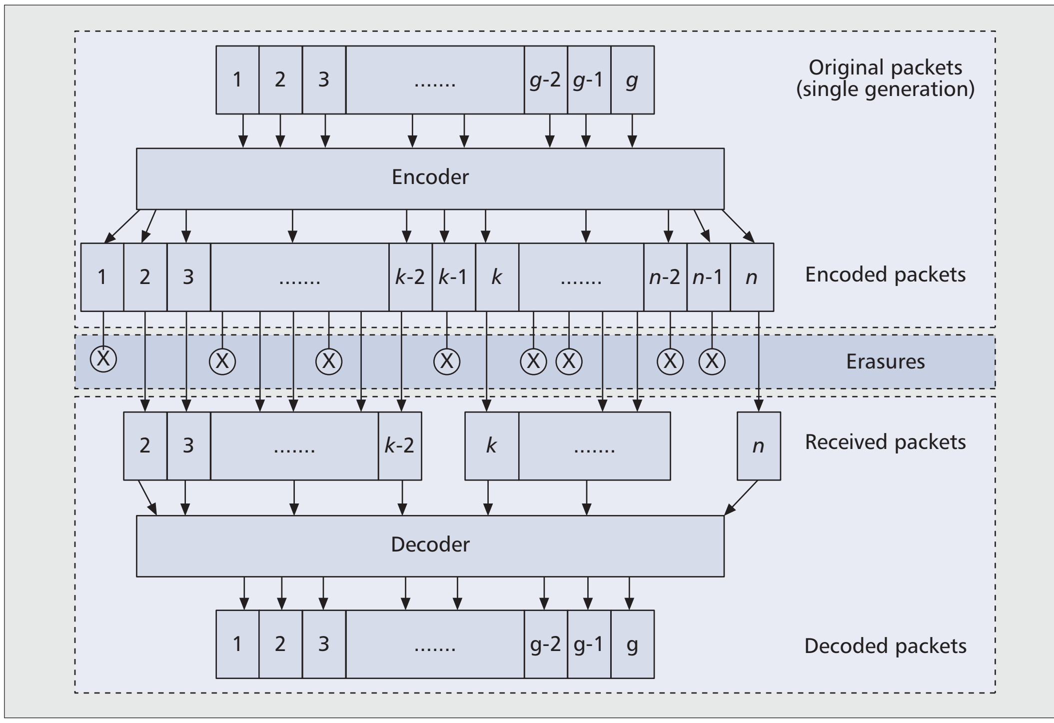


Figure 2. *Overview of network coding.*

work coding that is not directly shown in this figure. Recoding means that all network nodes are allowed to generate and send new encoded packets (i.e., new linear combinations of the packets they have previously received).

An obvious benefit of using network coding is that a network node is no longer required to gather all data packets one by one; instead, it only has to receive enough linearly independent encoded packets. This is important for our scenario since we can simultaneously serve multiple nodes with a single transmission by sending a linear combination instead of choosing a specific packet.

As does any other coding operation, network coding involves computational overhead, which might be prohibitive from a practical point of view. The authors in [7] proposed an efficient solution for mobile devices. We can use a systematic code to ensure reliability with low overhead. It is not necessary to always send encoded packets while using network coding. Uncoded packets can be considered primitive linear combinations, and thus can be processed by the decoder. A simple and efficient method is to transmit each generation in two stages. In the first stage, the source transmits all packets uncoded. Each of these packets will contain new useful information for the individual receivers. In the second stage, the source will generate and send random linear combinations of the original data in order to correct packet losses, which have occurred during the first stage. Note that a single encoded packet can potentially correct different losses at different nodes. With this approach we can maximize the number of nodes for which a packet is useful.

In order to illustrate the advantage of network coding for the envisioned use case, we show Fig. 3, where the leftmost mobile device would like to share a video file with three other devices in close proximity. One of the three receivers has a larger distance to the originating device. Based on the distance, the packet error rate differs significantly. Here we assume that the first-tier neighbors have a 10 percent loss rate, while the second-tier device has a packet error rate of 50 percent. Those values are reasonable and were reported in [8]. If communication is only allowed on the blue links (i.e., the originating device is transmitting packets), the overall time until all devices receive the file depends primarily on the second-tier neighbor.

If we enable multihop relaying (the red links in Fig. 3), the first-tier neighbors can also try to forward packets to the second tier. The problem is that the relaying devices need to be coordinated in order to not convey redundant information (i.e., to prevent the same packet being relayed twice). Here network coding enables the relaying devices to recode previously received packets in such a way that redundant information is minimal.

IMPLEMENTATION

This section discusses the implementation of the application based on the ideas outlined above. Our primary target platform is the iPhone, where the Objective-C language is mandatory for graphical user interface (GUI) development, but we chose to write most of our application in C++ in order to facilitate its porting to other mobile platforms. GCC 4.2 is the default internal compiler in the Xcode development environment, and it can be used to compile C++ source files in Objective-C++ mode and link with the generated object files. Consequently, the GUI that has to be written in Objective-C can call regular C++ code, and a high degree of platform independence can be achieved. Note that we used iPhone OS v. 3.1.3 for development.

STREAMING

When the streaming server starts, it enumerates its network interfaces in order to find out its IP and broadcast address corresponding to the wireless network interface.

Upon user request it opens the selected video or audio file, then determines its media type and overall size. The packet payload size is set to 1024 bytes and generation size is 64. Upon creating a new data stream, the server reads a data chunk (64 kbytes) from the file to fill the input buffer for the first generation. It also calculates the total number of generations based on the overall file size. Then it begins to send uncoded packets and some metadata with a specified data rate, which is slightly higher than the native data rate of the media file. All packets are sent to the broadcast address of the wireless interface. After sending 64 uncoded packets, the server transmits several encoded messages in order to repair all packet losses of the individual receivers. The actual number of these extra encoded packets can be adjusted based on the current network conditions. Of course, this approach would require periodic feedback from several receivers. A simple solution is to always add a large overhead (e.g., 50 percent) to combat packet losses even under the worst conditions. After sending a specific number of encoded packets, the server moves on to the next generation. It fills its input buffer with a new data chunk that is read from the input media file. Then it begins to broadcast uncoded (and later encoded) packets for the current generation, and this process continues until we reach the end of the input file.

Note that the streaming server can easily be integrated into the client application to enable users to stream multimedia content from their mobile devices.

PLAYBACK

The most important question is how to play the incoming media stream on the iPhone. The built-in media player (an instance of the MPMoviePlayerController class) can play a local file in the application bundle or open an HTTP network stream at a given URL. We intend to initiate playback when only a small part of the entire file is received, and this media player cannot play an incomplete file because it tries to buffer up a significant amount of data in the beginning. Using any other media player is not recommended by Apple, so the only solution is to run a web server on localhost that can continuously feed the incoming data into the player itself.

The media player issues standard HTTP range requests, which will be processed by the embedded web server in our application. Each range request is served by a new POSIX thread. The range in the first request is always 0-1, which means the first 2 bytes of the file. The response will contain the size of the whole file, and the media player uses this information and the media file header to issue further range requests to the web server. It can decide to close the current socket before the actual range request is fully served, and the web server must be able to detect this behavior in order to avoid sending packets that will not be processed by the player.

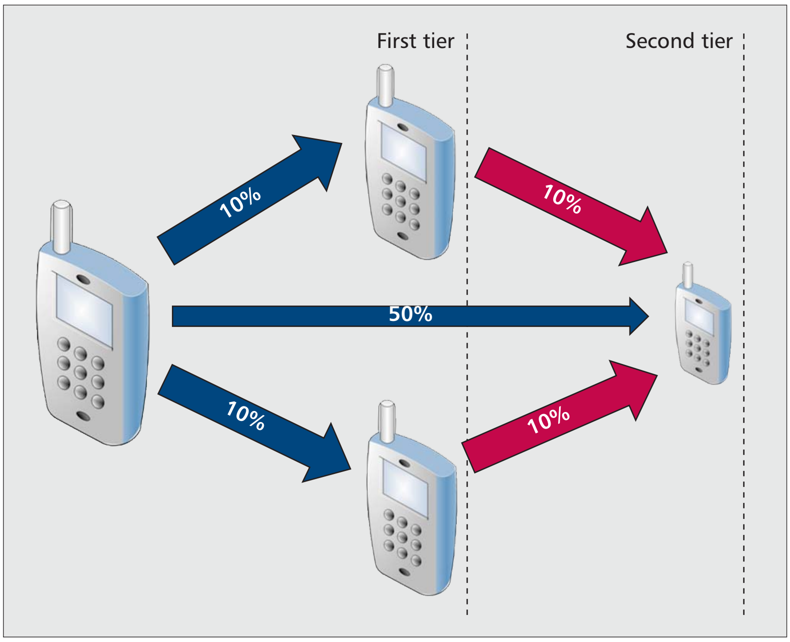


Figure 3. *Using network coding in a two-tier topology.*

The web server feeds the media player with raw data of a generation when it is completely decoded. Thirty seconds after transferring all bytes from a generation, it will be considered obsolete, and its data buffer will be deleted from memory. This way we can avoid memory leaks, which would quickly accumulate in our application during the playback of long media files.

The built-in media player can only play video files that use the QuickTime (.mov) container format and H.264 or Xvid video codecs with a suitable resolution for the iPhone/iPod Touch. Therefore, the usual AVI videos must be transcoded before streaming. This can be performed with open-source tools like ffmpeg.

The media player is configured with the control mode option Volume Only so that users cannot seek arbitrary offsets in the media stream. It is important to point out that our application has no control over the media player component after the playback is started. The player cannot be paused or stopped; there are only two callback functions that signal the host application when the media file is preloaded and when the playback is finished.

Starting the player imposes a significant load on the CPU, and during this period (1–2 seconds) a large number of incoming packets are lost on the receivers. A simple way to overcome this issue is to stop the streaming server for 2 seconds after sending out the very first packet. This will give time for the built-in media player to properly load up, and no packets will be lost on the client devices.

Note that the video and audio codecs are able to recover if some dirty data is transferred to the player. This can happen under extreme channel conditions, when the overhead sent from the server is not enough to repair all pack-

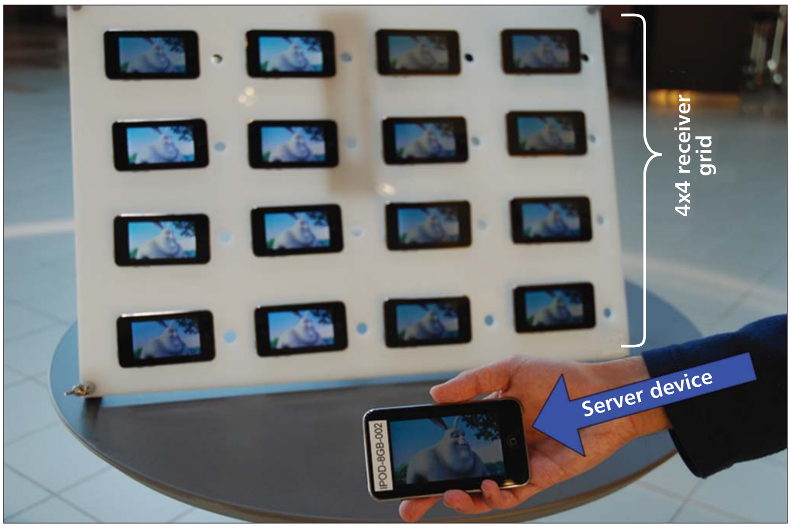


Figure 4. *The testbed consists of 16 third-generation iPod Touches, and the video is streamed from another iPod Touch.*

Processor	ARM CORTEX-A8
Clock rate	833 MHz (underclocked to 600 MHz)
Memory	256 MB DRAM
Flash memory	32 or 64 Gbytes
Display	320 × 480 px, 3.5 in
Playback time	Video: 6 hours, audio: 30 hours

Table 1. *Technical specifications of the 3rd generation iPod Touch.*

et losses of the receivers. In this case we typically observe green boxes on screen and hear some audio glitches, but the playback continues and the errors disappear when channel conditions return to normal.

Synchronization

The playback synchronization among multiple devices is based on the assumption that every client device decodes each generation at approximately the same time. Slight variations are possible due to the differences in propagation and processing delay and due to different packet erasures. The cumulative effect of these phenomena translates to a variation that is always smaller than 0.05 seconds. On the other hand, the startup of the media player can cause much bigger deviations.

As mentioned before, a generation cannot be fed to the player until it is completely decoded. Each generation is protected by a mutex that is initially locked by the main application thread. When the generation is decoded the mutex is unlocked, and the threads that are currently serving the HTTP range requests can proceed with reading the raw data of the decoded generation.

It is crucial that every player starts the playback at the same time. If a player starts later than the others, it will not be able to catch up, even though it decodes all subsequent generations at the right time. On the other hand, if a player starts early, it will resynchronize with the others (i.e., it will be stalled), since it cannot decode generations earlier than the others.

The synchronization can be further improved by performing explicit clock synchronization among the individual receivers. For example, all receivers can synchronize with the server's clock, and specific timestamps can be used to precisely control the moment when the generations are unlocked. Nevertheless, such high precision is not required in our setup, and its implementation would involve additional overhead.

Testbed

We have assembled a testbed to demonstrate the capabilities of our application. The receiver grid consists of 16 third-generation iPod Touch devices, and can be seen in Fig. 4. The iPod Touch technical specifications are shown in Table 1.

For demonstration purposes, we chose the 10-minute-long "Big Buck Bunny" animation movie that was released under the Creative Commons Attribution 3.0 license. The original HD movie was transcoded to a resolution of 480 × 320 pixels using the Xvid codec for video (bit rate: 372 kb/s) and the MPEG-4 AAC codec for audio (bit rate: 128 kb/s). The transcoded video can be played on the iPhone/iPod Touch using the built-in media player with a smooth playback and fine quality. In [1] the full-blown testbed is shown while playing this video using our application.

The iPod Touches were connected to an ad hoc wireless network that had been created using a Nokia N95 mobile phone, as the iPhone OS does not support the creation of ad hoc networks. The video stream was sent from another iPod Touch that was running our application in server mode. Although the streaming server has been integrated into our application, unfortunately it cannot access the video and audio files stored on the iPhone due to Apple's limitations. An application can only access files within its own application bundle. At the moment, the only solution is to store all media files we want to share in the application bundle.

In Fig. 5 we present the bandwidth usage of our streaming server compared to the unicast solutions mentioned earlier. All plots were generated by Wireshark during the playback of the same video file. As seen on these plots, the bandwidth usage is constant and never exceeds 100 kbytes/s, which is significantly lower than the others. Note that adding additional receivers will not have a negative impact on the performance of our application, since we only use minimal feedback from the clients. Moreover, the low bandwidth usage infers that streaming multiple videos at the same time might be a feasible idea. In theory, 10 simultaneous video channels can be broadcast if we assume a net throughput of 10–16 Mb/s in a typical wireless network.

FUTURE WORK

The application can be extended in the future to support user cooperation and multihop ad hoc networks.

The number of extra encoded packets sent by the server can be significantly reduced by using cooperation among the receivers. Assuming that the packet losses are uncorrelated, the cooperating devices can exchange missing packets with each other; thus, the server can send less overhead. It has been shown in [9] that using network coding in such a cooperative cluster is an efficient way of realizing packet exchange. The clients simply broadcast recoded packets (for the current generation), which most likely convey information that is useful to the other receivers in case of uncorrelated packet erasures. However, if we implement this approach, precautions must be taken to avoid massive collisions in the network. We can use carefully chosen backoff timers in the client applications to solve this issue. For example, these timeouts can be adjusted so that the receiver with the most packet losses is given priority to request missing packets from the others. Then the node with the most knowledge will reply first with a series of recoded packets.

So far we have only considered single-hop ad hoc (IEEE 802.11b/g) networks, where we have the convenience of the MAC layer performing a fair division of available channel capacity if multiple nodes are sending packets and can sense each other. In general, this is not true in multihop networks where the hidden node problem is responsible for many collisions. Request/clear to send (RTS/CTS) acknowledgment and handshake packets cannot be used for multicast transmissions; therefore, all nodes should follow the same cooperative protocol to facilitate efficient data dissemination.

The fundamental problem in multihop networks is that some nodes are not directly reachable by the source. Delivering the data stream to all receivers is only possible if some nodes, called relays, propagate the received data to other nodes that are farther away from the source. A relay node helps in the dissemination process by generating and transmitting re-encoded packets. Dynamically selecting these relays in an ad hoc network is not a trivial problem, as was shown in [10]. Currently the application is capable of generating and forwarding recoded packets. A severe limitation is that the physical data rate of broadcast transmissions is fixed to 1 Mb/s on the iPhone platform, so receiving and sending packets with a data rate close to 1 Mb/s (which is typical for a video) is not possible. At the moment, our solution is sufficient for propagating an audio stream (having a bit rate of 128 kb/s) in a linear multihop network, where the nodes are positioned to form a virtual line. But as the network topology becomes more complex or even dynamic, the nodes will broadcast a spate of packets if we continue to use this approach. This phenomenon was called the broadcast storm problem in [11].

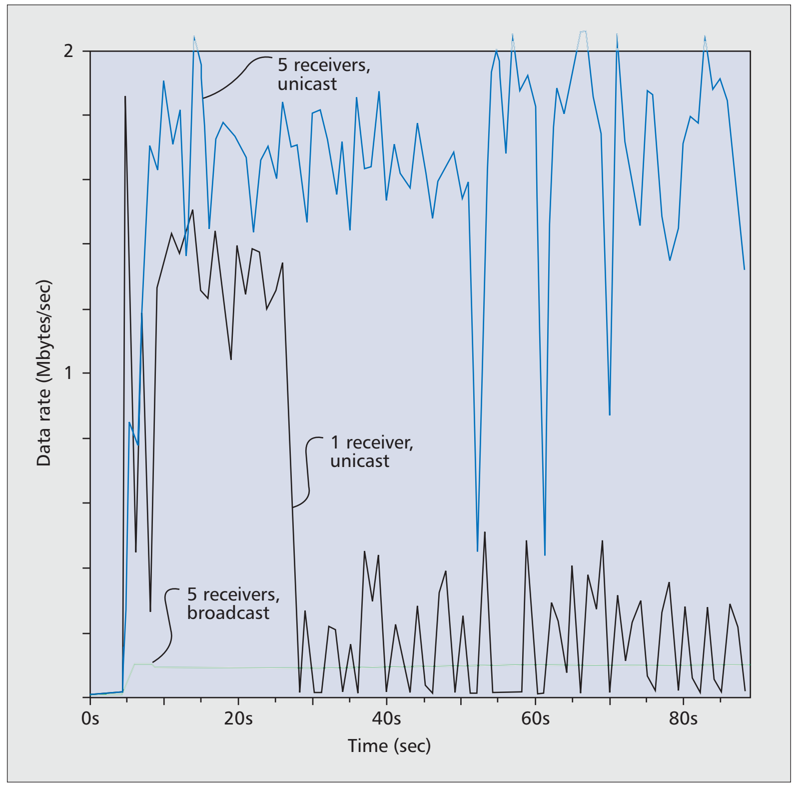


Figure 5. *AirVideo server sending rates with a single receiver (black) and with five receivers (blue) using TCP unicast connections in comparison with the bandwidth usage of our streaming server (green) using UDP multicast transmissions with network coding.*

CONCLUSION

In this article we have introduced a way to disseminate multimedia content in a synchronized manner. We propose a method based on network coding to efficiently deliver data from a single source to many receivers. An application running on Apple iPhone/iPod Touch devices has been presented to show the feasibility of this approach. We observe that the bandwidth usage of this application is remarkably low in comparison with other existing solutions. In [1] the full-blown testbed was shown while playing this video using our application. The first commercial implementation of this technology can be found at [12].

ACKNOWLEDGMENTS

This work was partially financed by the CONE project (Grant No. 09-066549/FTP) of the Danish Ministry of Science, Technology and Innovation as well as by collaboration with Nokia/Renesas, Oulu throughout the ENOC project.

REFERENCES

[1] Mobile Device Group of Aalborg Univ., Network Coding Videos, http://bit.ly/abcOrC, June 2010; collection of YouTube videos.
[2] H. Shojania and B. Li, "Random Network Coding on the iPhone: Fact or Fiction?," *ACM NOSSDAV 2009*, June 2009.
[3] R. Ahlswede *et al.*, "Network Information Flow," *IEEE Trans. Info. Theory*, vol. 46. no. 4, July 2000, pp. 1204–16.

[4] C. Fragouli, J.-Y. Le Boudec, and J. Widmer, "Network Coding: An Instant Primer," *SIGCOMM Comp. Commun. Rev.*, vol. 36, no. 1, 2006, pp. 63–68.
[5] T. Ho and D. Lun, *Network Coding: An Introduction*, Cambridge University Press, 2008.
[6] S. Katti *et al.*, "Xors in the Air: Practical Wireless Network Coding," *SIGCOMM Comp. Commun. Rev.*, vol. 36, no. 4, 2006, pp. 243–54.
[7] J. Heide *et al.*, "Network Coding for Mobile Devices — Systematic Binary Random Rateless Codes," IEEE Wksp. Cooperative Mobile Networks at ICC'09, June 2009.
[8] J. Heide *et al.*, "Know Your Neighbour: Packet Loss Correlation in IEEE 802.11b/g Multicast," *4th Int'l. Mobile Multimedia Commun. Conf.*, Oulu, Finland, July 2008.
[9] F. H. P. Fitzek and M. Katz, Eds., *Cooperation in Wireless Networks: Principles and Applications — Real Egoistic Behavior Is to Cooperate!*, Springer, Apr. 2006.
[10] P. Vingelmann, F. H. P. Fitzek, and D. E. Lucani, "Application-Level Data Dissemination in Multi-Hop Wireless Networks," *IEEE ICC 2010 — CoCoNet Wksp.*, May 2010.
[11] S.-Y. Ni *et al.*, "The Broadcast Storm Problem in A Mobile Ad Hoc Network," *ACM/IEEE MobiCom '99*, New York, NY, 1999, pp. 151–62.
[12] Steinwurf ApS, http://www.steinwurf.com

Biographies

Péter Vingelmann (peter.vingelmann@gmail.com) received his M.Sc. in technical informatics from Budapest University of Technology and Economics in 2009. He is currently pursuing a Ph.D. in wireless communication at the same university. Since January 2009 he has been working with the Mobile Device research group at Aalborg University, Denmark. His main research interests are programming, mobile networks, mobile applications, cooperative communication, network coding, and network performance evaluation.

Frank H. P. Fitzek is a professor in the Department of Electronic Systems, Aalborg University, heading the Mobile Device group. He has visited various research institutes including Massachusetts Institute of Technology (MIT), VTT, and Arizona State University. In 2005 he won the YRP award and received the Young Elite Researcher Award of Denmark. He was selected to receive the NOKIA Champion Award five times in a row from 2007 to 2011. In 2008 he was awarded the Nokia Achievement Award for his work on cooperative networks. In 2011 he received the SAPERE AUDE research grant from the Danish government. His current research interests are in the areas of wireless and mobile communication networks, mobile phone programming, cross-layer as well as energy-efficient protocol design, and cooperative networking.

Morten Videbæk Pedersen received his B.Sc. in electronics engineering in 2007 and M.Sc. in wireless communication in 2009, both from Aalborg University. He is currently pursuing a Ph.D. in wireless communication from Aalborg University. Since January 2006 he has been working in the Mobile Devices research group at Aalborg University, where his primary focus has been on implementation and performance evaluation of network coding algorithms, and cooperative networking protocols and methods. He has co-authored and published several peer-reviewed journal and conference papers, and multiple book chapters. In 2010 he was appointed Forum Nokia Champion. His main research interests are mobile programming, cooperative communication, network coding, and network performance evaluation.

Janus Heide received his M.Sc. in electrical engineering with specialization in wireless communication engineering in 2009 from Aalborg University. He is currently a Ph.D. Fellow within the research section of Antennas, Propagation, and Radio Networking (APNET) in the Department of Electronic Systems at Aalborg University. Since July 2007 he has been working in the Mobile Devices research group at Aalborg University. His main research interests are protocol analysis and design, wireless network systems, characteristics and models, cooperative communication, network coding, and data distribution in ad hoc and meshed networks.

Hassan Charaf came to Hungary in 1987 to study at the Electrical Engineering Faculty of Budapest University of Technology and received his M.Sc. degree in 1992, and later on his Ph.D. degree in 1998. He also earned an M.Sc. degree in economics at Budapest University of Economics. Presently he is an associate professor at the Department of Automation and Applied Informatics of Budapest University of Technology and Economics, where he is head of the IT Group. He holds a number of positions in professional bodies: Microsoft Regional Director in Hungary (1998–2004), IEEE Computer Society, John von Neumann Computer Science Society, Automation and Computer Science Committee of the Hungarian Academy of Sciences, Informatics Committee of the Hungarian Academy of Sciences, and Electrotechnics Committee of the Hungarian Academy of Sciences. He is the author of several books and lecture notes.

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page


Advertise your products
in IEEE Communications Magazine
Reach 55,000
Potential Customers!
Contact us today for
more information or rates
Tel. +1 212 705 8920
Fax. +1 212 705 8999
Email: e.levine@comsoc.org
Website: www.comsoc.org/adv

Real-Time Probing of Available Bandwidth in Home Networks

Archi Delphinanto, Eindhoven University of Technology and TNO
Ton Koonen, Eindhoven University of Technology
Frank den Hartog, TNO

ABSTRACT

Prioritization of flows in a home network based on traffic classification is still no guarantee that enough bandwidth will be available between a content server and a client. Besides, such QoS technologies need to be supported by every device in the end-to-end path to be effective, which is relatively expensive for the owners of home networks. In any small-scale IP network, best effort or QoS-enabled, at home or anywhere else, it is therefore preferable to diagnose the network in real time before admitting a new flow. In this article we analyze existing probing techniques, and demonstrate a new method to probe the available bandwidth between a server and a client in a heterogeneous IP-based home network. The tool works with existing end-user devices, is non-intrusive, has a short measurement time, does not require pre-knowledge of the link layer network topology, and is accurate enough to make decisions about the admission of high-throughput high-quality streams such as for IPTV services.

QOS IN HOME NETWORKS

Home networks are becoming ever more heterogeneous. This means that a single network consists of many different physical- and link-layer technologies and topologies, interconnecting many different devices with each other and the Internet, enabling many different services (e.g., [1, references therein]). The Internet Protocol (IP) suite is the main enabler for the required interoperability. Correspondingly, a growing amount of consumer electronics devices contain an IP stack. Unfortunately, IP has only limited support for quality of service (QoS), necessary to support many different services concurrently in a single shared home network. It mainly concerns reducing or avoiding jitter that occurs when multiple traffic streams contend for bandwidth. Furthermore, noticeable packet loss will happen when User Datagram Protocol (UDP) based services such as high-quality telephony and HDTV need to be supported. This is especially an issue for broadband service providers (e.g., [2, 3]).

Many additional QoS solutions for IP networks are available, but most of them operate on the principle of traffic classification, where each data packet is placed into a limited number of traffic classes, and each router on the network is configured to differentiate the traffic based on its class. These solutions have not gained large popularity in the home networking marketplace, because they need to be supported by every device in the end-to-end (e2e) path to be effective. This makes them relatively expensive for consumers with many non-depreciated devices: to enjoy QoS they have to buy new devices. Besides, current solutions are different for different layer 2 technologies. Intermediate translators would then be needed to guarantee e2e QoS in a heterogeneous path. Though implementations for this exist (e.g., [4]), they are deemed to be too expensive for mass-scale application today. Finally, prioritization of flows is still no guarantee that enough bandwidth (here synonymously used for application-level data throughput) will be available between a server and a client.

Before admitting a new flow to the home network, or rather any small-scale IP network (best effort or QoS-enabled), we therefore propose to diagnose the network in real time. The information can be used for admission control, have the content service pragmatically adapt its properties to the actual condition of the network, or report intermittent issues to the user and/or the service provider's remote service management server. A crucial part of a diagnosing tool is the real-time assessment of e2e available bandwidth between the relevant client and server in the network. Although many e2e speed test applications exist for the Internet [5], none of them fulfills all of the requirements for use in today's home networks. Among these requirements are the following. The tool must:

1 Be easy to implement. It should work with as many as possible existing devices in the home without firmware upgrades. Preferably the tool should only require a simple software module added on the server side of the e2e path of a flow. For most use cases we can therefore assume the diagnosing application to be a service running on the home gateway, serving various clients in the home network, which only need to have a regular IP stack.

2 Be non-intrusive. It should not disrupt other traffic in the home noticeably.
3 Have a short measurement time. It should have a low convergence time from an end-user perspective, and it should be fast enough to react to major changes in the home network traffic pattern. We assume this to be on the order of a few seconds; fluctuations within this time frame may be dealt with by application-layer buffering, for instance.
4 Not require pre-knowledge of the link-layer network topology. Home networks may contain link-layer technologies that are not standardized or widely known.
5 Be accurate enough to make informed decisions about the admission of delay- and jitter-critical applications. In the case of IPTV and IP telephony, that means an accuracy of ~1 Mb/s and ~50 kb/s, respectively.

In this article we propose a new tool, the Available Bandwidth Estimator (Allbest), which fulfills all the above requirements to the extent that it is a software upgrade of only the probe server, it injects less than ~ 1 Mb/s of probe traffic, produces results with an accuracy better than ~ 1 Mb/s within less than 10 s, and can be applied on any Ethernet-WiFi combined topology. This is the first diagnosing application that successfully applies probe round-trip time (RTT) measurement to wireless LANs and does not assume any home network topology a priori. The tool is based on a new probing method we developed, which is described in the following section. In the second half of the article we detail our testbed, followed by our test results. We finish with a conclusion and a few words on our current and future experiments.

Probing Home Networks

Bottleneck Capacity and Available Bandwidth

We follow [6] for defining capacity of a hop as the bit rate, measured at the IP layer, at which the hop can transfer maximum transmission unit (MTU)-sized IP packets. Therefore, the capacity of an e2e path is the maximum IP-layer rate the path can transfer from source to sink. In our work, we assume the MTU size to be 1500 bytes of Ethernet v2 (RFC 1191). As a path may consist of several links, the minimum link capacity in the path determines the path capacity. This link is called the narrow link. In contrast, the tight link is the link in the path with the maximum capacity utilization. This is the link with the least available bandwidth due to crossing traffic (i.e., other traffic in the path considered for admission of a new stream). In many cases, the tight link is in the narrow link, and the link is then referred to as the bottleneck.

For measuring bottleneck bandwidths in Internet paths, two types of tools can be distinguished: packet-pair dispersion tools (also called probe gap model, or PGM) and self-loading techniques. The latter probe the network with trains of packets [7] at an increasing rate, and thus rely on flooding the network. They therefore do not fulfill requirement 2. PGM techniques were first explored in [8], and send only a few packets at the rate C of the bottleneck capacity or somewhat slower. This allows crossing traffic to get in between the probe packets and disperse them (i.e., increase the difference in arrival time).

An issue with PGM techniques is that C needs to be known a priori. From requirement 4 follows that the tool must be able to estimate the available bandwidth A without such pre-knowledge of the path. This means that C needs to be determined first, and the estimation of A becomes a two-step process. In [9] we proposed and validated a new method for determining C in heterogeneous home networks based on RTT measurements of probe packets, and fulfilling all requirements listed in the previous section. However, we also learned that probing a wireless LAN with rate C may yield the correct average dispersion rate at the receiver when measuring in one direction, but will not if it needs to be derived from RTTs. This is easiest understood by looking at the details of our capacity estimation method first.

Capacity Estimation

Our capacity estimation method is based on the packet-pair dispersion technique, which is usually implemented by sending two packets back-to-back on the network, thus minimizing the chance that crossing traffic will disperse the packets. It is then the bottleneck that will delay the second packet with respect to the first. C can subsequently be calculated simply from the minimum dispersion D and the packet size L as $C = L/D$. One then should minimize the chance that crossing traffic increases or decreases the bottleneck dispersion of the packets further down the path. To do so, we perform a series of n packet-pair probes, assume the crossing traffic stochastic, and then calculate D from the minimum RTT of the first packet (RTT_1) of a probe pair and the minimum RTT of the second packet (RTT_2) of a probe pair. C is then given by

$$C = L / \left(\min_{i=1\ldots n} [RTT_2(i)] - \min_{i=1\ldots n} [RTT_1(i)] \right). \qquad (1)$$

RTTs can be measured without adaptation of the client side by using MTU-sized Internet Control Message Protocol (ICMP) Ping probe packets, or by sending MTU-sized UDP packets to a non-activated port. The client then automatically generates reply packets: ICMP Echo packets or ICMP Error packets (i.e., code 3 or "Destination port unreachable"), respectively. ICMP Error packets are much smaller than ICMP Echo packets and therefore experience hardly any delay on the way back to the probing sender/receiver, assuming that the return one-way capacity between the client and the server, $C_{reverse}$, is not much smaller than the sought-after one-way capacity $C_{forward}$ between server and client. The final result is then a good measure for $C_{forward}$. For symmetric media we may also use ICMP Ping probing packets, and assume that the delay and dispersion is the same for both directions of travel. Equation 1 then yields $C/2$ rather than C.

Equation 1 allows us to avoid unwanted contention of probe packets in the wireless medium. Existing packet-pair dispersion techniques will not

Our capacity estimation method is based on the packet-pair dispersion technique, which is usually implemented by sending two packets back-to-back on the network, thus minimizing the chance that crossing traffic will disperse the packets. It is then the bottleneck that will delay the second packet with respect to the first.

We assume that, for UDP probing, most of the random delay is experienced in the forward direction. This is justified by the fact that the reply packet is very small, and we assume that the queueing mechanism of the system is fair. The reply packet is therefore hindered relatively little by the crossing traffic.

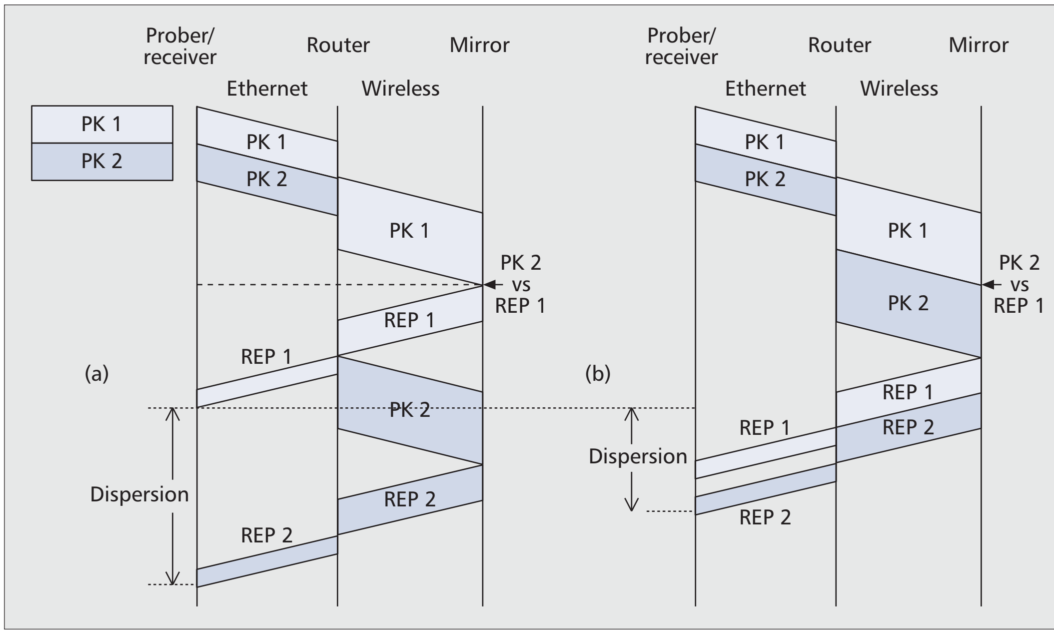


Figure 1. *Conventional back-to-back packet-pair probing in heterogeneous home networks. PK1 and PK2 are the probe packets sent back-to-back on the path. REP1 and REP2 are the respective reply packets. a) PK2 has to wait until REP1 is off the wireless medium; b) REP2 has to wait until REP1 is off the wireless medium.*

work in wireless media round-trip, because the reply packet of the first probe packet contends with the second probe packet on the air interface (Fig. 1). Irrespective of which packet wins, the reply packet of the second probe packet will eventually arrive at the probing sender/receiver too late. PK2 is acting as crossing traffic to the reply packets; the method is basically self-disturbing. As a result, C will be underestimated.

To avoid this contention, we need to prevent the first reply packet from being put on the network. We achieved this (Fig. 2) by sending a single packet with size 2 × MTU instead of two packets back-to-back. On the network, this packet will automatically be fragmented (and behave like two individual packets back-to-back), and only after defragmentation will a single reply be sent back by the client. This will provide us the correct RTT_2, that is, the RTT_2 only delayed by bottleneck dispersion, not by additional contention. Because we are not directly measuring D, but separate RTTs (Eq. 1), we can find the correct RTT_1 by sending a different series of single probe packets, well separated from each other and with size MTU.

Contention of probe and reply packets is the main reason why conventional packet-pair probing is not suitable for determining A. We can neither probe back-to-back nor with rate C without creating extra delay caused by contention. Unfortunately, we cannot solve the latter by using fragmentation, because that can only mimic back-to-back probing.

Available Bandwidth Estimation

A deep analysis of the various delays that constitute the RTTs observed during capacity estimation allowed us to make a good estimation of A also. In Fig. 3 a typical histogram is shown of RTT_1 that we measured in an IEEE 802.11b network with 1.5 Mb/s crossing traffic. Besides a clear minimum value, the RTT undergoes two random effects: the random back-off mechanism of IEEE 802.11 (mostly at short additional delays) and the delay caused by queuing due to crossing traffic.

We assume that, for UDP probing, most of the random delay is experienced in the forward direction. This is justified by the fact that the reply packet is very small, and we assume that the queuing mechanism of the system is fair. The reply packet is therefore hindered relatively little by the crossing traffic. We further assume that any systemic delay in the network (for instance processing delay) is either negligible or canceled when subtracting RTT_1 from RTT_2 [10], and that the delay caused by random effects is mainly happening in the bottleneck. A is then given by

$$A = L / \left(\frac{L}{C} + \bar{d}_r \right) \qquad (2)$$

with L/C the delay in the bottleneck without crossing traffic following from Eq. 1, and the average delay caused by random effects in the bottleneck. The latter can be derived from RTT_1 as

$$\bar{d}_r = \underset{i=1\ldots n}{avg}\,[RTT_1(i)] - \min_{i=1\ldots n}\,[RTT_1(i)]. \qquad (3)$$

Allbest Testbed

The setup of our testbed is schematically drawn in Fig. 4. The Allbest server runs on the "prober/receiver" computer, and probes the "mirror" via any heterogeneous topology of interest. The results presented in this article were obtained by configuring a WLAN IEEE 802.11b or 802.11g with a Linksys WRT54GL v.

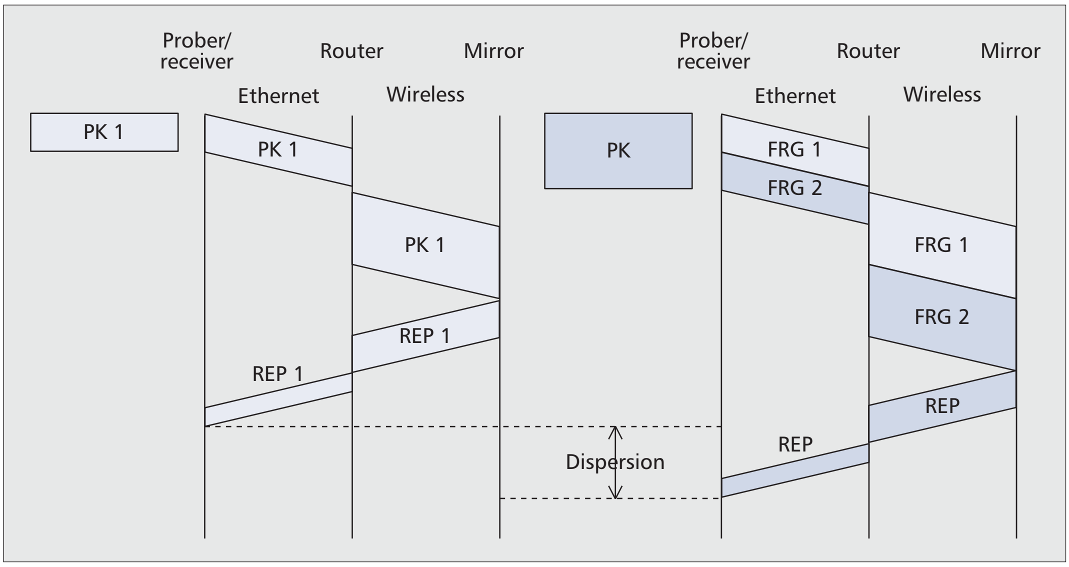


Figure 2. *Allbest's method of active probing in heterogeneous home networks. Probe packets PK1 and PK are sent far apart from each other. PK has size 2 × MTU and is automatically fragmented on the network. The fragments FRG are dispersed. A reply REP is sent only after defragmentation; thus, no contention has occurred.*

1.1 access point as a bottleneck link. We switched off the automatic rate adaptation and clear to send (CTS) protection mode, and run both networks on their maximum physical rates of 11 Mb/s and 54 Mb/s, respectively. The measurements were carried out in a Faraday cage to avoid uncontrolled interference.

We benchmarked Allbest against the well-known testing tool Iperf and Wbest [11]. Wbest is the only other real-time probing tool we know that is applicable to wireless networks. It requires the wireless hop to be in the last link, because it needs to be sure that the probing packets arrive at the bottleneck with rate *C*. For the estimation of *C* it uses standard PGM and packet-pair dispersion. Both Wbest and Iperf need to be installed on both the prober/receiver (which for Wbest and Iperf just acts as a prober) and the mirror (which for Wbest and Iperf acts as a receiver). The prober/receiver and mirror are laptop computers with a 2.0 GHz processor. Allbest and Iperf run on Windows XP Service Pack 3, and Wbest runs on Linux Ubuntu 10.04. To maximize the performance of the software, other processes running in the computers' background memory were switched off whenever possible.

Allbest basically consists of a home-built configurable UDP packet generator and a home-built configurable ICMP Ping packet generator, combined with Wireshark to measure high-precision RTTs. Any RTT > 2 × min[RTT(i)] is discarded, and we have verified that most of those long RTTs are caused by uncontrollable processing delay in the laptops due to other tasks of the operating system. A measurement takes 90 probe pairs and is repeated 6 times.

With Iperf we measured at which UDP injection rate which packet loss occurs with 1472-byte payload per packet. The result is fitted linearly, and the point where the fitted line crosses the transmission rate axes is interpreted as being the available bandwidth. Each Iperf measurement is set for 10 s with 1 s interval. The UDP packet loss is averaged over 8–10 similar measurements, leading to 1 percent standard deviation. Also, Wbest was configured to use 1472-byte UDP payload. Each measurement of 90 packet pairs was repeated 30 times. Like Allbest, Wbest filters and discards unreliable results.

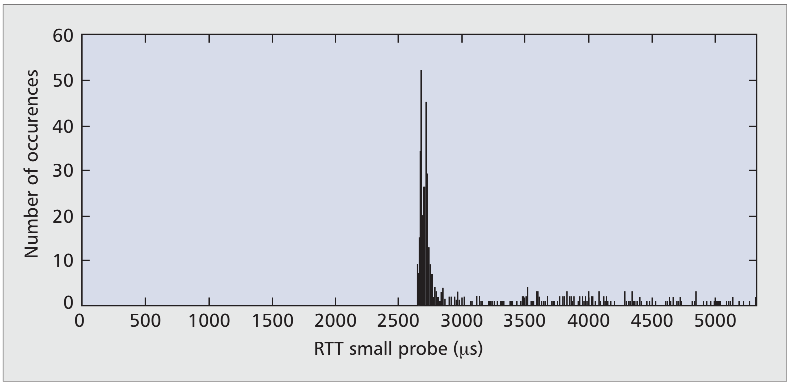


Figure 3. *Histogram of 1000 RTTs of MTU-sized probe packets in an IEEE 802.11b network with 1.5 Mb/s crossing traffic. Bin size = 20 μs.*

Random UDP crossing traffic is generated with the Distributed Internet Traffic Generator (D-ITG), available from the Universita degli Studi di Napoli "Federico II." The UDP packets have uniformly distributed packet sizes (40–1472 bytes) and are sent at Poisson-distributed exponential time intervals. We distinguished crossing traffic and contending traffic, and follow [11] for their definitions.

Available Bandwidth in Ethernet-WiFi Networks

We have obtained the available bandwidth for three different topologies with Iperf, Wbest as well as Allbest (in its UDP probing variety):

1 Prober/receiver→100BASE-TX→IEEE 802.11b→mirror

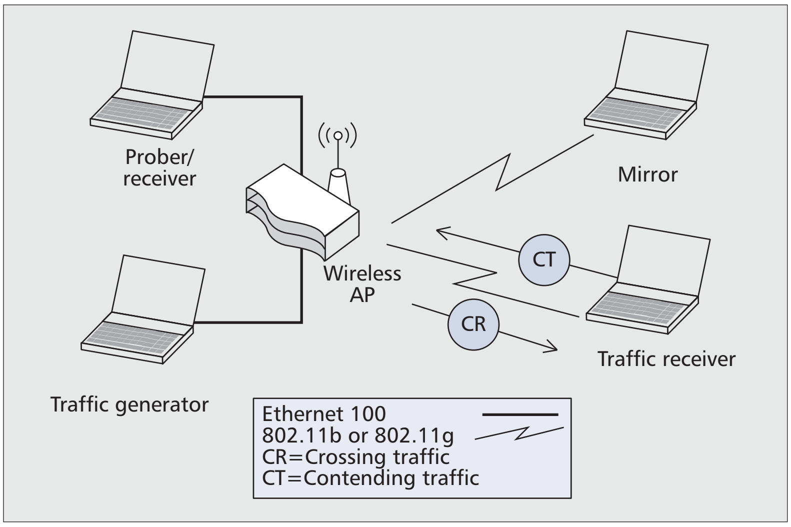


Figure 4. *Schematic view of our heterogeneous wired/wireless LAN probing testbed. Allbest runs on the prober/receiver. The traffic generator generates crossing or contending traffic.*

2 Prober/receiver→100BASE-TX→IEEE 802.11g→mirror
3 Prober/receiver→IEEE 802.11g→100BASE-TX→mirror

For every topology we generated three different amounts of crossing traffic (X), at about 0, 25, and 50 percent of the capacity, and 25 percent of contending traffic.

For topology 1, the results are summarized in Fig. 5a. For this topology, with IEEE 802.11b, all tools yield similar results at first sight and within the error margins. Allbest estimates the capacity C of 802.11b on 7.6 ± 0.2 Mb/s, which is equal to the theoretical value [9]. For all tools, the available bandwidth A is lower than C for $X = 0$. This is caused by the random backoff mechanism of WLAN. If the random backoff algorithm were to be active for all packets, we would expect $A = 6.4$ Mb/s [9]. Since all tools estimate available bandwidths somewhat higher than that, we suspect there are still many packets that do not undergo random backoff.

On the whole, Allbest seems to find larger available bandwidths than Iperf, and Wbest finds even larger ones. Even though Iperf is a well-known benchmarking tool, it is probably underestimating A in our experiments. Because the crossing traffic is stochastic, some packet loss will already be recorded at relatively low Iperf injection rates. For both Allbest and Wbest, the values for A at $X > 0$ are also closer than Iperf to the expected value of A (obtained by simply subtracting the crossing traffic rate X from the available bandwidths A at $X = 0$ Mb/s). The error margins of the Allbest results are remarkably lower than the ones for Wbest, although we tried to have the results based on the same number of probes. We do not have an explanation for this yet.

For Allbest, the value for A at $X = 3$ Mb/s was calculated by discarding any RTT > 3 × min[RTT(i)], rather than using the default cutoff time of 2 × min[RTT(i)], as stated in the previous section. We found that with the latter, too many packets had been discarded that were clearly delayed by crossing traffic, and A was grossly overestimated (5.1 ± 0.2 Mb/s). Many PGM techniques use a default cutoff time of 2 × min [RTT(i)]. This follows from their assumption of fair queueing congestion management in the router. This means that bottleneck delays can never be larger than $2L/C$, even if the utilization by crossing traffic is larger than 50 percent (which then just results in larger packet loss). Crossing traffic of 3 Mb/s with randomly distributed time intervals will utilize the bottleneck more than 50 percent for at least part of the time. Fortunately, the actual congestion management mechanism of the router (most probably store and forward) still allowed us to capture relevant packets with larger RTTs and compute a realistic value for A.

For topology 2, the results are summarized in Fig. 5b. For this topology, with IEEE 802.11g, Allbest shows clear supremacy. Allbest estimates the capacity C of 802.11g on 38 ± 2 Mb/s, which is equal to the theoretical value [9]. The value of A at $X = 0$ Mb/s is then expected to be 26 Mb/s if the random backoff algorithm is active for all packets [9]. Iperf and Allbest estimate somewhat higher again, but Wbest significantly underestimates A, also for larger X. The inventors of Wbest warn of underestimation when the probe packets arrive at the bottleneck at a rate larger than C [11]. Surprisingly, Wbest's capacity estimation for topology 2 is quite good, 38 Mb/s. The fact that Wbest arrived at plausible answers for A with topology 1 can be explained by it grossly overestimating the C of topology 1 (8.8 Mb/s). The results for Allbest are very close to the ones for Iperf, and have the lowest error margins of all. But as in Fig. 5a, it is not clear whether Iperf yields the correct values. More than Iperf, Allbest yields values for A at $X > 0$ close to what one obtains by subtracting X from $A(X = 0)$.

For topology 3, the results are much the same as for topology 2 (Fig. 5c). This shows that Allbest can function in either order of physical- and link-layer technologies. Unfortunately, we could not get any UDP results from Iperf, because it cannot inject faster than about 10 Mb/s when directly connected to an 802.11g network.

Conclusions and Future Work

We achieved a breakthrough in available bandwidth probing of heterogeneous home networks by understanding and then solving the contention issues that PGMs traditionally had with wireless links in the e2e path. This allowed us to design a new probing method based on round-trip time measurements, with low intrusiveness and short convergence time, and without the need to know the home network topology a priori. Our tool, Allbest, is accurate enough to make informed decisions about the admission of IPTV streams and the like, and gives the service provider no more information than strictly needed. We have built a prototype and a testbed, and our performance measurements indicate that Allbest works well and outperforms Iperf and Wbest for various topologies based on 100BASE-TX, IEEE 802.11b, and 802.11g, for up to 50 percent crossing traffic.

Our work is opening up a whole field of research related to diagnostics of heterogeneous

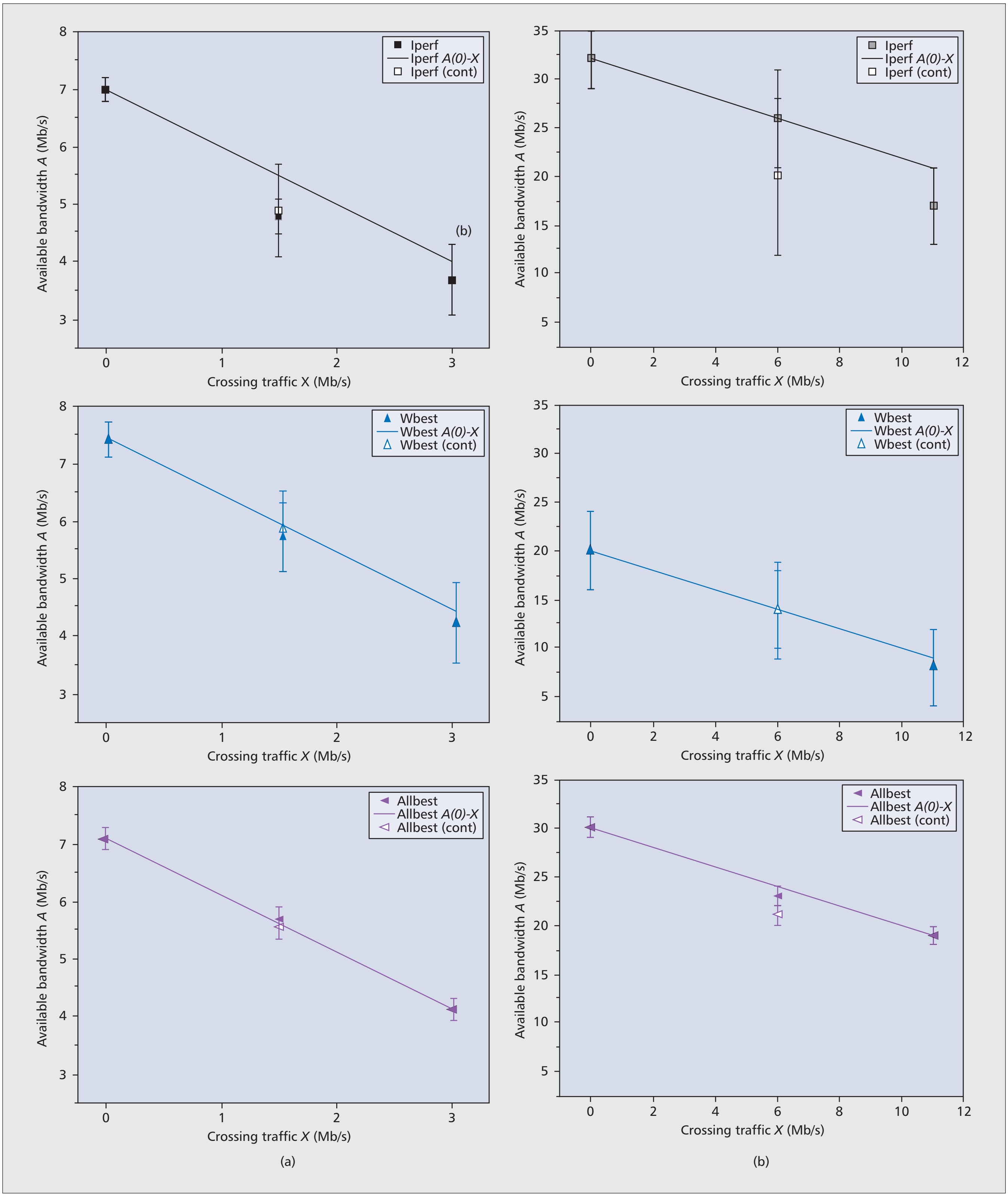


Figure 5. *Part 1. Available bandwidth measured with different tools and different amounts of crossing and contending (cont) traffic. a) topology 1; b) topology 2.*

home networks. Many different configurations and network parameters need to be investigated. New network technologies supporting IP are currently entering the home, such as HomePlug, MoCa, IEEE 1901, IEEE 802.11n, and G.hn. Some of them (e.g., HomePlug) are exhibiting very different physical- and link-layer properties (e.g., fast rate adaptation) from the networks studied in this article. Our method needs to be improved to include these novel techniques. Also different queuing mechanisms than fair queuing should be studied. To increase the accuracy of our method up to a level at which it can be used for voice-over-IP services, network tomography techniques may be applied.

Another matter is how the obtained values should lead to intelligent decisions on the service level, which probably needs some form of cross-layer optimization. One of the questions following from this is how frequent a measurement should be repeated. It will certainly depend on the dynamics of the traffic in the home network. It can be safely assumed that crossing traffic in homes cannot be modeled with the stochastic properties of Internet traffic. We are currently performing a series of experiments in Dutch homes in order to understand the in-home traffic dynamics.

Finally, the applicability of our method to other consumer networks should be studied. In-car networks, personal area networks, hotel networks, and others exhibit similar properties and management issues as home networks. At the IEEE CCNC 2011 conference, operators also showed interest in using Allbest for mobile access networks.

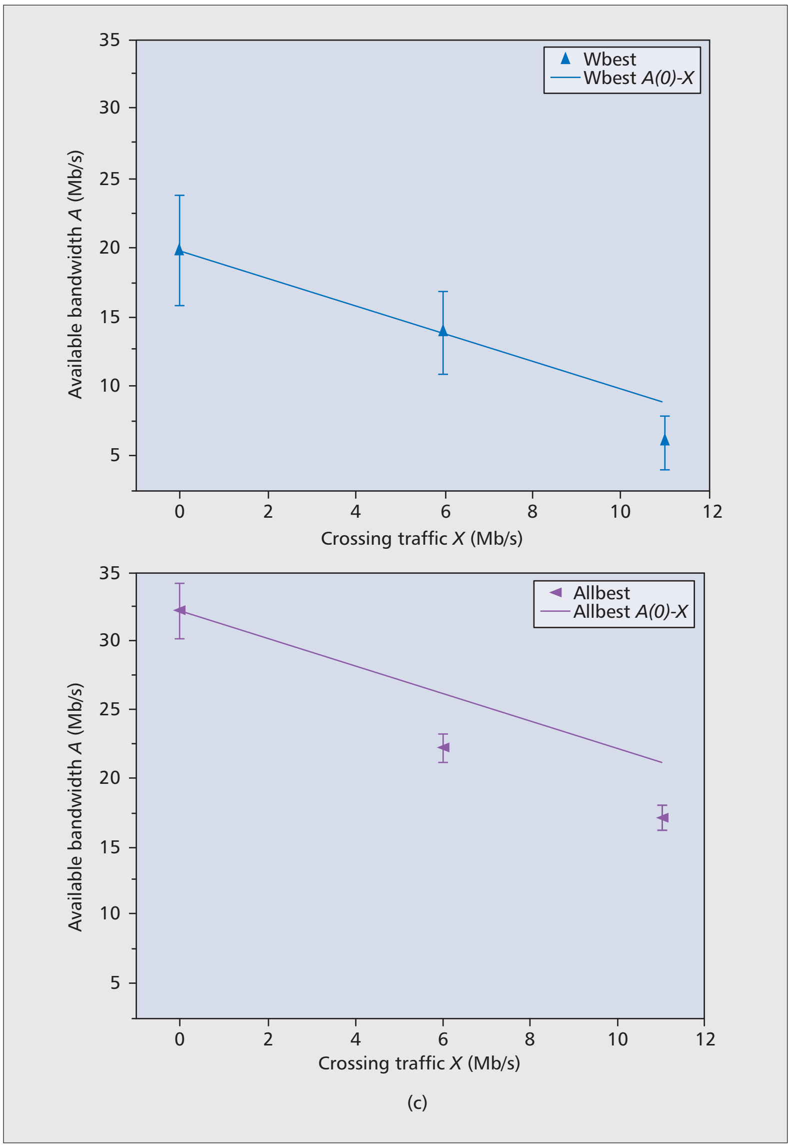


Figure 5. *Part 2. c) Topology 3.*

Acknowledgment

This work was partly done in the collaborative Integrated Project FIGARO, which is supported by the European Commission under the 7th Framework Programme, Grant Agreement Number 258378.

References

[1] F. T. H. den Hartog *et al.*, "First Experiences with Personal Networks as An Enabling Platform for Service Providers," *Proc. 4th Annual Int'l. Conf. Mobile and Ubiquitous Systems: Networking & Services*, Philadelphia, PA, Aug. 2007, pp. 1–8.

[2] Home Gateway Initiative, "HGI Guideline Document: QoS white paper," HGI-GD013-R2, 2009, http://www.homegateway.org.

[3] C. Develder *et al.*, "Delivering Scalable Video with QoS to the Home," *Telecommunication Systems*, Springer, Online FirstTM, 9 June 2010.

[4] D. Jiang and M. Li, "Quality of Service in the Home Network," *Proc. 2nd Int'l. Conf. Future Generation Commun. and Networking*, Hainan Island, PR China, Dec. 2008, pp. 473–76.

[5] E. Goldoni and M. Schivi, "End-to-End Available Bandwidth Estimation Tools, An Experimental Comparison," *Traffic Monitoring and Analysis*, *LNCS*, Springer, vol. 6003, 2010, pp. 171–82.

[6] R. Prasad *et al.*, "Bandwidth Estimation: Metrics, Measurement Techniques and Tools," *IEEE Network*, vol. 17, no. 6, Nov./Dec. 2003, pp. 27–35.

[7] B. Melander, M. Bjorkman, and P. Gunninberg, "A New End-to-End Probing and Analysis Method for Estimating Bandwidth Bottlenecks," *Proc. IEEE GLOBECOM 2000*, San Francisco, CA, Nov. 2000, vol. 1, pp. 415–20.

[8] V. Jacobson, "Congestion Avoidance and Control," *Proc. ACM Symp. Commun. Architectures and Protocols (SIGCOMM'88)*, Stanford CA, Aug. 1988, pp. 314–29.

[9] A. Delphinanto *et al.*, "Path Capacity Estimation in Heterogeneous, Best-effort, Small-scale IP Networks," *Proc. 35th IEEE Conf. Local Computer Networks*, Denver, CO, Oct. 2010.

[10] W. Kampichler and K. M. Goeschka, "On measuring Quality of Service Limitations in Local Area Networks," *Proc. IEEE ICC '03*, Anchorage, AK, May 2003, vol. 1, pp. 291–95.

[11] M. Li, M. Claypool, and R. Kinicki, "WBest: A Bandwidth Estimation Tool for IEEE 802.11 Wireless Networks," *Proc. 33rd IEEE Conf. Local Computer Networks*, Montreal, Canada, Oct. 2008, pp. 374–81.

Biographies

Archi Delphinanto [M] received his B.Sc. in engineering physics from Bandung Institute of Technology, Indonesia, in 1998, and his PD.Eng. in ICT from Eindhoven University of Technology, the Netherlands, in 2003. He currently works as a scientist at TNO, the Netherlands, and is also pursuing his Ph.D. degree in the Electrical Engineering Department at Eindhoven University of Technology. His research interests focus on service discovery and access in heterogeneous networks, including interoperability, quality of service, and service monitoring.

Ton Koonen [F] is a full professor at Eindhoven University of Technology (TU/e) since 2001, and chairman of the Electro-Optical Communication Systems group. Before joining TU/e, he worked for more than 20 years in applied research in industry at Bell Laboratories in Lucent Technologies, and as a part-time professor at Twente University. His current research interests are in optical fiber access and in-building networks, including radio-over-fiber techniques. He is a Bell Labs Fellow.

Frank den Hartog [M] (Frank.den.Hartog@ieee.org) is a senior scientist, Future Internet at TNO, and deputy chair of the Technical Working Group of the worldwide Home Gateway Initiative (HGI). Between 1998 and 2003 he was an applied scientist at KPN Research, where he pioneered the home networking research area. He now acquires and manages large collaborative research projects on every aspect of heterogeneous consumer networking. He has authored over 100 conference articles, journal papers, patents,ß and contributions to standardization.

Using Home Routers as Proxies for Energy-Efficient BitTorrent Downloads to Mobile Phones

Imre Kelényi and Ákos Ludányi, Budapest University of Technology and Economics
Jukka K. Nurminen, Aalto University

This work is connected to the scientific program of the "The Development of Quality-Oriented and Cooperative R+D+I Strategy and Functional Model at BUTE" project. This project is supported by Nokia Research Center and the New Hungary Development Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002). Most of this research was done when Jukka K. Nurminen was working for Nokia Research Center.

ABSTRACT

Using proxy servers to cache and shape network traffic can significantly improve the energy efficiency of the participating mobile clients. In this article, we analyze the implications of hosting a BitTorrent proxy on a broadband router, which pushes the content to a mobile phone over wireless radio (WLAN or 3G). The amount of memory in a router is limited; therefore, our interest is in how to efficiently use memory to download BitTorrent content as fast as possible and at the same time transfer it to the mobile phone in an energy-efficient way. We investigate these aspects via a series of measurements. The results show that the proxy-based solution outperforms the torrent client running on the phone in terms of energy and download time. We also simulate the BitTorrent operation to understand how these memory-limited devices would influence the operation of the whole community.

INTRODUCTION

Most peer-to-peer applications, including BitTorrent, were originally designed for PC devices. When they are used on handheld devices, such as mobile phones, new problems arise because of limited resources and more constrained communication. A particular worry, which is also the target of this article, is how the applications influence the energy consumption of the device. Frequent need to recharge is inconvenient for the user, and at times the electricity grid can be inaccessible.

BitTorrent is currently the most popular peer-to-peer content sharing solution, used all around the Internet. BitTorrent's key idea is that clients downloading content, which is often referred to as *torrents*, also become uploaders, and share the downloaded pieces with other peers. This enables scalability and robustness, making BitTorrent a serious alternative to centralized content sharing solutions such as FTP or HTTP.

Running BitTorrent on mobile phones is clearly interesting for users as evidenced by close to 100,000 downloads of the native BitTorrent client for Symbian-based mobile phones in 2009–2010 [1]. In our earlier work to improve the energy efficiency of content downloading [2] we investigated ways of using an intermediate server hosted in the Internet to download BitTorrent content. In that case, the transfer was split into two parts. Regular BitTorrent was used to download the content to the server, which was hosted on the Amazon EC2 cloud computing platform, and normal HTTP was used to transfer the complete file from the server to the phone. When receiving content from the proxy, the mobile phone experienced a higher bit rate, which, as earlier studies show [3], improves the energy efficiency of file transfer. The speed of BitTorrent download can vary considerably and is often well below the capacity of the wireless interface. In contrast, a proxy equipped with a high bandwidth upload link can send the data to the mobile phone in a fast energy-efficient burst.

A practical problem of the proxy solution is where to host the proxies. One possibility is that the proxy is hosted on the Internet, but it is unclear how to cover the hosting costs of such a service. Another alternative is that users run the proxies on their PCs. This requires that PCs be accessible from mobile devices and powered on all the time, requiring careful security and energy consumption considerations. The third option, and the focus of this article, is that proxies are run on broadband routers in homes.

The router platform is attractive for a number of reasons. First, since most homes are equipped with them, the installed base of routers is large. Second, many of the router platforms allow modifying the operating system. Third, routers are typically powered up all the time, and the energy consumption of the router is almost constant no matter how actively it is used. Running proxies on broadband routers would thus introduce no new costs.

A number of new problems, however, arise because a broadband router has more limited resources than an average PC computer. Although some models allow hardware extensions with USB devices, and some high-end router models even have built-in support for tor-

 0163-6804/11/$25.00

rent download, typically, the memory size is limited and no mass memory is available. Future routers are likely to be more capable but we think that it is important to be able to work on minimalistic hardware, especially if we consider the generality of the idea, other extension cases in the future, and utilization of today's home routers.

Prior research has investigated the use of helper nodes, peers that assist in content downloading without their own interest in the content [4, 5]. These ideas are related to our research but, in contrast to our work, they assume that the helper nodes are regular PCs with plentiful resources.

In this research, we extend the earlier work in two directions. First, our explicit interest is in how the system can use a proxy peer to energy-efficiently deliver a torrent to a battery powered, wirelessly connected mobile device. Second, in our solution we use regular router models, which do not have space to store the whole torrent. The ability to store and share pieces that a peer has already downloaded is one of the key concepts of BitTorrent. If the storage space is limited, downloaded pieces have to be discarded when the memory fills up, and thus cannot be shared any longer. Consequently, we investigate how this influences BitTorrent's behavior and what policies and mechanisms are needed to manage the memory. Our target is that these solutions are compatible with the existing BitTorrent clients and that they do not harm the download performance of regular peers.

We explore these mechanisms in detail. We describe our prototype implementation and the measurement results of the performance of the system. We discuss our simulations for analyzing the effect of our key mechanisms to the performance of both proxy and regular peers on a large scale. We briefly review related studies, and then conclude the article and discuss some ideas for further research.

The Router Proxy Solution and Its Difficulties

With the proxy solution, the torrent download is divided into two overlapping activities. To the *swarm* (the set of peers downloading and sharing a particular torrent) the proxy looks like a regular BitTorrent peer; it downloads and uploads content with the normal mechanisms. At the same time, it forwards downloaded content to the mobile phone. Because these two activities have to be done using the limited memory of the router, two interesting questions arise: how to serve other peers when the whole torrent does not fit into the memory and when to communicate with the mobile device.

A key assumption of BitTorrent operation is that when a peer has completely downloaded a piece of the torrent, it announces the availability of the piece to its peers. The peers can then assume that the announced piece is available for download. This is, however, not the case if only part of the content fits in the router memory. To be able to download the whole torrent, the router has to delete some pieces after they have been sent to the mobile device, and then reuse the memory to download additional pieces. The assumption that a piece a peer has downloaded is available for others is thus no longer valid.

The trivial solution to this problem is to ignore it. The BitTorrent client in the router would announce the pieces normally. Then it would later receive requests for pieces that have already been deleted from the memory and cannot be uploaded to other peers anymore. This would have a negative effect on the download speed because most clients stop serving or even disconnect peers that cannot serve a requested piece.

Another alternative is to announce no pieces at all and miss the opportunity to serve other peers. However, this free-rider behavior is bad for the proxy itself because BitTorrent's tit-for-tat algorithm ranks it low, resulting in slow download speed. Furthermore, if the number of such proxies in the BitTorrent swarm is high, it can have a negative effect on the download speed of other peers.

In order to both serve other peers and ensure that all of the piece announcements the proxy peer makes are valid, our solution divides the available memory into two buffers. The *download buffer* holds transient data on the way to the mobile; the pieces are downloaded from peers, sent to the mobile device, and discarded. Then the same memory space is reused to download other pieces. The content of the download buffer is thus constantly changing as the download of the torrent progresses. The *upload buffer*, on the other hand, stores pieces that are served to the swarm. After a piece in the upload buffer has been downloaded and sent to the mobile device, it remains in the memory and is made available for other peers with the normal BitTorrent piece transfer mechanisms. The concept is illustrated in Fig. 1.

The system also has to decide when to communicate with the mobile device. We use a parameter, *chunk size*, to indicate how many pieces should be completely available before they are sent to the mobile. Intuitively, a bigger chunk size would reduce download speed and waste memory because a number of completely downloaded pieces are waiting for other pieces to complete. On the other hand, sending a bigger set of data to the mobile phone in one pass would improve energy-efficiency.

Upload/Download Buffer Allocation

Determining how to allocate the router's memory to upload and the download buffers is not trivial. A large download buffer would allow downloading more pieces in parallel. However, if the peer cannot serve others, BitTorrent's tit-for-tat mechanism may prevent it from utilizing its full download capacity.

As a rule of thumb, the peer should allocate as much of its storage capacity for downloading as needed to maintain the maximum number of parallel downloads, while using the remaining capacity for uploading to maintain a good tit-for-tat ranking. Because the number of parallel downloads depends on the dynamic characteristics of the swarm, a static upload/download buffer ratio does not work in all cases. There-

> *Future routers are likely to be more capable but we think that it is important to be able to work on minimalistic hardware, especially if we consider the generality of the idea, other extension cases in the future, and utilization of today's home routers.*

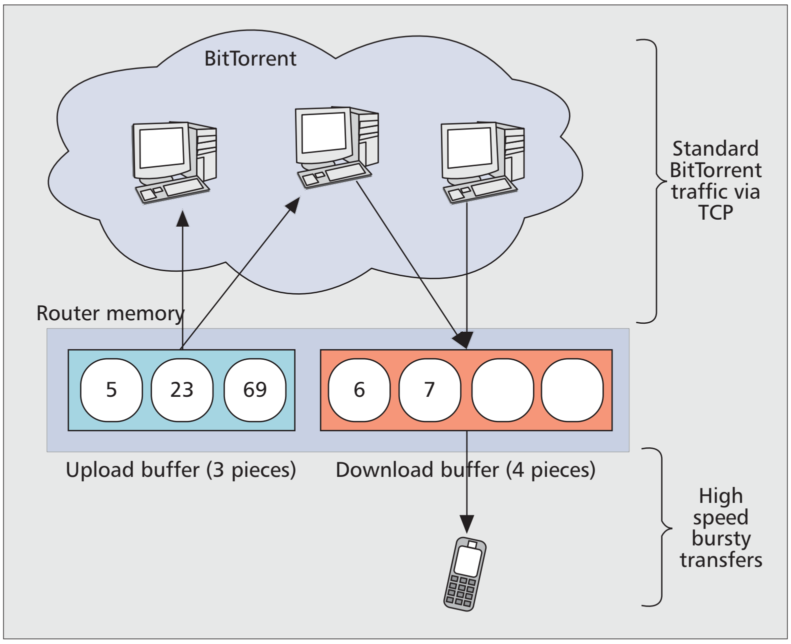


Figure 1. *ProxyTorrent architecture.*

fore, we developed an adaptive buffer allocation algorithm to optimize the download speed of torrents with different swarm characteristics.

The main idea of the algorithm is to try to start downloading from as many peers as possible, and then dynamically reallocate the unused part of the download buffer to the upload buffer. Thus, after joining the swarm and establishing the connections, we set the number of download pieces to be equal to the number of accessible peers with pieces we like to download. Then we periodically check the utilization of the download buffer. If it is underutilized, we reallocate some space from the download buffer to the upload buffer. In this way, the download buffer size can only shrink during the download process. This is in line with the fact that the number of peers we are interested in is also likely to decrease as the download progresses because when the peer is getting more pieces, the number of peers in possession of still missing pieces gets smaller.

Energy Measurements and Results

The Router-Based BitTorrent Proxy Prototype

In our prototype system we used a Linux-based router to host the proxy. The Linux distribution we installed on the routers is DD-WRT [6], which is an open source embedded operating system especially tailored for routers. In addition to giving us the ability to fully control the router's network configuration setup, DD-WRT allows us to build and execute custom applications on a wide range of commercially available routers. We used Asus WL-500gP, which has a 266 MHz CPU, 8 Mbytes flash storage, and 32 Mbytes RAM.

For the BitTorrent downloads we used the Enhanced CTorrent BitTorrent client [7], which we extended with the proxy functionality. CTorrent is a standard BitTorrent client designed to be quick and lightweight, which are the key requirements for a client running on a router with limited resources. The modified version allows us to specify the size of the upload and download buffers plus the chunk size.

We used Nokia N82 phones to perform the measurements. The energy consumption was measured using Nokia Energy Profiler [8]. We developed a mobile client with Java ME, which can start torrent downloads and receive the downloaded pieces from the proxy.

The mobile client establishes multiple TCP connections with the proxy and uses them in parallel when downloading new pieces from the proxy. We varied the number of parallel TCP connections and found that more connections resulted in higher speeds. However, increasing the number of connections beyond five did not improve the speed anymore. Therefore, we used five sessions for the measurements.

We allocated roughly 13 Mbytes of the router's RAM memory for ProxyTorrent to have capacity for 100 BitTorrent pieces (of 128 kbyte in size); the remaining memory was used by the operating system and services running on the router. We performed a set of tests with different memory buffer allocations, which showed that for good performance the majority of the memory should be allocated for the upload buffer. This was also verified by the simulations. Consequently, we used a fixed allocation of 75 percent upload buffer and 25 percent download buffer.

We performed all of the measurements in a one-week period. We repeated each measurement at least three times, and calculated the average download time and energy consumption. The scenario was to download a popular, heavily seeded 105-Mbyte torrent with 128-kbyte piece size to the phone. During the tests, the swarm had around 500 peers and around 1:5 leecher/seeder ratio. For comparison, we also downloaded the same torrent with SymTorrent, the phone-based standard BitTorrent client, and CloudTorrent, our Amazon EC2-based centralized proxy solution.

Measurement Results

The third-generaton (3G) and WLAN-based measurement results are shown in Fig. 2. The router-based proxy results are labeled "ProxyTorrent" plus the chunk size used for the transfer (e.g., CS 5 means five pieces are sent to the phone in one chunk).

ProxyTorrent outperforms both SymTorrent and CloudTorrent. Compared with SymTorrent, using the router-based proxy consumes 40 percent less energy with 3G and 55 percent less with WLAN. As expected, doing transfers at higher speeds significantly improves the energy efficiency. In addition to better bandwidth utilization, shorter download times and lower protocol overhead contribute to energy savings.

Figure 2 also shows that transferring data in larger chunks reduces the energy consumption of 3G downloads to some degree (the difference between using 1 and 10 chunks is 9 percent), but with WLAN the chunk size does not influence energy consumption. The reason for the different behavior is the power saving mechanisms of these two radio technologies; WLAN can switch the radio interface to idle mode rapidly after a transfer has finished, while with 3G returning to idle state after a completed transfer typically takes around 10 s (the exact time depends on the operator). Thus, with WLAN the best strategy is to send each received piece immediately to the mobile because it reduces the total download time. In contrast, for 3G bigger chunk sizes are better.

One strange phenomenon is the relatively poor performance of CloudTorrent. CloudTorrent first downloads the full torrent to the cloud server; then the client starts requesting the content file by file via HTTP requests. During the measurements, we experienced transfers to the mobile from the cloud being much slower than transfers from the proxy. A possible explanation is the different geographical locations (the router proxy was in the same city as the phone) and the different technologies used for transferring the data (CloudTorrent used parallel HTTP transfers with complete files; ProxyTorrent transferred smaller chunks of data via TCP connections). Nevertheless, CloudTorrent still outperformed the dedicated mobile client with both 3G and WLAN, and the proxy-based solution was more energy-efficient and faster than both alternatives.

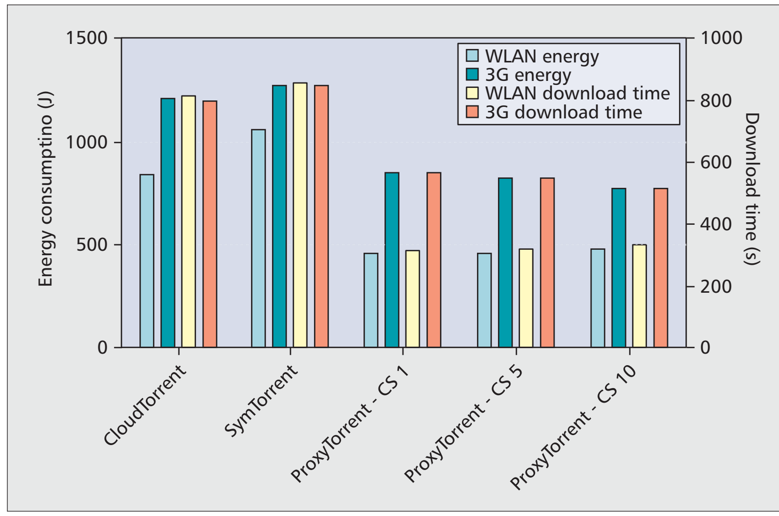


Figure 2. *Measured energy consumption and download time of a 105-Mbyte torrent with different systems and access technologies. Different chunk sizes (CS) were used in the three ProxyTorrent measurements.*

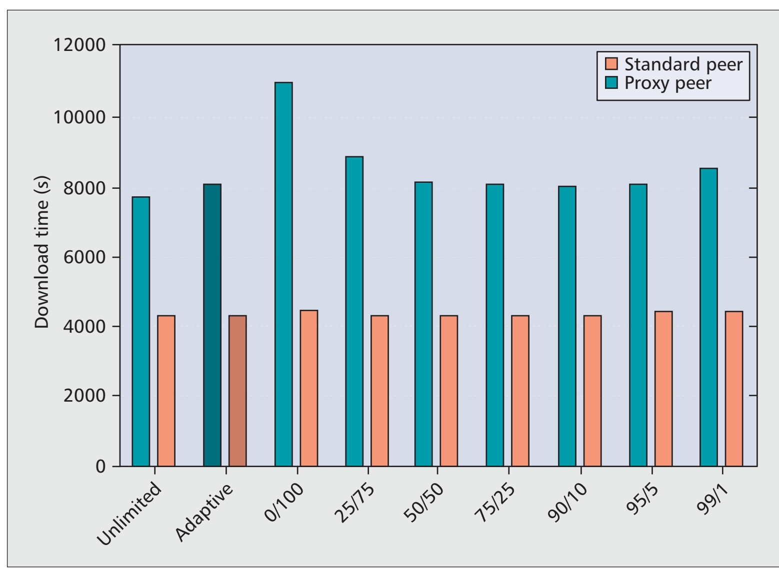


Figure 3. *The simulated effect of different memory allocation alternatives to torrent download times with 5 percent proxy peers.*

BitTorrent Swarm Simulations

Simulation Setup

The goal of the simulations was to investigate how the system would work on a larger scale and to validate the proposed adaptive upload/download buffer allocation algorithm. We used a modified version of a discrete-event BitTorrent simulator originally developed by Microsoft Research [9]. The simulator models most BitTorrent mechanisms, including tit-for-tat, choking, and rarest first piece selection. Support for end-game mode is missing, which increases the download time of the last few pieces of the torrent; however, since all of the simulated peers suffer from this, their relative performance is not affected significantly. The network model is based on a fluid model of connections, which assumes that the flows traversing a link share the link bandwidth equally. The dynamics of TCP connections are not modeled. Thus, it is assumed that the bottleneck link is either the uplink of the sending node or the downlink of the receiving node.

The simulated proxy peers were configured to use chunk size 1 and stopped uploading to other BitTorrent peers while uploading to the mobile.

Simulation Results

In the simulations, we focused on two cases: a swarm with a small number of proxy peers (5 percent) and a swarm with 50 percent of proxy peers. We used different static upload/download buffer allocation ratios as well as the adaptive strategy introduced earlier. It should also be noted that in contrast with the measurements, a larger, 500-Mbyte torrent was used, so that different buffer allocation strategies have a more visible effect. The results are shown in Figs. 3 and 4. The labels on the X axis refer to the different allocation strategies; for example, 25/75 means that 25 percent of the storage was used for the upload buffer and 75 percent for the download buffer. We also simulated "unlimited" buffer capacity, which equals using standard BitTorrent without any limitations.

In both the 5 and 50 percent simulations, choosing the right allocation strategy has a significant effect on the download speed of the proxy peers (the download time difference between 0/100 and 75/25 is more than 25 percent). Furthermore, it is clearly visible that increasing the size of the upload buffer (and decreasing the size of the download buffer) is beneficial until some point (somewhere around 90/10), but after that, further decreasing the size of the download buffer starts to have a negative effect on the proxy peer since the client cannot do enough parallel transfers. Using only 25 percent of the memory (3 Mbytes) for uploads, which is less than 1 percent of the full size of the torrent, already resulted in a major increase in download performance, compared with the free-rider case.

In the 5 percent case, the download speed of the standard peers is only slightly affected by the buffer allocation strategy, which is not surprising, since the low number of limited peers cannot have a real impact on the swarm's performance. In the 50 percent case, however, using different buffer sizes has a more significant effect, since a larger portion of the peers and their capacities are affected. Although proxy peers start to suffer if they have an undersized download buffer, standard peers always benefit from a larger number of upload slots.

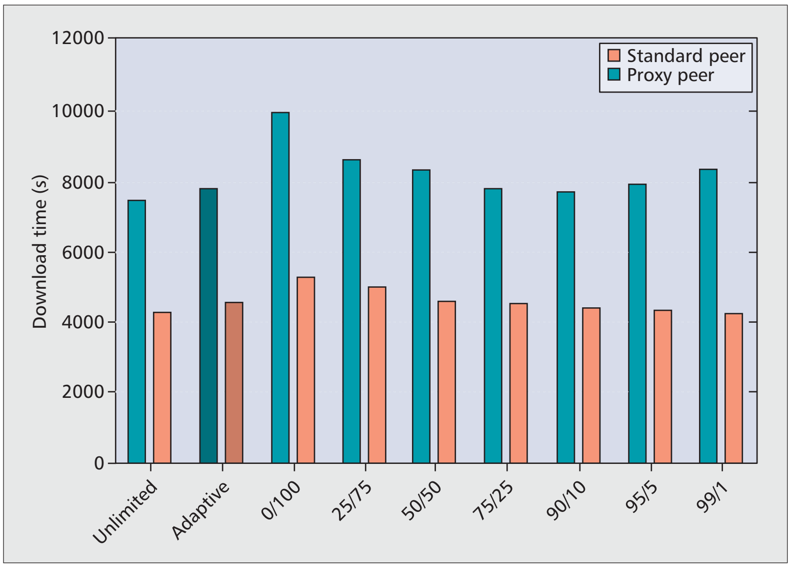


Figure 4. *The simulated effect of different memory allocation alternatives on torrent download times: the case of 50 percent proxy peers.*

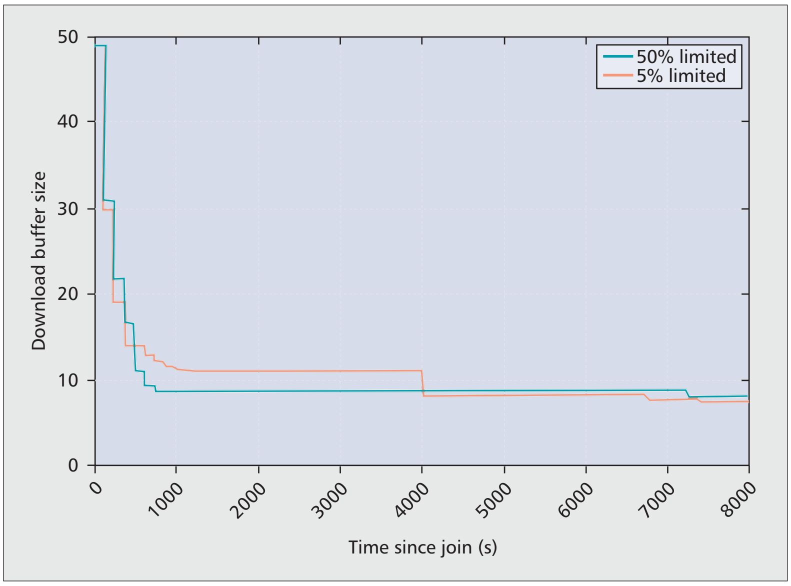


Figure 5. *Simulated effect of the adaptive buffer allocation algorithm. The average share of download buffer of proxy peers as a function of the time spent since joining the network.*

The adaptive strategy performed well, matching the best download times achieved with static allocation strategies. Figure 5 shows how the adaptive algorithm controls the download buffer size as time passes (the charts show average values over all proxy peers). The buffer allocation gradually settles to around 90/10 upload/download ratio. A slight difference between the 5 percent and 50 percent cases is also visible: in the swarm with the larger number of standard peers, the proxy peers are also served a bit more frequently, thus their download buffer size remains at a higher level for a longer period.

Related Work

So far the energy efficiency of BitTorrent, or peer-to-peer file downloading in general, has mainly been investigated from two different angles: how to reduce the standby energy consumption of PC peers [10, 11] and how to perform active content download energy-efficiently with a mobile peer [2, 12, 13]. In another track, the use of helper nodes to speed up torrent downloads has shown promising results [5, 4]. The use of memory-limited devices to help torrent downloads seems to be absent in prior work.

In general, the use of proxies to improve energy efficiency of mobile applications has been widely studied. For instance, Flinn and Satyanarayanan [14] discuss how to create energy-adaptive applications based on proxies and Shenov and Radkov [15] use a proxy to improve the energy-efficiency of streaming content to mobile devices.

The nano data center concept [16] provides services with hardware distributed at homes. This operator-controlled community solution is based on enhanced versions of normal routers. While savings in the overall energy consumption is a key target, the focus is on the infrastructure side, not on the energy-efficient operation of the terminal devices.

Conclusions

In this article, we have analyzed how proxies running on resource-limited broadband routers can help mobile phones download content with BitTorrent. The particular challenge of hosting the proxy on a router is that the memory capacity is very limited. Therefore, new mechanisms for efficient memory use are needed.

The most important conclusions of this study are:

- The proxy peer solution helps the mobile device save battery power when downloading BitTorrent content. The amount of energy saved depends on several factors, including the popularity of the torrent, and the bandwidth of the proxy and the mobile phone. In comparison to downloading the torrent directly to the phone with a native client, we can save around 40–50 percent energy according to our measurement results.
- The performance of the router proxy depends strongly on efficient use of the limited memory space of the router. The amount of memory allocated to serve uploads to other peers strongly influences the torrent download speed. We introduced an adaptive buffer allocation algorithm

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

that can be used to determine the optimal allocation ratio. The download speed is also influenced by the amount of data that is sent to the mobile in one pass, especially in 3G networks.

• The energy consumption of the mobile phone depends on both the torrent download speed and the chunk size. Higher download speeds and bigger chunk sizes result in energy savings. This is especially true with 3G cellular networks. With WLAN connections, using optimal parameters is less important.

REFERENCES

[1] SymTorrent, http://symtorrent.aut.bme.hu.

[2] I. Kelényi and J. K. Nurminen, "CloudTorrent – Energy-Efficient BitTorrent Content Sharing for Mobile Devices via Cloud Services," *Proc. 7th IEEE CCNC*, Las Vegas, NV, 2010.

[3] J. K. Nurminen, "Parallel Connections and their Effect on the Battery Consumption of a Mobile Phone," *Proc. 7th IEEE CCNC*, Las Vegas, NV, 2010.

[4] P. Garbacki *et al.*, "2Fast: Collaborative Downloads in P2P Networks," *Proc. 6th IEEE P2P*, Los Alamitos, CA, 2006.

[5] J., Wong, "Enhancing collaborative Content Delivery with Helpers," Master's yhesis, Univ. of British Columbia, 2004.

[6] DD-WRT, http://www.dd-wrt.com/.

[7] Enhanced CTorrent, http://www.rahul.net/dholmes/ctorrent/.

[8] G. Bosch Creus and M. Kuulusa, "Optimizing Mobile Software with Built-in Power Profiling," F. Fitzek and F. Reichert, Eds., *Mobile Phone Programming and Uts Application to Wireless Networking*, Springer, 2007.

[9] A. Bharambe, C. Herley, and V. N. Padmanabhan, "Analyzing and Improving a BitTorrent Network's Performance Mechanisms," *Proc. 25th IEEE INFOCOM*, 2006.

[10] J. Blackburn and K. Christensen, "A Simulation Study of a New Green BitTorrent," *Proc. 1st Int'l. Wksp. Green Commun.*, 2009.

[11] M. Jimeno and K. Christensen, "A Prototype Power Management Proxy for Gnutella Peer-to-Peer File Sharing," *Proc. 32nd IEEE Conf. Local Computer Networks*, 2007.

[12] J. K. Nurminen and J. Nöyränen, "Energy-Consumption in Mobile Peer-to-Peer — Quantitative Results from File Sharing," *Proc. 5th IEEE CCNC*, Las Vegas, NV, 2008.

[13] I. Kelényi and J. K. Nurminen. Energy-Efficient Peer-to-Peer Content Sharing Using Application-Level Traffic Sharing," *Proc. 4th ACM PM2HW2N*, Tenerife, Spain, 2009.

[14] J. Flinn and M. Satyanarayanan, "Energy-Aware Adaptation for Mobile Applications," *Proc. 17th ACM Symp. Operating Systems Principles*, Charleston, SC, 1999.

[15] P. J. Shenoy and P. Radkov, "Proxy-Assisted Power-Friendly Streaming to Mobile Devices," *Proc. SPIE*, 2003.

[16] V. Valancius *et al.*, Greening the Internet with Nano Data Centers," *Proc. CoNEXT '09*, 2009.

BIOGRAPHIES

IMRE KELÉNYI (imre.kelenyi@aut.bme.hu) is currently an assistant professor in the Department of Automation and Applied Informatics, Budapest University of Technology and Economics (BME), Hungary. He received his Master's degree in technical informatics from BME in 2007. He completed BME's Ph.D. course in 2010. Currently he is preparing his Ph.D. dissertation on energy-efficient mobile peer-to-peer systems, which is also his primary research interest. He is the creator of the world's first ever mobile BitTorrent and Gnutella clients, SymTorrent and Symella. He has worked as a contract software engineer for various companies including Nokia, Nokia Siemens Networks, and T-Mobile.

ÁKOS LUDANYI (llutyo@gmail.com) received an outstanding B.Sc. degree in computer engineering from BME in 2010. Now he is attending an M.Sc. course on applied informatics at BME. His research interest is mobile P2P applications. He worked at NavNGo, developing car navigation systems. Currently he is working as a software engineer for BME's Applied Mobile Research Group and Nokia.

JUKKA K. NURMINEN (jukka.k.nurminen@aalto.fi) is a professor of data communication software at Aalto University in Finland. Before returning to academia he spent almost 25 years at Nokia Research Center working on a variety of topics related to mobile applications and services. His current interests are mobile computing, peer-to-peer applications, and the energy efficiency of distributed solutions. He received his M.Sc. degree in 1986 and Ph.D. degree in 2003 from Helsinki University of Technology.

The energy consumption of the mobile phone depends on both the torrent download speed and the chunk size. Higher download speeds and bigger chunk sizes result in energy savings. This is especially true with 3G cellular networks. With WLAN connections, using optimal parameters is less important.

TakeoverTV: Facilitating the Social Negotiation of Television Content in Public Spaces

Greg T. Elliott, David Carr, and Henry Holtzman, MIT Media Lab

ABSTRACT

Social television, in essence, is a way to share and experience media with others. Rather than attempt to bridge physical distance by creating new social experiences, we have built a system called TakeoverTV that augments canonical social interactions by addressing three key questions: How can we better support social gatherings around media in public and private places?; How can we enable conversation across our devices rather than forcing us to communicate through them?; and How can we help people come to agreements about what media they want to watch together? Primarily aimed at public spaces like bars, TakeoverTV is a system that lets local users influence and interact with the movies and television programs shown on public displays. Our system collects the media preferences of all people physically present at a given location, lets users start a vote among all members of the space, and allows everyone to participate using tangible physical objects like tables in addition to smart devices. Over time, public establishments can evolve complex identities based on the preferences and votes of their patrons while generating valuable analytic data for patrons, establishment owners, content-providers, and advertisers alike.

INTRODUCTION

TakeoverTV is a platform that allows public displays to become aware of the content preferences of nearby users, allows these users to visualize the collective preferences of those present, and to start and participate in votes that determine what is shown on the display (Fig. 1). When a user arrives at a location, their likes and dislikes automatically influence what can be shown on local displays (Fig. 2). Those that want more control can start a vote to choose a new show using their beer glass or phone. The system supports a continuum of lean-forward and lean-backward interaction strategies, allowing public displays to scale from passive broadcasting (current model) all the way up to full social negotiation where all users vote and interact via mobile phones and augmented tables to decide what should be shown and when.

Specifically, our system supports these key features:

- Exposing common interests among co-located users
- Allowing users to collectively decide between a simplified set of common interests (to prevent information overload) using both lean-back and lean-forward interactions
- Evolving media identities for physical locations; for example, bar X becomes a place to watch "Lost" [1] and bar Y becomes a place for fans of the show "24," [2] creating unique opportunities like allowing owners to co-create and evolve their establishment along with their clientele

SCENARIOS

PHASE I: ONE TAKEOVERTV LOCATION

John and his friends head out on a Wednesday evening for a casual night out. After grabbing a seat, they notice that the TV's in the bar are airing an episode of American Idol. They dislike the show and decide to change it by starting a vote with TakeoverTV. Next, they slam their drinks down on the table to begin a vote, and all patrons are encouraged to choose what they would like to watch. TakeoverTV has been silently collecting the preferences of all current patrons, so when a vote is started the system uses a heuristic to narrow down voting options to the top three common interests among the crowd. People vote by knocking their beers on the table or by mobile phone, and as the timer counts down to zero, the crowd gets rowdy as patrons try to coerce others to change their votes. The vote ends, a show is chosen and begins to play. Patrons turn their local table volume up or down, depending on the outcome of the vote.

PHASE II: MANY TAKEOVERTV LOCATIONS

Now that TakeoverTV is installed in multiple bars and restaurants throughout their town, John

0163-6804/11/$25.00 © 2011 IEEE

and his friends have begun to identify with specific establishments. They know that Jackie's Diner tends to attract fans of trashy dramas and that The River Lounge pulls in a college crowd that prefers animated comedies in the evening like *Futurama* [2]. TakeoverTV has aggregated the preferences of patrons, allowing John and his friends to check the TakeoverTV analytics on their phone as they search for a bar. Since they are in the mood for *The Simpsons* [2], they decide that The River Lounge is a good spot. After arriving at The River Lounge, they see that another animated show, *King of the Hill*, is playing. John and his friends begin talking to others at the bar and find that they seem to have a lot in common, due to the bar having an identity that helps like-minded individuals find each other.

Phase III: TakeoverTV Locations in Multiple States

Content providers and advertisers use TakeoverTV as a way to introduce new content to proper demographics. Companies that create new shows can sponsor pilot show nights where they pay bar owners to place a new pilot show as a voting option. If patrons vote to see the pilot, the bar buys a round of drinks. Establishments that tend to prefer stand-up comedy routines get hand-selected for private shows from the comedian most popular at that location, and only loyal attendees get tickets. Bar owners use anonymized TakeoverTV metrics and analytics to better understand their customers and have drink special nights for the most influential voters. The identities of locations are complex time-based entities that grow and change as the patrons change, and users, owners, and content providers all benefit from the feedback loop provided by TakeoverTV.

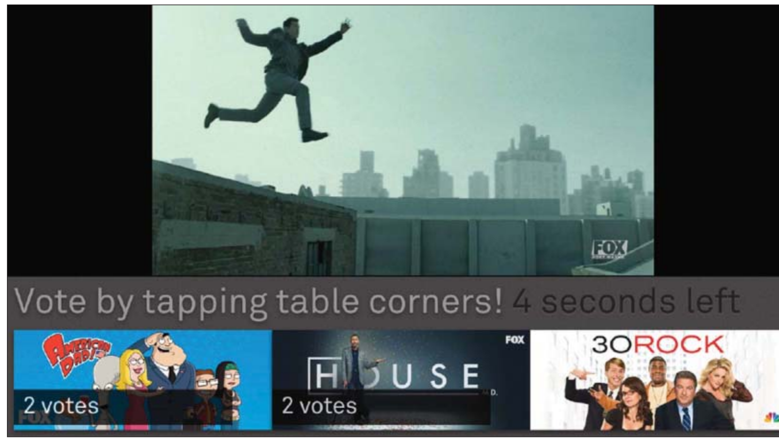


Figure 1. *TakeoverTV large display interface while a vote is active. The tiles at the bottom represent the three most popular shows among local patrons.*

Figure 2. *TakeoverTV in non-voting state. Local user photos are shown at bottom right so that they know their preferences are taken into account. Photos scale with larger numbers of users to prevent crowding.*

Key Features of the System

In order to answer the original questions postulated in the abstract, we first discuss the key features of the system and then follow with a section that addresses the questions.

Exposing Common Interests

TakeoverTV collects the interests of all users present at a location, and uses a heuristic to determine which shows should be included as options in a vote. The system narrows down the options to prevent obvious information overload, and we offload this work onto the system. First, the heuristic filters out noise by ignoring transient patrons that come and go quickly. Second, the algorithm orders each user's preferences by the last time that show was played while the user was present. Finally, it finds the most common elements of the top choices of each user and presents the top three choices for all users to vote on.

Furthermore, the preferences, locations, and votes cast by users create valuable data that can be used for metrics and analytics. For instance, TakeoverTV can begin to disentangle who is voting for which shows at which times in which locations, as well as how the voting is occurring. Are shows interrupted? Is there a high rate of churn? What patterns in the data expose user preferences and tastes? Our metrics allow us to answer these questions. Additionally, we maintain privacy of the system by anonymizing data and presenting metrics only in aggregate.

Group decision making is a rich field of exploration all to itself [3–5], and we have built an architecture that supports multiple decision making processes. We chose to start with voting as a first-order decision to provide legible feedback from the system, allowing users to understand and unravel how the system might work simply by using it. More advanced techniques likely will provide better results, although they risk losing transparency. We plan to explore additional techniques in the future, and we encourage others to explore novel techniques to apply to the platform.

Collective Decisions

At any time, users can attempt to change the movie or show being played by starting a vote. They do this by either tapping their glass on the table corner that starts a vote (lean-back) or using the TakeoverTV application on their

TakeoverTV adds an additional dimension that moves beyond the superficial by providing metrics and analytics on the movie and show interests of guests. A lounge may attract sports fans in the afternoon, while it attracts a college crowd interested in animated comedies at night. Our system exposes these trends.

phone (lean-forward). The tables at a TakeoverTV location are able to detect tapping in distinct regions, allowing users to interact with the system using only their hands or a glass.

Once a vote is started, a two-minute countdown begins, and all displays switch to voting mode (Fig. 1) visible to all users in the location. Users then vote for their favorite option by tapping their glass on the table corner that corresponds with that option, or selecting the option on their phone. When the countdown ends, the winning show begins playing. For now, we allow voting at any time, although we may decide to limit it to start only at the end of a show.

It is important to note that the system scales easily to support several interaction styles in order not to disrupt patrons' experiences. We define and support multiple interaction strategies: *passive*, *semi-passive*, *semi-active*, and *active*. Passive users do not pay attention to the displays in public locations and thus TakeoverTV has no effect on their expected experience because it does not subsume their experience. Semi-passive users might cast a vote by tapping their glass on a corner, but won't start a vote. Semi-passive users' preferences are still included in the vote, and can interact if they wish. Semi-active users start votes by phone or tapping on the table, and will try to convince their friends to vote for their choice. Active users will not only attempt to convince friends, but also will attempt to convince strangers in the area by shouting to strangers.

By design, if no one starts a vote, TakeoverTV completely fades into the background.

Public Location Identities

TakeoverTV becomes significantly more valuable over time — locations gain an additional dimension of identity from the history of their patrons' preferences and votes. Currently, a person may characterize a public location by the decor, the dress of the patrons, the music choices, and so on. TakeoverTV adds an additional dimension that moves beyond the superficial by providing metrics and analytics on the movie and show interests of guests. A lounge may attract sports fans in the afternoon, while it attracts a college crowd interested in animated comedies at night. Our system exposes these trends.

This has implications on how users might interact. In the way that dating websites often rely on questions about the types of music or television shows a person likes as a way to extrapolate taste [6–8], we believe these tastes can be useful in choosing a location for a night out. Users may choose one bar over another if the inhabitants of that bar tend to like the shows and movies they like, as these preferences indicate taste in humor, style, and attitude [6–8]. If users can expect similarities in taste, they may be more willing to strike up conversations with strangers. We plan to test the system in key locations around Cambridge, Massachusetts, in hopes of supporting these claims.

Furthermore, the metrics provided by TakeoverTV can help establishment owners co-create and co-evolve their location with their clientele. Currently, owners can only guess at the preferences of guests, but our system can provide actual evidence for media preferences. Not only can they better understand their target audience, but they can also explore niche groups that are perhaps poorly supported.

Finally, content providers are able to screen pilot shows to appropriate demographics and gauge reception by audience response. A content provider pays a bar owner a set amount to have their pilot show appear as a vote option for a limited time, and if that pilot is voted for and watched, patrons get a free drink or some other incentive. We can then offer interactions like voting off the pilot if it is reprehensible, and pass that information on to the content provider. Similarly, advertisers can better understand when their targets are visiting locations and how to better appeal to them. By analyzing the metrics provided by TakeoverTV, advertisers can learn from real-time preferences and trends in a currently unrepresented market.

How TakeoverTV Addresses Key Questions Proposed

In the abstract, we introduced three primary questions that transform everyday situations into beneficial social television scenarios. We now discuss how our system addresses these questions:

Encouraging Social Gatherings around Media

TakeoverTV helps people collectively vote on what to watch both at home and in public locations like bars. It aggregates their preferences and simplifies the choices to reduce strain, allowing people to focus on each other and the social gathering instead of the technology. Using a well-known design style to which we refer as "Just-in-Time Interface Design," the technology only becomes visible when it is needed, and when it appears, it is clearly identifiable as technology. The system encourages discussion about show preferences, instigates negotiation of decisions in a controlled, simplified way, and creates common ground among local users. This provides unique opportunities for conversations between strangers and public groups. Over time, we expect the identities locations develop will create a desire to choose one location over another, and motivate people to explore the social circles around them.

Conversation across Devices, Not through Them

We could have designed the system to only support mobile phones so that interaction with the system required users to stare at a small screen and disconnect themselves from their immediate social environment. Instead, we provide multiple access points — the table and the phone — that allow interaction to become part of the social setting and even enhance it. Convincing friends to vote for shows can cause users to physically grab another's glass and tap it, or to encroach on their space and tap the table. We purposefully designed the system to evoke playful behavior that breaks people away from the myopic interactions associated with mobile phones.

Lastly, we designed the tables to look ordinary to avoid detracting and disrupting the local ambience [9]. The technology that underlies the table could be added to any standard bar table, and users would not know that anything had changed. This is critical to meshing with the existing social norms at public places like bars.

Coming to Agreements about What to Watch

As noted earlier, group decision making is an extensive field, and our architecture supports a wide variety of decision making styles. By first narrowing the choices to a good fit for local users, then offering ways to change what is shown through social negotiation, we reduce the complexity of choice while increasing the social engagement required to get others to vote for your choice. This first-order scaffolding for group decision making is introduced into an area that currently has no support for decision making (rarely do patrons have any influence over what is shown), and we plan to apply more advanced techniques in future work.

Implementation

Software

Our prototype system is powered by an Apple Mac Mini connected to a 37-in display mounted at the MIT Media Lab (Fig. 3). The server side consists of a Tornado python server (running Nginx) with a MySQL database. The front-end display is a Cocoa/Objective-C application that renders video files natively and displays voting content via a WebKit object that fetches content to display from the server. This python server determines which users are present, reads their preferences, and determines voting choices, rendering all this data to the WebKit object. The use of a web-based front- and back-end system allows for cross-platform deployment. The Mac Mini provides a nice wrapper for media playback, but is not critical and could be replaced by other options.

Currently, each user manually specifies show preferences on their iPhone; however, in the future we plan to use the Netflix application programming interface (API) to read in users' preferences automatically.

The iPhone application is written in Objective-C and communicates with the server via the same RESTful interface the displays use. It allows users to log in via Foursquare or a custom RFID number, to start a vote for the location they are at, and to cast and change their vote.

User detection is done by both RFID and via the Foursquare API. Users can "check in" to the TakeoverTV location at the Media Lab, and their picture is automatically shown on the displays. While this is a more realistic interface for current scenarios, we also built a passive login system using RFIDs. We use ThingMagic [10] Vega readers — operating at an EPC GEN2 compliant UHF (850–950 mHz) — and Alien Technology's ALN-9640 Squiggle(R) inlays, encoded and individualized using a Zebra model R100Xi label printer. The tags look like normal name tags, and we also support credit-card-sized battery-powered RFID cards that can be read through a wallet in a pocket. This allows anyone with a name tag or card to simply be near the system, and have their preferences included in the set and start to cast votes.

We chose to use Foursquare (1 million users) because location-based "check-in" services like Foursquare, Gowalla, Brightkite, and Yelp are growing at exponential rates, allowing people to "check in" at locations they visit as a way to find other friends, coordinate gatherings, or simply gloat about the places they go. Additionally, Google's Latitude (3 million users) is another service that lets users find each other based on their location and is said to be moving toward "check-in" behavior [11].

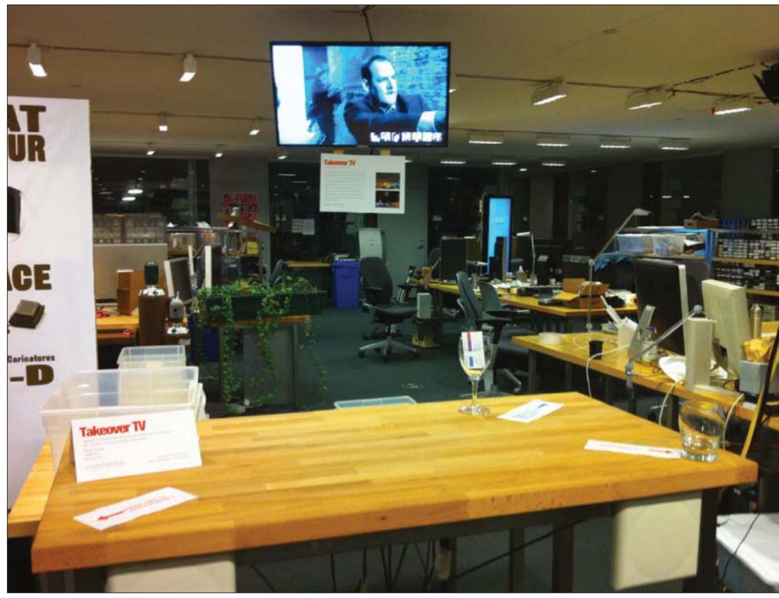


Figure 3. *The TakeoverTV table equipped with Mac Mini, tap sensors, integrated audio transducer, RFID readers (final version should have readers mounted in establishment, not at table), and offset display to mimic ambient placement in public spaces.*

Hardware

The TakeoverTV table hardware is tasked with delivering the television audio experience to each table, and also providing a means for patrons to vote without using the phone interface. It was also important that the hardware feel natural in a bar setting. The audio had to be delivered in a clearly audible fashion at each table, but not be intrusive enough to overpower normal bar conversation. It was also desired that the voting mechanism be invisible when not wanted, and not feel "electronic" even when in use. Lastly, it had to be resistant to spilled drinks and rough handling by bar patrons.

IEEE Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

In the near future, we plan to deploy this system and study how these emerging conversations and identities affect the locations people choose, how they characterize themselves, and how they blend their personality with the personality of the space.

The prototype audio system receives wireless audio over a Bluetooth link and delivers the audio signal to a vibration transducer mounted on the underside of the table. Using a wireless protocol allows the table to move around the bar freely and eliminates any visible wires. The audio transducer actually radiates sound through the wooden table surface itself. This creates a sound field that is clear within a two- to three-foot radius, but rapidly dissipates outside that distance. The transducer is completely shielded by the tabletop and dispatches the need for any grills or exposed electronics.

Continuing on the theme of the audio system, the voting hardware is also hidden completely on the lower table surface. It consists of four bimorph piezoelectric sensors located near the table corners [12], an AVR series microcontroller for signal processing, and a Bluetooth serial link to the TakeoverTV control system. During a vote, the three show choices are mapped to three of the table corners. By tapping (or slamming) a drink near one of these corners, a patron can vote for the corresponding show. The fourth table corner is reserved for other functions (e.g., initiating a vote). The outputs of each piezoelectric sensor are un-amplified but are biased to approximately midscale on the microcontroller analog-to-digital converter. The microcontroller samples all four sensors at approximately 35 kHz to detect the rising (or falling) edge of the drink impact waveform. The first corner sensor to receive this signal is the closest to the impact site, and this selected corner is transmitted back to the TakeoverTV system.

We chose to hide the electronic underpinnings to encourage patrons to emphatically slam their drinks down when voting for their favorite show to fit with bar culture. In our limited experience with approximately 40 visitors, this voting mechanism is durable, accurate, and generates positive user interactions.

While the hardware system outlined above works well, there are several opportunities for improvement. First, our testing has shown that Bluetooth is a suboptimal communication mechanism in this application. It is not robust in challenging RF environments, requires stateful connections, has relatively high power consumption, and is expensive compared to other low-data-rate radio technologies. More promising candidates are low-power stateless technologies like those employed in low-cost data chipsets by Texas Instruments (formerly Chipcon) and Nordic Semiconductor. On the audio delivery front, low-power analog FM radio links seem to be a viable alternative. While somewhat dated, this technology is extremely robust, offers a handful of audio channels, low power consumption, and low cost.

Another shortcoming of the current system is that each table contains a battery that must be periodically charged or replaced. With a bit of labor one could simply replace low battery packs after closing or before opening each day. Alternatively, table power consumption could be greatly reduced by moving to an overhead fixed audio system. This potentially limits table positions and/or mandates more intrusive audio, but could extend battery replacement intervals to months or years. Lastly, a wireless power distribution system could be employed. Traditionally these systems have had limited range because of the size limits imposed by portable devices, but a table-sized power reception antenna might greatly improve their performance [13].

Future Work

Despite our in-depth implementation, the system is not yet ready for deployment. We must further refine the implementation and improve the implementation so that it is more robust before we can test the system in locations around Cambridge, Massachusetts. We plan to compare the system to other media voting mechanisms such as iTunes DJ, and how the social connections and demographic analysis of TakeoverTV benefit or hurt the experience. We also hope to incorporate an Internet-enabled DirecTV DVR box to provide a wider array of fresh content and present different ways to slice group preferences. Additionally, we hope to learn how to improve user acceptance of the system and combat privacy issues by focusing on opt-in features.

Conclusion

TakeoverTV replaces the typical one-way television experience in public spaces like bars and gyms with a conversation about others' interests and identities. These conversations take place at several levels. First, as people begin to choose locations based on others' media preferences, they are implicitly negotiating territory with friends and strangers. Second, within a given location, TakeoverTV creates an electronically mediated conversation about show preferences. Third, and mostly important, the system encourages and instigates direct person-to-person interactions at a given table and the immediate space around it.

This last point is significant because TakeoverTV avoids funneling all interaction through the system and instead attempts to encourage interaction through increased direct human side channels. In the near future, we plan to deploy this system, and study how these emerging conversations and identities affect the locations people choose, how they characterize themselves, and how they blend their personality with the personality of the space.

Acknowledgment

We thank the MIT Media Lab MAS 960 Social TV community for comments and feedback.

References

[1] "Lost," http://abc.com.
[2] "Futurama," "The Simpsons," "24," http://fox.com.
[3] S. Kiesler and L. Sproull, "Group Decision Making and Communication Technology," *Organizational Behavior and Human Decision Processes*, vol. 52, issue 1, June 1992, pp. 96–123.
[4] K. L. Kraemer and J. L. King, "Computer-based Systems for Cooperative Work and Group Decision Making," *ACM Computing Surveys*, vol. 20, no. 2, July 1988, pp. 115–46.
[5] J. Grudin, "Groupware and Social Dynamics: Eight Challenges for Developers," *Commun. ACM*, vol. 37, no. 1, Jan. 1994, pp. 92–105.

[6] P. Litle, M. Zuckerman, "Sensation Seeking and Music Preferences," *Personality and Individual Differences*, vol. 7, issue 4, Oct. 1986, pp. 575–78.

[7] D. Rawlings and V. Ciancarelli, "Music Preference and the Five-Factor Model of the NEO Personality Inventory," *Psychology of Music*, vol. 25, no. 2, Jan. 1993, pp. 120–32.

[8] K. Schwartz and G. Fouts, "Music Preferences, Personality Style, and Developmental Issues of Adolescents," *J. Youth and Adolescence*, vol. 32, 3, June 2003, pp. 205–13.

[9] P. Cesar, K. Chorianopoulos, and J. F. Jensen, "Social Television and User Interaction," *Computers in Entertainment*, Jan. 2008.

[10] ThingMagic, http://www.thingmagic.com/.

[11] MG Siegler, "Google Latitude Has 3 Million Active Users, Check-Ins Likely On the Way," TechCrunch.com, May 2010, http://techcrunch.com/2010/05/06/google-latitude-users-check-in.

[12] H. Ishii *et al.*, "PingPongPlus: Design of An Athletic-Tangible Interface for Computer-Supported Cooperative Play," *Proc. SIGCHI Conf. Human Factors in Computing Systems: The CHI Is the Limit*, New York, NY, May 1999, pp. 394–401.

[13] A. Karalis, J. D. Joannopoulos, and M. Soljacic, "Efficient Wireless Non-Radiative Mid-Range Energy Transfer," *Annals of Physics*, vol. 323, issue 1, special issue, Jan. 2008, pp. 34–48.

BIOGRAPHIES

GREG T. ELLIOTT (greg@media.mit.edu) worked with all sorts of companies in the industry, from large corporations like IBM, Dell, the BBC, and Steelcase to counter-culture innovators like Thunderdog Studios, Behance, and Obsessable. He received his M.S. from ACE (Arts Computation Engineering interdisciplinary program) in the Informatics & Computer Science Department at the University of California, Irvine. He received a B.S. in cognitive science and computation (a combination of computer science, neuroscience, and psychology) from the University of California, San Diego. Aside from Konbit, he works on novel user experiences that deal with information overload, at-home 3D printing, and a social TV project, TakeoverTV, that lets local users influence and interact with the movies and television programs shown on public displays.

DAVID CARR (dcarr@media.mit.edu) received a B.S. in electrical engineering from Rice University in 2007 and is currently a Master's student at the MIT Media Lab. His interests span all things electronic, although he is currently working on improving the accessibility and performance of personal fabrication machines. He also enjoys working on and flying airplanes near his home in Cambridge, Massachusetts.

HENRY HOLTZMAN (holtzman@media.mit.edu) is MIT Media Lab's chief knowledge officer, co-director of the Digital Life consortium, and director of the Information Ecology research group. In addition, he directs the Lab's CE 2.0 initiative, a collaboration with more than 40 Media Lab sponsor companies to formulate the principles for a new generation of consumer electronics that are highly connected, seamlessly interoperable, situation-aware, and radically simple. His research in tangible networking and RFID led to coining the term "Internet of Things." He has also led research projects on multicast network architectures for multimedia, IP television, scale-free image representation, and knowledge-based video representation. As a member of the MPEG standardization committee, he helped to define MPEG-2 video technology, used in DirecTV, DVD, digital cable, and digital TV broadcasting. He has been granted multiple patents for his inventions.

ACCEPTED FROM OPEN CALL

Architecture of a Network-Aware P2P-TV Application: The NAPA-WINE Approach

Robert Birke, Emilio Leonardi, Marco Mellia, DELEN – Politecnico di Torino
Arpad Bakay and Tivadar Szemethy, Netvisor Ltd.
Csaba Kiraly and Renato Lo Cigno, DISI – Università di Trento
Fabien Mathieu and Luca Muscariello, Orange Labs
Saverio Niccolini and Jan Seedorf, NEC Laboratories Europe
Giuseppe Tropea, Lightcomm SRL

ABSTRACT

Peer to Peer streaming (P2P-TV) applications have recently emerged as cheap and efficient solutions to provide real time streaming services over the Internet. For the sake of simplicity, typical P2P-TV systems are designed and optimized following a pure layered approach, thus ignoring the effect of design choices on the underlying transport network. This simple approach, however, may constitute a threat for the network providers, due to the congestion that P2P-TV traffic can potentially generate. In this article, we present and discuss the architecture of an innovative, network cooperative P2P-TV application that is being designed and developed within the STREP Project NAPA-WINE.[1] Our application is explicitly targeted to favor cooperation between the application and the transport network layer.

INTRODUCTION

Peer-to-Peer (P2P) technology has been recently exploited to offer video service over IP (P2P-TV), as for example done by PPLive, SopCast, TVAnts, CoolStreming/UUSee, and TVUplayer, to name a few commercial systems. Recently, the interest of the research community has started to rise thanks to the opportunities P2P-TV systems offer [1]. In this context NAPA-WINE (Network Aware Peer-to-peer Application over WIse NEtwork), a three-year project (STREP) within the 7-th Research Framework of the European Commission, focuses on improving the performance of P2P live streaming distribution, while protecting network resources from excessive usage. In a traditional P2P-TV system, a video source chops the video stream into chunks of data, which are then exchanged among nodes distributing them to all participating peers. Peers form an overlay topology at the application layer, where neighbor peers are connected by virtual links, over which chunks are exchanged using the underlying IP network. Two networking layers can be identified: the overlay layer in which peers exchange content, and the IP layer in which routers and links transfer the packets. The main goal of NAPA-WINE is the study of novel Peer-to-Peer architectures and systems suitable for High Quality TV live streaming over the Internet: a *P2P-HQTV* system [2]. NAPA-WINE focuses on overcoming the limitations of today's approaches, where the two layers are completely independent and at times antagonists. Instead, we envision a cooperative paradigm in which the application and network layers interoperate to optimize the service offered to end users, as depicted in Fig. 1.

We believe that an approach where the overlay and the IP networks ignore each other, while acceptable for elastic applications or low-bandwidth streaming applications, may cause intolerable performance degradations for high-quality real-time (even for soft real-time) applications such as P2P-TV. If the network becomes heavily congested, the application will never be able to meet its minimum requirements. As a consequence, the only successful paradigm for a large scale P2P-TV architecture is a cooperative paradigm, where the network and the application join forces to meet the quality of service requirements, and to reach the largest possible population of users.

Finally, there is an additional incentive to such cooperation. In traditional P2P systems, only two main actors can be identified — the network providers and the application users — and their interests are often in contrast. In the case of P2PTV, a new main actor comes into play: the content producer, whose interests are mainly in distributing its content through the Internet in the safest possible way, thus avoiding every possible problem, either technical or legal, which could induce users to migrate away. This is a major shift with respect to current P2P applications, where users provide content, a shift that offers further incentives for network operators

[1] *NAPA-WINE (http://napa-wine.eu) is a three-year STREP project started on February 1, 2008 and supported by the European Commission within Framework Programme 7 (ICT Call 1 FP7-ICT-2007-1, 1.5 Networked Media, grant No. 214412). This article reflects its evolution after about two years.*

 0163-6804/11/$25.00 © 2011 IEEE

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

to cooperate with the content providers and the users to successfully deliver the video. Note that cooperation between network providers and P2P applications has already been investigated in the context of file sharing P2P applications (see for example the pioneering work in [3]), but it has never been applied in the context of P2P-TV system design and performance evaluation.

In this context, the NAPA-WINE project is proposing an innovative network cooperative P2P-TV architecture that explicitly targets the optimization of the quality perceived by the users while minimizing the impact on the underlying transport network. The focus is on the study, design, and development of a P2P-HQTV system, in which peers set up a *generic mesh* based overlay topology, over which video *chunks* are exchanged according to a swarming-like approach. A source peer produces the video stream, chops it into chunks, and injects them into the overlay where peers cooperate to distribute them.

The architecture we envision is schematically represented in the Fig 1. An important element of the cooperative P2P-HQTV system is represented by the built-in distributed monitoring tool that allows the application to continuously gather real time information both on network conditions and on users' perceived performance. Information collected by the monitoring tool can be used to trigger reconfiguration algorithms acting both at the level of the chunk distribution mechanism (scheduling) and at the level of the overlay topology reconfiguration. In addition, potentially useful information on the system state can also be exported to the network, so that it is aware of the status of the P2P-HQTV system.

The architecture we propose also allows the network layer to expose useful information to the application layer. As pursued by the IETF in the ALTO [6] working group with the contribution of NAPA-WINE partners, the network provider is given the capability to guide the P2P application, for example by explicitly publishing information about the status of its network, like link congestion or AS routing preferences.

The High Level Architecture

This section describes the main functions and requirements of a P2P-TV system, proposing the high-level architecture for the NAPA-WINE system, whose block diagram is depicted in Fig. 2.

User Module

The "user module" (top left block in Fig. 2) implements the interface between the user and the application. Besides the Graphical User Interface (GUI), it is responsible of video coding and decoding. If the considered peer is the source node, the input video signal has to be converted into a sequence of chunks, which are the atomic unit of data that will be exchanged in the P2P overlay. Depending on the type of video source, this may include analog/digital conversion, encoding, adding forward error correction (FEC) information, etc. The user module supports any available CODEC thanks to a standard interface. Furthermore, it implements a flexible

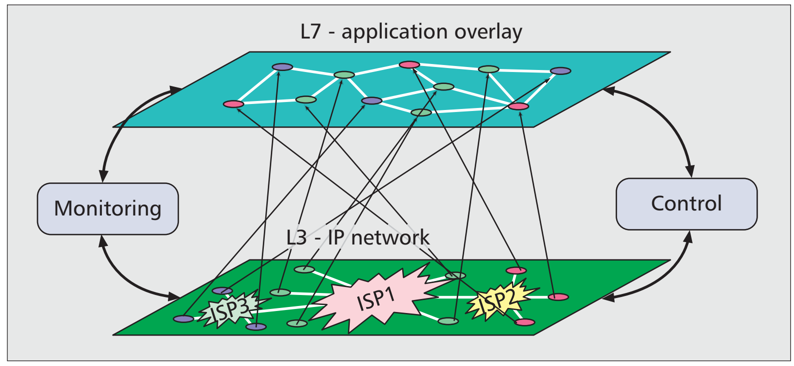


Figure 1. *Schematic representation of the NAPA-WINE approach. Nodes at the network layer (L3) and peers at the application layer (L7) form the network and overlay topology, respectively. Cooperation among the two layers is possible thanks to monitoring and control capabilities offered by the NAPA-WINE architecture.*

chunking mechanism that can explicitly consider the nature of the video, e.g., forming chunks considering frame boundaries and types. When the peer acts as a receiver only, the user module reassembles the audio and video stream from received chunks, so that, after decoding and resynchronization, the video can be displayed.

The user module also runs video quality monitoring algorithms, constantly monitoring the quality of experience the user is perceiving. Finally, several instances of the user module can run at the same time, making it possible to watch a channel while recording another one, or maintaining pre-views of other channels with reduced quality.

Scheduler Module

The "scheduler module" (see Fig. 2) is the core of the information distribution engine. It is responsible for receiving and sending chunks, both from/to other peers via the network, and to/from the local "user modules." Different instances of this module might run at the same time, each taking part in one overlay serving one channel. Cross-channel information diffusion is under consideration and may help to optimize service and channel switching, but is not under development so far.

The scheduler module hosts the chunk and peer scheduling algorithms, which select both what should be offered (or requested) and to (from) whom. The decision is based on distributed information: a local database of the status of other peers stored in the active peers database. Options for designing the scheduler and the impact on overall system performance are detailed later.

Overlay Module

We recall that in unstructured systems peers form a generic highly mesh topology, which guarantees good resilience to churn and great flexibility.

The "overlay module" (top-right block in Fig. 2) selects and updates the peer *Neighborhood*, i.e., the set of peers the local peer exchanges information with. A fairly large number of neighboring peers are maintained to form a well connected overlay structure, as peers may leave an

The selection of peers may be based on the information stored in the repositories, which, as detailed later, store information about both peers and the transport network. The overlay and the scheduler modules are interdependent.

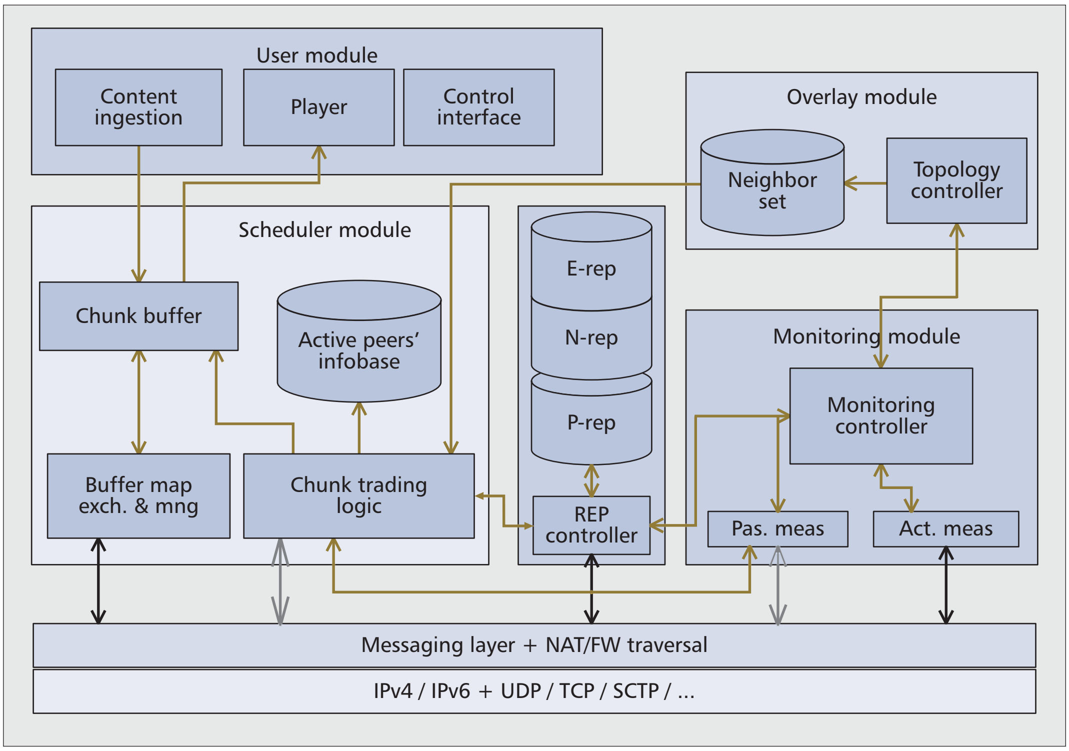


Figure 2. *Global NAPA-WINE architecture. User, Scheduler, Overlay, Messaging, Monitoring, and Repository modules are highlighted using different colors. Arrows represent the relationship among modules, distinguishing data path (gray), signaling path (black), and internal communication (brown).*

overlay at any time, e. g., when the user switches channel. Discovering new peers and establishing connections to them, as well as forwarding information about them to other peers, is the task of the overlay module.

The selection of peers may be based on the information stored in the repositories, which, as detailed later, store information about both peers and the transport network. The overlay and the scheduler modules are interdependent, and for this reason they are detailed together later.

Monitoring Module

Both the chunk scheduler and the overlay manager can greatly benefit from information about the quality of the connectivity with other peers. This includes, but is not limited to, the distance and the available bandwidth between two peers, or the presence of Network Address Translation (NAT). The "monitoring module" (see Fig. 2) gathers this kind of information. It basically has two modes of operation: passive measurements are performed by observing the messages that are exchanged anyway between two peers, e. g., when exchanging video chunks or signaling information; active measurements, in contrast, craft special probe messages which are sent to other peers at the discretion of the monitoring module. The design of this monitoring module is one of the most innovative goals of NAPA-WINE and it will be documented later in detail.

Repositories and Repository Controller

"Repositories" (in the center of Fig. 2) are databases where information about each peer is shared with all other peers. Indeed, the information generated by the monitoring module is not only useful for the local peer, but also for other peers that can benefit from it. Therefore, mechanisms for exchanging measurement values with other peers are useful. Furthermore, it may be useful to have access to information from entities that are not part of the P2P overlay, such as the so-called "network oracles," through which, for example, network providers can share information about the status of the network.

The information acquired by each peer is locally stored and readily available, and it is summarized and exported (published) to the (logically) central repository. The contents of the repository, and data import and export, are managed by the repository controller, which implements a set of APIs to push and fetch information from the repository database.

Messaging Layer

The "messaging layer" (see Fig. 2) offers primitives to the other modules for sending and receiving data to/from other peers. It abstracts the access to transport (UDP/TCP) protocols and the corresponding service access points offered by the operating system by extending their capabilities and providing an overlay level addressing, i.e., assigning a unique identifier to each peer. For example, it provides the ability to send a chunk of data to another peer, which has to be segmented and then reassembled to fit into UDP segments. The messaging layer also provides an interface to the monitoring module, which is used for passive measurements: whenever a message is sent or received, an indication will be given to the monitoring module, which can then update its statistics.

Index	Accuracy	Usage/Criticalities
Hop Count	High	Topology optimization
Packet Loss	High	Triggering scheduling and topology update and QoE monitoring
Chunk Loss	High	Triggering scheduling and topology update and QoE monitoring
Delay Measurements		
Latency	> 10ms	Source and stream absolute time alignment/clock synchronization below NTP precision is hard to obtain
RTT	< 1ms	Chunk and peer scheduling/timestamping accuracy limits precision
Jitter	< 1ms	Delay jitter may be an indication of congestion
Bandwidth Measurements		
Mid term throughput	High	Neighborhood selection and chunk/peer scheduling
Capacity	Low	Pacing chunk scheduling for altruistic peers to avoid clogging a PHY bottleneck (e.g., ADSL links beyond a LAN)
Available bandwidth	Low	Fundamental to avoid congestion in the network/difficult to achieve high accuracy due to correlations and the lack of efficient algorithms to perform the measure

Table 1. *Summary of the network layer measurement available in the Monitoring module.*

Another important feature of the messaging layer is the presence of mechanisms for the traversal of Network Address Translation (NAT) boxes. NAT makes it possible to attach several computers to the Internet using only one public IP address. Therefore, it is very popular with private customers, who are also the main target audience for P2P TV. However, the presence of a NAT device may prevent peers from establishing connections to other peers. Therefore, special NAT traversal functions are offered by the messaging layer.

The Monitoring Module and Repositories

The monitoring module is conceived to collect and share information useful for the construction and maintenance of the P2P overlay topology and for the scheduling process of video chunks. The main objective is to maintain up-to-date information on the quality of the peers' end-to-end path, and to infer information on the network's available resources. Repositories, both local and global, are used to store metrics, estimates, and measured data requiring additional elaboration. Furthermore, a certain number of functions act on these data to produce up-to-date estimates of various parameters that can be published on global repositories through an interface named "Repository Controller." Particular care is taken to optimize both the computational and the memory requirement of the previously mentioned functionalities. The repository controller limits the amount of information that is exchanged among peers through the network to avoid waste of bandwidth. It also verifies that a peer can access the information it is requesting. Indeed, each repository stores data that can remain private and local to the peer, or information that is made publicly available. The distinction between public and private information could be based on security considerations, and also on its usefulness and the cost of its dissemination. For example, the ISP can share information about the status of its network only with clients connected to its network.

In the following, we detail the functionalities implemented in the monitoring module.

Monitoring Components

Measurements are available at the chunk and segment levels, i.e., above and below the messaging layer. Several network layer metrics can be monitored:
- Delay between peers (e.g., Round Trip Times (RTT), Delay Jitter).
- Loss probability.
- Path capacity and available bandwidth.
- Number of hops traversed.

Other more sophisticated measures are under study. Thanks to a simple and modular architecture, measurements can be added as (compile-time) plugins, and activated on demand. The monitoring layer is implemented at every peer, and information it collects is made available to all the peers. Table 1 summarizes the network layer measurements that are currently available in the monitoring module.

The monitoring module is built of three main components.

The Passive Measurements Component (PaM) — The PaM implements the set of measurement functionalities needed to passively monitor exchanges between peers. This func-

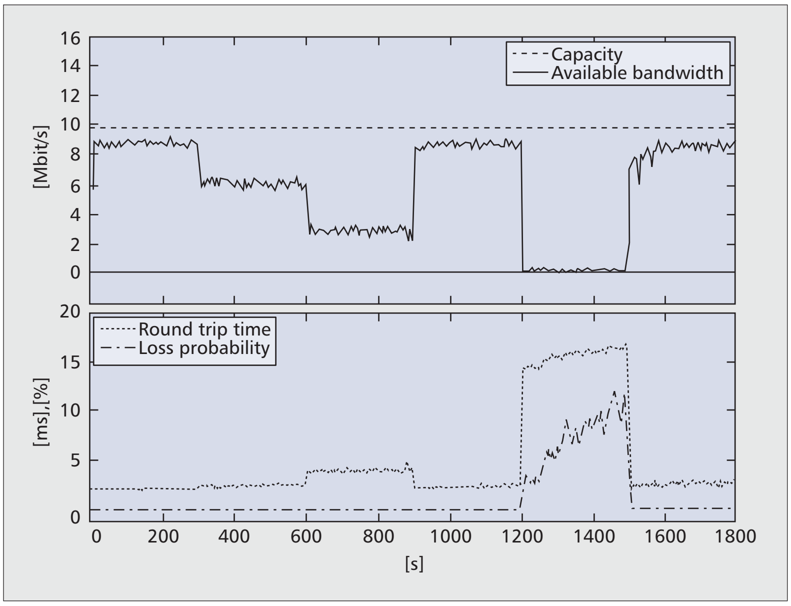


Figure 3. *Measurements obtained from the monitoring module.*

tion is passive as it relies on peers' communications to reduce as much as possible any overhead. Each measure is initiated by the sender, and the receiver cooperates to complete it, e.g., sending back acknowledgment messages. Results of each measurement are then fed to the sender Monitoring Controller, which possibly elaborates them before pushing results in the P-REP (Peer-Repository) and N-Rep (Network-Repository).

The Active Measurement Component (AcM) — The AcM performs active measurements, i.e., it can inject pure measurement messages and data that do not carry any user information. This component possibly runs a number of different measurement probes, either periodically or on request. For example, it may be invoked to estimate the end-to-end status between the local peer and a remote peer to which no chunk is being sent. Therefore, AcM may also be seen as a bootstrap function when limited knowledge of the peers and the network is stored in the P-REP. Active measurement results in a waste of bandwidth, and therefore care must be taken when activating them.

The Monitoring Controller (MON-Controller) — The MON-Controller is in charge of managing all the measurements performed by the PaM and the AcM. It implements the algorithms to decide when to trigger a particular measurement, and to process the results of each end-to-end measurement. For example, the MON-Controller can evaluate average, standard deviation and confidence intervals of a given index; it can identify and possibly discard wrong samples, etc. Considering the global network knowledge, it is supposed to implement the algorithms that infer the network status, e.g., by implementing some network tomography or virtual coordinate systems that are then stored in the N-REP. It is also responsible to push locally cached measurement results into the centralized repository by summarizing them to avoid excessive traffic overhead. The communication of the MON-Controller with the repositories is granted by the Rep-Controller.

Monitoring Example

Figure 3 provides a simple example of measurements obtained with the current implementation of the monitoring module. It shows the measurements of the Capacity and of Available Bandwidth (top plot), and the measurements of Round Trip Time and of Loss Probability (bottom plot). A controlled test bed has been used, in which a 10 Mb/s Ethernet link connects a source peer to a receiver peer, while interfering traffic is injected. In particular, a 4 Mb/s CBR source is active from time 300s to 600s; at time 600s, a 7 Mb/s CBR source is active up to time 900s. Then from time 1200s up to time 1500s, TCP connections are started every 60s. From the figure, it is possible to observe the variation of available bandwidth, the increase of the RTT, and the losses induced by interfering traffic.

Repositories

In the NAPA-WINE architecture, *repositories* are envisioned as global databases, containing information to aid peer selection decisions.

Repositories are key elements in the design to achieve network awareness. They have one additional important role in solving the *bootstrapping problem* of P2P systems: namely, upon startup, a peer needs to obtain a list of neighbor candidates. Repositories are well suited for this role: the peer only needs to have access to a few well-known repository servers in its initial configuration. To increase repository resilience, various approaches (such as database replication and DHT-like technologies) can be studied, which are outside of the NAPA-WINE prototype design scope.

The NAPA-WINE architecture defines three different repositories based on the information they contain.

- **P-REP**: The *peer repository* is a distributed repository storing information about the peers' status and end-to-end measurements performed by the peers' monitoring modules, for instance, average bandwidth over different periods of time, round-trip time, download/upload success rates, etc. Each peer measures and stores locally these quantities, for instance as an $N \times N$ matrix, N being the number of neighbor peers (those discovered and selected by the topology manager). Measures can be exported in an aggregated form to a global (distributed or centralized) database, the global P-REP, for performance monitoring and statistical analysis, possibly including information on obtained quality, visited channels, etc. A centralized prototype has been developed and is used to display on a webservice the swarm topology, its characteristics, and how different architectural choices affect the P2P-TV application.
- **N-REP**: *Network repositories* store network-wide information inferred from P-REP values. The information stored by P-REP is hardly complete about all possible N^2 peers' end-to-end paths. Indeed, each peer typically collects measurements over a subset of all peers. In particu-

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page

lar, when joining a given channel, the new peer has little or no information on the status of other peers, and of the end-to-end path characteristics. To solve this problem, "virtual coordinate" [4] systems and network tomography [5] have been studied in the research community, whose goal is to map peers to nodes into a virtual space, in which distances among nodes reflect the actual distances in the real system, or to infer network properties from a subset of measurements. Both make it possible, for example, to "predict" the quality of an end-to-end path which has never been tested in the past. The N-REP stores results generated by these systems. The topology manager uses this information together with distributed algorithms like Tman to achieve the desired topological properties.

- **E-REP**: The *external repository* is dedicated to storing *out-of-the-box*, semi-static information on the network. It is conceived for network operators that can fill it with non public information, e.g., AS graph and peering points, the ranking of routing paths, routing tables, information on the routing optimization, and the network topology. It can also store intra-domain information of one or multiple ASs. Clearly, access to this information must be protected and granted only to a subset of peers, e.g., to only clients in the AS of the ISP. Therefore, the EREP runs on dedicated equipment installed and managed directly by the network operator or content provider. For example, in the ideal case, the E-REP stores the full knowledge on the AS network; the whole ISP topology is therefore exposed to the P2P application.

The E-REP also has an interface for connecting to Application Layer Traffic Optimization (ALTO) servers. The goal of an ALTO service — as currently being standardized in the IETF — is to provide applications with information they can use to optimize their peer selection [6]. The kind of information that is meaningful to convey to applications via an out-of-band ALTO service is any information that applications cannot easily obtain themselves (e.g., through measurements) and which changes on a much longer time scale than the instantaneous information used for congestion control on the transport layer. Examples of such information are operator's policies, geographical location or network proximity (e.g., the topological distance between two peers), the transmission costs associated with sending/ receiving a certain amount of data to/from a peer, or the remaining amount of traffic allowed by a peer's operator (e.g., in case of quotas or limited flat-rate pricing models). ALTO services can be provided by network operators, third parties (e.g., entities that have collected network information or have arrangements with network operators that enable them to learn such information), or even by user communities. The ALTO-interface gives such parties a concrete method for making this kind of information available to the E-REP. The E-REP can use the standard interface of ALTO servers, providing such ALTO-information to NAPA-WINE peers.

- **Rep-Controller**: The peer accesses repositories via its *Repository Controller* (*Rep-Controller*) modules. This software component implements the repository protocol and exposes it over a client API. A single Rep-Controller is able to communicate to multiple repository servers concurrently, and has advanced services such as data caching and batch publishing of measurement data (to save precious bandwidth).

The repository access protocol is designed for the following basic use cases:

- Publish measurement results made by peers' monitoring modules (or NREP processors) into the global database. This is a simple operation, publishing an *(originator*, $peer_1$, $peer_2$, $measurement_{ID}$, *value*) record. The originator is typically a peer (whose Monitoring Module has generated the data), equal to $peer_1$; $peer_2$ is non null only for peer-pair measurements.
- Retrieve the list of "best" Peer IDs according to given criteria. This is the main "retrieval" primitive, designed to support the topology manager in the neighborhood selection and ranking. The operation is specified as

$$(maxPeers, constr_1, \ldots, constr_n, weight_1, \ldots, weight_m)$$

where constraints are specified on measurement values, and weights form a utility function the result list is ordered by. This allows the formulating of queries such as "return a list of at most 10 peer IDs which are closer to $peer_2$ than 5 TTL hops and have smaller round trip time to $peer_3$ than 500ms and sort the list by the peers' own reported upload bandwidth."
- Retrieve measurement results or inferred values for fine-tuning trading algorithms. This primitive allows the direct query of published (measurement or other) values relevant to a given peer.
- Retrieve repository meta-data such as the list of $measurement_{ID}$s the repository has records for. This allows client peers to parameterize their peer listing queries accordingly.

Since information stored in the repositories is sensitive, the repository controller enforces security and policies so that only authorized peers can retrieve information. For example, access to the information about ISP topology or AS routing stored in the E-REP is granted only to peers actually connected to that ISP.

Scheduler and Overlay Modules

A P2P-TV client needs to communicate very efficiently with other peers to receive and redistribute the huge amount of information embedded in a video stream. In addition, TV being a real time application, information must arrive in a short time and with small delay variation. The application goal is then to deliver all the video information to all peers in the system in the smallest possible amount of time. To reach this goal, a distributed system and algorithm must be adopted. The key enabling factors for efficient communication by a peer are: *who* are the peers to communicate with, i.e., its neighboring peers, and *what* data is exchanged within this neighborhood. The first factor drives the overlay management strategy, while the second factor dictates

Since information stored in the repositories is sensitive, the repository controller enforces security and policies so that only authorized peers can retrieve information. For example, access to the information about ISP topology or AS routing stored in the E-REP is granted only to peers actually connected to that ISP.

The design architecture followed allows flexibility and adaptivity, so that, for instance, chunk and peer selection strategies can be tuned even at run time based on feedback received from other peers either via gossiping or through the repositories.

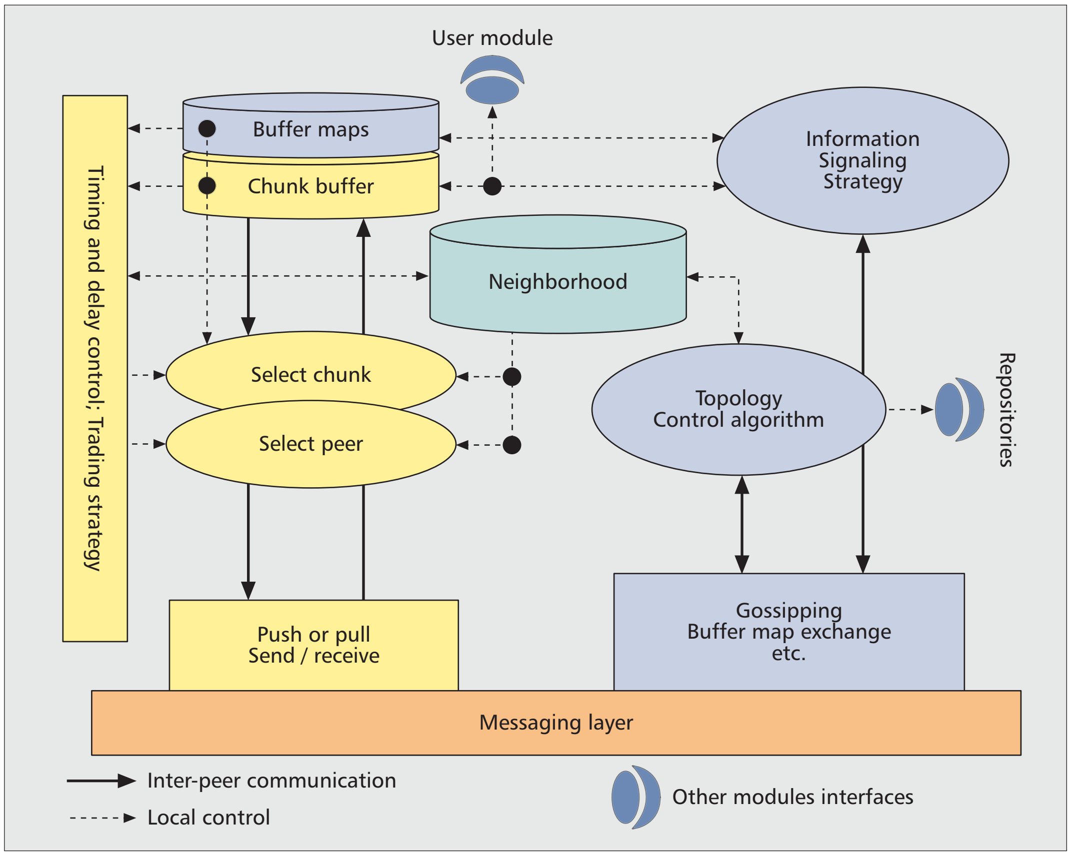


Figure 4. *Logical organization of the scheduling and topology management modules, around the neighborhood data-base.*

the goals of the scheduling algorithms at each peer.

Figure 4 sketches the relationship between the functions that compose the overlay management, and the functions that compose the overall strategy for offering and searching information within the neighborhood of each NAPA-WINE node. Interfaces toward the repositories, the user module, and the messaging layer are also indicated.

The design architecture followed allows flexibility and adaptivity, so that, for instance, chunk and peer selection strategies can be tuned even at run time based on feedback received from other peers either via gossiping or through the repositories.

The Neighborhood database (see Fig. 4) is populated as soon as a peer participates in the distribution of a TV channel. For the sake of clarity we describe both scheduling and overlay management as if only one channel is present, but we are also studying the extension to the joint management of multiple channels.

Always referring to Fig. 4, the components in yellow are related to chunk scheduling, transmission and reception, while those in blue refer to topology management and signaling in general (exchange of buffer maps, i.e., the list of chunks available at each peer, availability to service chunks, etc.). The two functionalities interact through the Neighborhood database as well as the chunk buffer (i.e., the structure where chunks are stored for trading and before play-out), and the related buffer maps of neighbors.

Building and Maintaining Neighborhoods

The overlay network in P2P systems is the result of a distributed algorithm that builds and maintains the neighborhood at each peer. When a peer joins the system, it selects an initial set of neighbors (we call this phase bootstrapping); then the set of neighbors of every node in the system is dynamically optimized over time (overlay maintenance).

The bootstrapping phase is most naturally helped by the Repositories. For the maintenance phase, basing everything only on the Repositories could result in limited scalability and resilience. This is the reason why topology maintenance is performed exploiting information retrieved from the repositories as well as information locally distributed via a gossiping-based mechanism (the latter mechanism will allow the system to work even with Repositories absent).

Culling of neighbors is mainly based on the perceived "quality" of the neighbors; indices that impact this choice are the inter-peer throughput, the measured RTT, the chunk loss rate, etc. The addition of neighbors is instead based on a mixed strategy of optimization (e.g., adding the most stable and resourceful peers) and randomization to avoid fragile topologies (e.g., a group of peers with too few connections toward the rest of the overlay).

Our architecture will make overlay creation and maintenance algorithms much more effective since it offers through the repositories continuously fresh information to the peers about the presence of new resourceful peers. As an

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

example, Fig. 5 shows different overlay topologies obtained using different neighbor selection algorithms. In 5(a), a complete random choice is performed; in 5(b), peers are selected according to the Autonomous Systems they belong to, as reported by the E-REP; in 5(c) peers prefer selected nodes with large upload bandwidth, as reported by the P-REP; in 5(d) high upload bandwidth peers within the same AS are preferred.

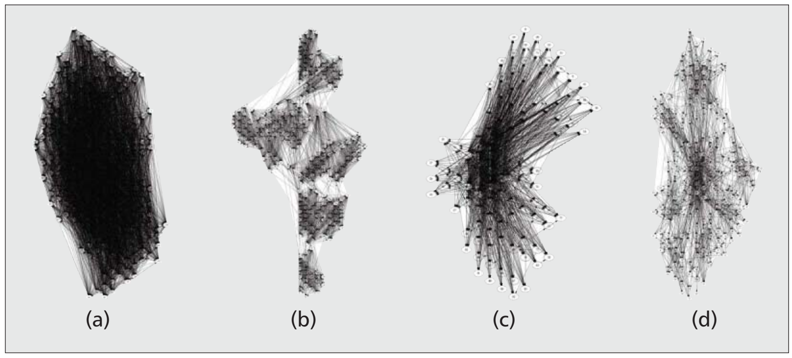


Figure 5. *a) completely randomly chosen neighbors; b) neighbors chosen based on locality information provided by E-REP; c) neighbors chosen based on upload bandwidth peer information; d) combination of b) and c).*

Scheduling Chunks and Peers

Once a topology has been set up, each peer participates in the chunk exchange procedure, with the twin goal of receiving the stream smoothly and in time and cooperating in the distribution procedure. We do not discuss the problems of fairness and security here, because they are not the main focus of NAPA-WINE, and therefore we assume peers are honest and cooperative.

Scheduling the information transfer is probably the single function that most affects performance and network friendliness. This function includes protocol as well as algorithmic problems. First, peers need to exchange information about their current status to enable scheduling decisions. This is a requirement in an unstructured system, where the stream flow does not follow strictly the overlay topology (e.g., a tree). The information exchanged refers to the state of the peer with respect to the flow, i.e., a map of which chunks are "owned" by a peer, and which are missing. This task is carried out by the Information Signaling Strategy block in Fig. 4. This block is in charge of:

- Sending buffer maps to other nodes with the proper timing.
- Receiving them from other nodes and merging the information in the local buffer map data base.
- Negotiating if this and other information should be spread by gossiping protocols or not, and to which depth it should spread in the topology.

Besides the buffer map exchange, the signaling includes Offer/Request/Select primitives used to trade chunks. These messages can be piggybacked on chunks for efficiency.

Another key protocol decision is about *Pushing* or *Pulling* information. A chunk is pushed when the peer owning the chunk decides to send it to some other peer, while it is pulled when a peer needing the chunk requests it from another peer. From a theoretical point of view, as shown in [8], pushing is more effective. Assuming a synchronous system in which peers coordinate to avoid sending more than one copy of the same chunk, in [8] we have demonstrated the existence of an optimal distributed scheduling that achieves the minimum delivery time to all peers, $t_d = \lceil \log_2(N) \rceil + 1$ chunk times, where N is the number of peers, while no such algorithms are known for pulling systems. Practical implementations, however, often prefer a pull based mechanism, because it guarantees that no conflicts arise at the receiver. Other options include mixed Push/Pull strategies [10], and Offer/Select chunk trading [11] that can be associated to both Push and Pull strategies.

Regardless of the protocol and the signaling strategy used, the core of a scheduler is the algorithm to choose the *chunk* to be pulled or pushed and the *peer* to communicate with. During the project we have developed many algorithms and strategies that optimize the performance of the system and that are being implemented in our prototype [7–9].

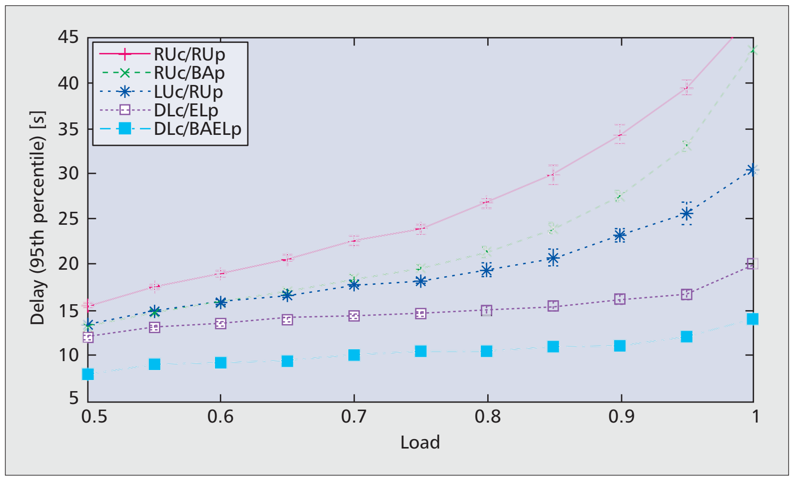


Figure 6. *Performance of several scheduling policies for immediate Push. 95th percentile of the chunk delivery delay versus load.*

For example, Fig. 6 reports the 95th percentile of the delivery delay of all chunks to all peers in an overlay topology in which each peer has 20 neighbors selected at random. 1000 peers are considered (with different upload bandwidth), chunks last 1s, so that propagation delay is negligible compared to chunk transmission time. We report the performance of five schedulers:

- **RUc/RUp**: a random chunk is sent to a random peer that needs it.
- **RUc/BAp**: a random chunk is preferentially sent to high upload bandwidth peers that need it.
- **LUc/RUp**: the youngest chunk is sent to a random peer that needs it.
- **DLc/ELp**: deadline based algorithm that diffuses the chunk with the smallest deadline (see [8] for details) to a peer that is in the best possible state to further diffuse it.
- **DLc/BAELp**: same as above, but preferentially selecting high upload bandwidth peers.

These key features must, however, be supported by measurements that in most cases are dynamic and can only be implemented in the application overlay itself. As a consequence, the Measurement Module assumes a central role in the system.

As can be seen, more advanced schedulers make it possible to reduce the chunk delivery time, with improvements up to a factor of 5 at high load. Notice that information provided by the P-REP, such as the peer upload bandwidth, leads to a significant improvement.

Immediate Pushing of chunks may have drawbacks in some scenarios (e.g., when the RTTs is not negligible compared to chunk transmission time) and explicit acceptance of the chunk transfer may guarantee better usage of resources. We have investigated signaling mechanisms where the transmitter and the receiver peers explicitly exchange signaling messages to agree on the chunk to be transferred; the scheme requires the sender to maintain a local queue of already scheduled chunks to seamlessly exploit peer upload bandwidth [11].

Conclusions and Ongoing Work

TV applications exploiting the P2P communication paradigm are gaining momentum and are already a commercial reality. Their overall architecture is, however, imprinted by file-sharing applications and they operate without any coordination with the IP network, often resulting in poor, even wasteful resource usage, to the detriment of both network and users, and preventing them from the possibility to scale to High Quality (let alone High Definition) TV.

We have presented in this article the architecture of a P2P-TV system that is being developed in the NAPA-WINE project, that is designed with the goal of efficiency and cooperation between the application and both the network operator and the content provider, aiming at optimizing resource usage and minimizing the P2P-TV application impacts on the transport network.

The key features are:

- The development of algorithms for topology management and information scheduling that, starting from network measures, minimize the usage of network resources while preserving the application performance.
- The presence of shared repositories that can be used to exchange information between the network and the application, so that the decisions taken by peers regarding connectivity, topology, signaling, and information transfer can be taken with the appropriate knowledge-base.

The overlay topology management and the chunk scheduling of information has been identified as key features for the application to be network-friendly. The first function makes it possible to build efficient and rational overlay topologies that are correctly mapped on top of the transport network structure (e.g., considering the minimal number of hops between neighbors, locality w.r.t. Autonomous Systems, etc.). The second function guarantees that the network capacity is exploited without waste (e.g., by minimizing retransmissions and pursuing an efficient distribution of chunks, etc.).

These key features must, however, be supported by measurements that in most cases are dynamic and can only be implemented in the application overlay itself. As a consequence, the Measurement Module assumes a central role in the system, and the development of efficient algorithms and methods for measurement is just as fundamental to the overall system as efficient topology management and scheduling.

The described architecture is being implemented in the NAPA-WINE project and is made publicly available as software libraries under LGPL license, freely downloadable from the project web site.

Acknowledgments

We are deeply in debt to all the people participating in NAPA-WINE and contributing to the project's success with their work and research.

References

[1] L. Jiangchuan *et al.*, "Opportunities and Challenges of Peer-to-Peer Internet Video Broadcast," *Proc. IEEE*, vol. 96, no. 1, Jan. 2008, pp. 11–24.

[2] E. Leonardi *et al.*, "Building a Cooperative P2P-TV Application over A Wise Network: the Approach of the European FP-7 STREP NAPA-WINE," *IEEE Commun. Mag.*, vol. 46, no. 20, Apr. 2008, pp. 20–22.

[3] V. Aggarwal, A. Feldmann, and C. Scheideler, "Can ISPS and P2P Users Cooperate for Improved Performance?" *SIGCOMM Comp. Commun. Rev.*, vol. 37, no. 3, July 2007, pp. 29–40.

[4] T. S. E. Ng and H. Zhang, "Predicting Internet Network Distance with Coordinates-based Approaches," *Proc. IEEE Infocom*, 2001, vol. 1, New York, NY, USA, June 2002, pp. 170–79.

[5] Y. Yardi, "Network Tomography: Estimating Source-Destination Traffic Intensities From Link Data," *J. American Statistics Association*, 1996, pp. 365–77.

[6] J.Seedorf, S. Kiesel, and M. Stiemerling, "Traffic Localization for P2P Applications: The ALTO Approach," *Proc. IEEE P2P 2009*, Seattle, WA, USA, Sept. 2009.

[7] A. Couto da Silva *et al.*, "A Bandwidth-Aware Scheduling Strategy for P2P-TV Systems," *Proc. IEEE P2P 2008*, Aachen, Germany, Sept. 2008.

[8] L. Abeni, C. Kiraly, and R. Lo Cigno, "On the Optimal Scheduling of Streaming Applications in Unstructured Meshes," *Proc. IFIP Networking 2009*, Aachen, Germany, May 11–15, 2009

[9] L. Abeni, C. Kiraly, and R. Lo Cigno, "Robust Scheduling of Video Streams in Network-Aware P2P Applications," *Proc. IEEE Int'l. Conf. Communications (ICC'09)*, Cape Town, South Africa, May 23–27, 2010.

[10] A. Russo and R. Lo Cigno, "Delay-Aware Push/Pull Protocols for Live Video Streaming in P2P Systems," *Proc. IEEE Int'l. Conf. Commun. (ICC'09)*, Cape Town, South Africa, May 23–27, 2010.

[11] R. Fortuna *et al.*, "QoE in Pull Based P2P-TV Systems: Overlay Topology Design Tradeoffs," *Proc. IEEE P2P 2010*, Delft, The Netherlands, Aug. 2010.

Biographies

Robert Birke received his Ph.D. in 2009 from Politecnico di Torino, where he is currently holding a PostDoc position. In the past years he participated in different Italian and European funded research projects and currently he is working on the European Napa-Wine project about P2P television. His research interests are in the field of network measurements and peer to peer streaming application design.

Emilio Leonardi is Associate Professor at the Dipartimento di Elettronica of Politecnico di Torino. His research interests are in the field of: performance evaluation of wireless networks, P2P systems, packet switching. He is the scientific coordinator of the European 7-th FP STREP project NAPA-WINE.

Marco Mellia [SM'08] (mellia@tlc.polito.it) received his Ph.D. degree in 2001 from Politecnico di Torino where he is now Assistant Professor. His research interests are in the fields of traffic measurement, P2P applications and green network design. He has co-authored over 150 papers published in international journals and conferences, and he

participated in the program committees of several conferences including IEEE Infocom and ACM Sigcomm. He is the Work Packages coordinator in the NAPA-WINE project.

CSABA KIRALY is research associate at Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento, Italy, where he is responsible for the coordination of NAPA-WINE activities and chief developer of the NAPA wine streamers and schedulers. He is co-author of several scientific papers.

RENATO LO CIGNO is Associate Professor at Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento, Italy. He is co-author of more than 120 scientific publications in international conferences and journals. He is Area Editor of Computer Networks (Elseveir). His interests are in design and performance evaluation of wired and wireless networking protocols and applications. He is the coordinator of WP2 in NAPA-WINE.

ARPAD BAKAY received both his M.Sc. and University Doctorate degrees from the Electrical Engineering School of the Budapest Technical University. Since then he has worked for various commercial, for-profit companies as project and products architect, and later also as executive responsible for R&D. In the meantime he also kept academic positions as researcher and lecturer at various Hungarian and foreign universities (ELTE — Hungary, Vanderbilt — USA, BTU — Hungary). He is the coordinator of WP4 in NAPA-WINE.

TIVADAR SZEMETHY obtained is Ph.D. in 2006 from Vanderbilt University, USA, where he participated in projects related to Model-Integrated Computing, Domain- Specific Modeling and Model Verification. Since 2006, he works as software R&D consultant (NETvisor Plc, Hungary) and part-time lecturer at the University of West Hungary. Fabien Mathieu was a student at the Ecole Normale Superieure, from 1998 to 2002. He completed a Ph.D. in Computer Science in 2004. He currently works as a researcher at Orange Labs. His research interests include P2P content distribution modeling and performance evaluation.

LUCA MUSCARIELLO is senior researcher at Orange Labs on performance evaluation of communication networks. His research interests include simulation and analytical modeling of wired and wireless networks. He received a M.Sc. and a Ph.D. degree in Communication Engineering from Politecnico di Torino.

SAVERIO NICCOLINI obtained his Master degree in Telecommunications and his Ph.D. in Telematics from the University of Pisa in 2000 and 2004 respectively. He is currently the technical manager of the Real-Time Communications group at NEC Laboratories Europe focusing his research interests on application-level monitoring technologies for trustworthiness applications, cost-efficient traffic management and programmable networks based on OpenFlow.

JAN SEEDORF is Research Scientist at NEC Laboratories Europe, working in the real-time communications group. He received a B.Sc. degree in 2001 and a Diploma degree in 2002 in Computer Science from the University of Hamburg, Germany. His research interests include decentralized content distribution, P2P technologies, and security.

GIUSEPPE TROPEA received his Laurea Degree (M.Sc.) in Informatics Engineering from University of Catania, Italy, in October 2002. He is involved in several European projects, both with a small enterprise (Lightcomm S.R.L.) and the Italian Consorzio Nazionale Inter-universitario per le Telecomunicazioni (CNIT). He is co-author of several research papers dealing with privacy protection in digital services and video distribution for P2P communities.

ACCEPTED FROM OPEN CALL

A Brief Survey of Radio Access Network Backhaul Evolution: Part I

Humair Raza

ABSTRACT

As radio access networks (RAN) are evolving, the accompanying backhaul technologies are also being adapted constantly to meet the required cost-performance curve. The evolution in the RAN backhaul is being triggered by the adoption of Ethernet as the physical interface and the proven advantages of IP in simplifying the network layer. This article attempts to survey some of the prevalent and upcoming backhaul technology trends based on the aforementioned evolution within RAN. Wherever possible, a critical analysis of a particular technology trend for its technical and commercial feasibility is also presented.

INTRODUCTION

All forms of radio access technologies typically require a collection of wired or wireless links to transport traffic from the radio sites to control or switching sites. These transport links collectively comprise the radio access network (RAN) backhaul. In this article we will restrict the scope of discussion to commercial cellular (mobile) RAN technologies, such as GSM and IS-95 (CDMA).

Over the course of the last 20 years, backhaul networks have evolved from a collection of simple time division multiplexed (TDM) links, supporting voice traffic in 1G and 2G mobile networks, to fairly complicated packet based architectures to accommodate new data centric applications within 3G networks. This technology evolution is coupled with a constant increase in the bandwidth requirement, attributed to the increased reliance on mobile voice and data devices. Recent market reports are projecting an almost exponential increase in the bandwidth requirements for the RAN backhaul [1, 2]. Even though one can debate the exact magnitude of the growth rate, the sequential growth trend can be easily explained through the increased carrier capacity of 3G radio technologies.

Since the capacity and complexity of the backhaul links dictate the capital and operating cost of the network, mobile operators are putting much more emphasis on optimizing the RAN backhaul. Therefore, the backhaul network design now becomes an important part of the planning process leading to the rollout of new services. An increased attention in the backhaul part of the network placed by the operators has triggered an increased investment by the vendors, leading to an array of technologies now at the disposal of the operators to match a variety of business models.

The main focus of this article is to provide a brief survey of the evolution of backhaul technologies, starting from legacy TDM digital cross-connect switches (DCS), to MPLS enabled multi-service routers (MSR). We start this survey by first introducing the generic backhaul reference architecture, followed by a brief description of various steps involved in the design of a properly dimensioned backhaul network. Various backhaul technology trends are captured. We present an overview of several open issues facing the operators going through 2G to 3G migration and a brief introduction of the impact of Long Term Evolution (LTE) on RAN backhaul design. Since some aspects of LTE standards are still being discussed, a detailed survey of RAN backhaul evolution to address LTE requirements, specifically the issues of synchronization, security, QoS and bearer path optimization, will be covered in Part II of the article, to appear in a future issue of *IEEE Communication Magazine*.

GENERIC BACKHAUL REFERENCE ARCHITECTURE

The discussion of various backhaul technology alternatives, presented later in this article, will be based on a generic representation of a typical RAN backhaul segment, as shown in Fig. 1.

The three main points of interest for the design of a RAN backhaul are the cell, hub, and mobile switching office (MSO) sites. At some or all of these sites, aggregation, switching, and routing functions are performed to optimally transport traffic between the radio access and core network. Even though the technologies involved in performing the necessary functions at these points of interest are evolving over a period of time, the network architecture remains relatively constant.

Based on the reference architecture shown in Fig. 1, a particular RAN backhaul implementation scheme can be defined in terms of available

0163-6804/11/$25.00 © 2011 IEEE

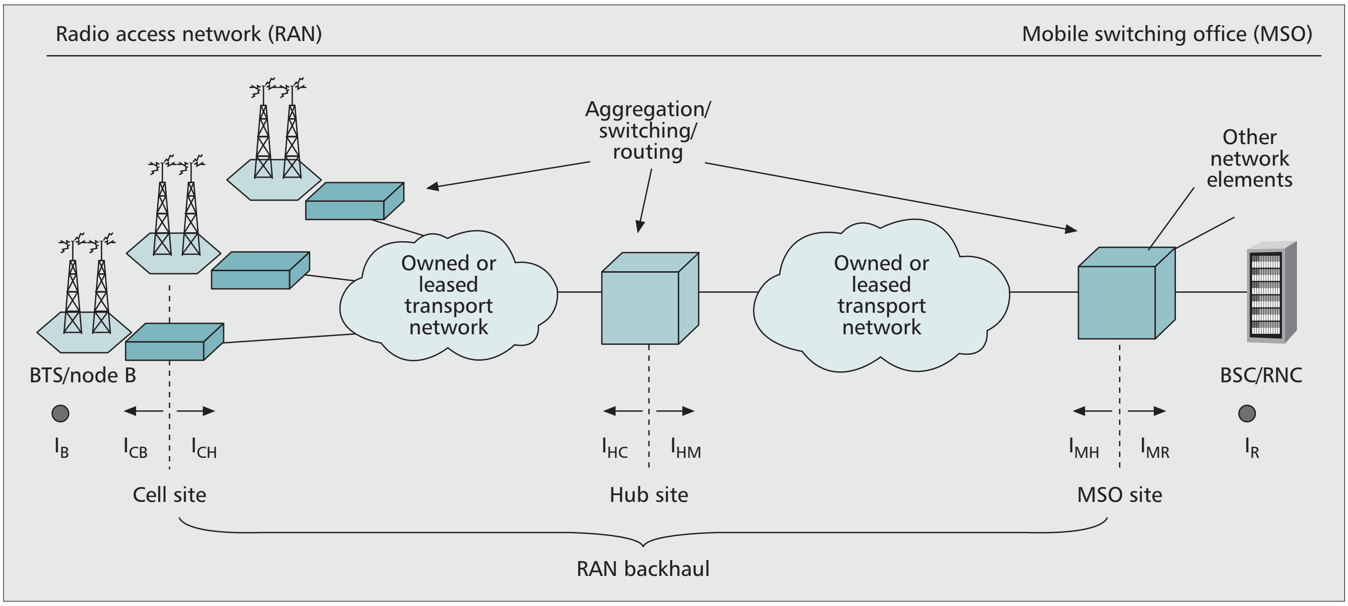


Figure 1. *Generic RAN backhaul reference architecture.*

interfaces at the cell, hub, and MSO site devices, denoted as (I_{CB}, I_{CH}), (I_{HC}, I_{HM}), (I_{MH}, I_{MR}), respectively, in Fig. 1. The selection of interfaces I_{XX} in most cases is dictated by the choice of transport network, e.g., circuit switched or packet switched. In addition, the processing performed between the ingress and egress interfaces of these devices may also depend on whether the transport network is owned or leased.

An important step involved in the design of a RAN backhaul network is the estimation of the required capacity at various interfaces along the reference architecture, which is explained in the next section.

RAN Backhaul Design Process

Typically, a RAN backhaul design process is closely related to the traffic engineering process. The traffic engineering within mobile networks involves modeling of the subscriber behavior, such as the number of voice calls per user, duration of the voice calls, data usage in number of bits per application types, to generate estimates of the peak and average traffic across the access and core network. In addition to the subscriber and application usage behavior, the traffic engineering also takes into account the RF design and processing capacity limits of various network elements, such as base transceiver station (BTS), base station controller (BSC), etc.

At the end of the traffic engineering process a collection of logical A–Z demands between a group of BTS and BSCs are estimated. The RAN backhaul design process then utilizes this set of logical demands to minimize the cost of transport while maintaining the required QoS constraints, such as delay, packet loss, and jitter, of the supported applications. The transport cost minimization is typically a joint optimization problem, where the cost of the local switching is traded off for the cost of a transport link to carry the traffic to a more optimal switching location.

A RAN backhaul design and the associated traffic engineering process are depicted in Fig. 2.

As shown in Fig. 2, the traffic engineering process is conventionally broken down into two separate categories addressing the voice and data access requirements. With the introduction of VoIP, voice will eventually become another application within the set of current data applications, such as web browsing and email access, eliminating the need for the separation shown in Fig. 2.

A brief overview of the RAN backhaul reference architecture and a design process, given so far, will help motivate the discussion of various technology trends presented in the next section.

RAN Backhaul Technologies Evolution

Traditionally, design requirements of RAN backhaul networks have been driven by the voice traffic. However, as the end-user devices and applications are evolving, emphasis is now placed on supporting mobile data services. Specifically, the backhaul networks are being adapted to satisfy stringent bandwidth requirements of high-speed packet access (HSPA) and 1x evolution-data optimized (EV-DO) enhancements to WCDMA and CDMA2000 networks. These enhancements, with the potential to increase the peak data rates by at least an order of magnitude compared to the existing rates, can significantly increase the transport cost, if backhaul networks are designed through the conventional TDM/circuit switched approach. Therefore, operators are looking for ways to reduce this cost by introducing packet level aggregation and/or switching technologies at various points between the RAN and core networks.

The evolution of cell, hub, and MSO site devices, to accommodate the stated shift in the backhaul requirements, is presented in the following sections.

As IP has evolved from its early perception as the best effort, connectionless transport to QoS aware and traffic engineered transport, the recent 3GPP standards (R6 and beyond) dictate IP as the required transport technology to scale radio access networks.

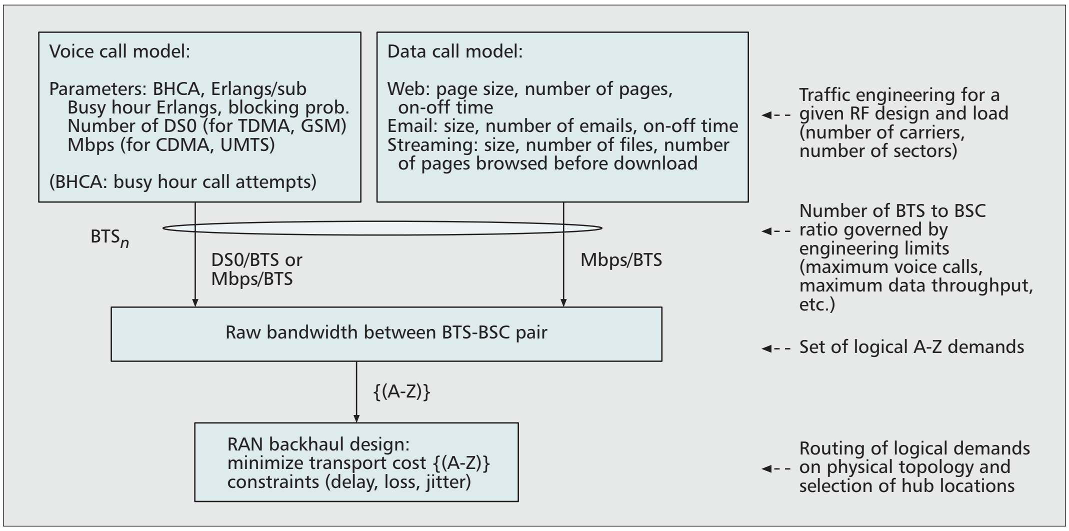


Figure 2. *RF and traffic engineering process providing the required inputs for a RAN backhaul design.*

CELL SITE DEVICE

As shown in Fig. 1, a cell site device aggregates traffic from one or multiple BTS/Node Bs at a given location. Typically for 2G networks, cell site devices were not considered essential, as the traffic that originated from each of the BTS was of the order of few Mb/s, not enough to justify a specialized grooming device in many cases. The instances where such devices existed at the cell site were primarily justified for operational reasons, such as fault isolation, remote troubleshooting, etc. However, as the traffic demand from each Node B is expected to grow many folds, traffic grooming and payload optimization functions are becoming essential, necessitating a dedicated cell site device. The requirement for managing transport links of multiple generation radio technologies is further accelerating this demand.

As the cell site device directly interfaces with the BTS/Node B, its evolution is closely related to the evolution of available BTS/Node B interfaces. Therefore, it is worthwhile to explore the trends of available interfaces at BTS/Node Bs to motivate the corresponding evolution for cell site devices. The related aspects of BTS/Node B evolution, such as physical interface, link/network layer interface, and bandwidth requirement for a typical deployment scenario, are shown in Fig. 3.

Figure 3 attempts to capture two major interface evolution trends, originating from GSM and IS-95 (CDMA) as the base 2G technologies. The standardization efforts for the evolution of GSM under 3GPP [3] mandated ATM as the link layer between Node B and radio network controller (RNC), as IP transport was not perceived to meet the stringent reliability and QoS associated with the predecessor TDM link layer available in GSM networks. However, as IP has evolved from its early perception as the best effort, connectionless transport to QoS aware and traffic engineered transport, the recent 3GPP standards (R6 and beyond) dictate IP as the required transport technology to scale radio access networks.

On the other hand, IS-95 (CDMA) standardization under TIA [4] has not specified a particular link layer technology to be used between BTS and BSC. The common implementation choices were high-level data link control (HDLC) or point-to-point protocol (PPP) based framing to transport Abis payload. Due to this non-TDM bias, the subsequent CDMA2000 standardizations under 3GPP2 [5] have quickly converged to the IP based transport, skipping the ATM layer altogether in this evolution.

It should be noted that the physical interfaces are not specified by the standards, but left as an implementation choice. Due to wide availability of T1/E1 based transport in the first mile of access networks, the most common physical interfaces on BTS/Node Bs are T1/E1 based. However, as the access networks are being overhauled everywhere to accommodate higher data rate requirements for fixed broadband, Ethernet based transport is emerging as a technology of choice even among the radio vendors. Therefore, the common perception is that once the peak data rates exceed well beyond the capacity of T1/E1 carrier, Ethernet interfaces will become widely available.

To summarize, two evolution trends can be easily drawn from the representation in Fig. 3:

ET1 — Physical interfaces evolving from T1/E1 to Ethernet.

ET2 — Network layer interfaces converging to IP.

These two evolution trends are somewhat related to two commonly used practices to choose an appropriate cell site device. The first approach is to provide minimal functions at the cell site device, complemented with the feature rich devices in front of BSC/RNC. In this case the cell site device in essence maps the physical interface of BTS to the transport technology available at the cell site. The second approach is to perform additional payload processing, grooming and switching functions at the cell site. In this case the cell site device has to terminate the link/network layer to process the payload before transporting it to BSC/RNC.

Several generic implementations of cell site devices based on these evolution trends are shown in Fig. 4. The "Present Mode" implementation as a DCS merely represents a common

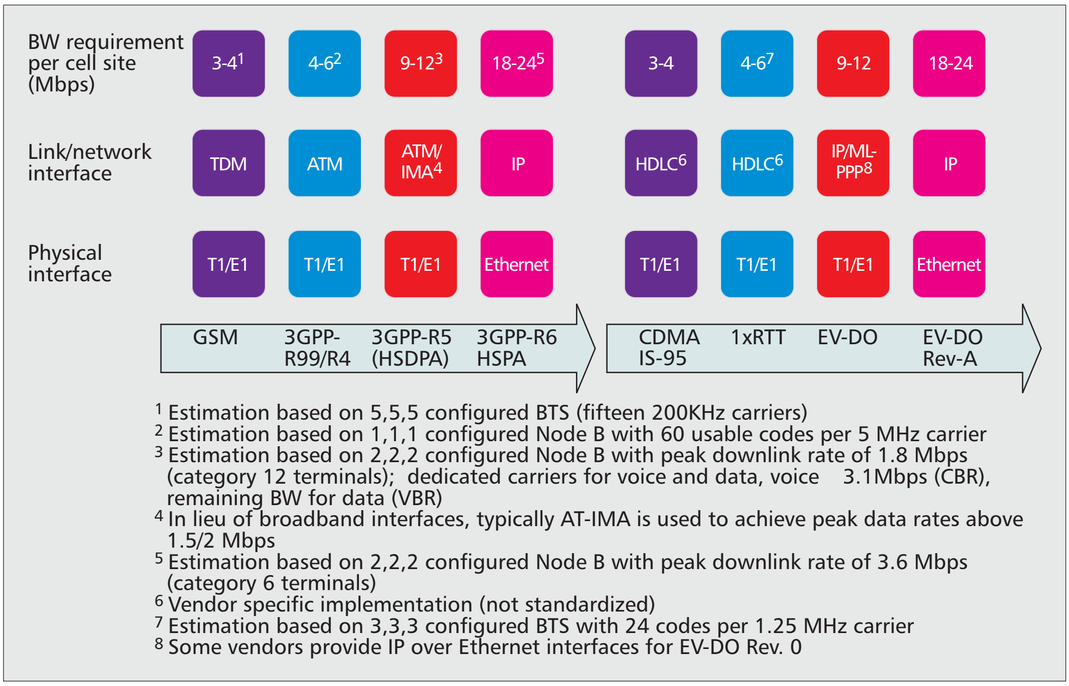


Figure 3. *Evolution of BTS/Node B interfaces.*

choice and should not be assumed to be the only available option. Based on the evolution trend ET1, the top half of Fig. 4 shows two common implementation choices: Multi-Service Provisioning Platform (MSPP)/Next-Gen DCS, and Multi-Service Switch (MSS)/MSR. Specifically, MSPP/Next-Gen DCS relies on extensions of SONET/SDH standards that deal with the transport of multi-service payloads, such as GFP [6]. On the other hand, an MSS/MSR may utilize MPLS to transport multiple payload types over Ethernet uplinks. In this context, Pseudo-wire emulation edge-to-edge (PWE3) [7], and the associated set of protocols, can be used to transport TDM or Ethernet payloads. To provide capability of multi-service transport over T1/E1 uplinks, multilink-PPP (ML-PPP) may also be required.

Based on the evolution trend ET2, the bottom half of Fig. 4 shows two implementation schemes, specifically optimized for 3GPP and 3GPP2 networks. Since in this mode of operation the cell site device is terminating the logical link or network layer between BTS and BSC, the typical implementation choices are limited to a MSS or MSR. In addition to pure multi-service transport functions, these devices also perform a number of functions labeled as "Payload Processing" in Fig. 4. For GSM EDGE RAN (GERAN), this may encompass Abis compression, whereby proprietary techniques are used to process Abis payload (between BTS and BSC) to remove idle time slots, perform silence suppression, etc. For Universal Terrestrial RAN (UTRAN), this involves processing Iub payload to terminate inverse multiplexing over ATM (IMA) links or to perform ATM adaptation layer 2 (AAL2) switching, etc. After appropriate payload processing, an optimized payload can be transported over Ethernet or T1/E1 uplinks using PWE3 or ML-PPP protocols, as described earlier.

Even though the trends shown in Fig. 4 are explained in terms of wireline transport links, the same trends and discussion are applicable when wireless backhaul links, such as point-point or point-to-multipoint microwave, are used to carry T1/E1 or Ethernet uplinks from the cell sites to hub sites.

Hub Site Device

As shown in the generic RAN backhaul reference architecture (Fig. 1), the main purpose of a hub site device is to aggregate traffic from multiple cell sites. The incoming links to a hub site device are originated either from the associated cell site devices or direct point-point wireline or wireless links from BTS/Node Bs. This intermediate level of aggregation typically reduces the overall cost of transport by combining a large number of partially filled incoming links to a few highly filled uplinks to MSO sites.

Along the lines of deployment modes for the cell site devices, the hub site devices can also be deployed to process physical or link/network layer interfaces. The ongoing evolution in the hub site devices is very similar to the evolution described for cell site devices in the previous section and will not be repeated here. The cell and hub site devices, however, differ in the required processing capacity and the achieved level of optimization. Since a hub site device typically terminates traffic from multiple cell sites, the processing capacity requirements are much higher than the cell site device. Secondly, the hub site device operates on traffic originated from a large geographic area, typically consisting of a diverse subscriber population. Therefore the statistical multiplexing gains are much higher than achieved through the individual cell site devices.

Based on the reasons quoted above, some operators may prefer to deploy hub site devices strictly for optimization reasons and the cell site

The ongoing evolution in the hub site devices is very similar to the evolution described for cell site devices. The cell and hub site devices, however, differ in the required processing capacity and the achieved level of optimization.

Native Ethernet, or even Carrier Ethernet, is not enough to provide the required QoS, traffic engineering and resiliency to transport mixed voice and data traffic within the RAN. In the absence of sophisticated traffic engineering capabilities, over-provisioning of Ethernet will be required, which may be even more expensive.

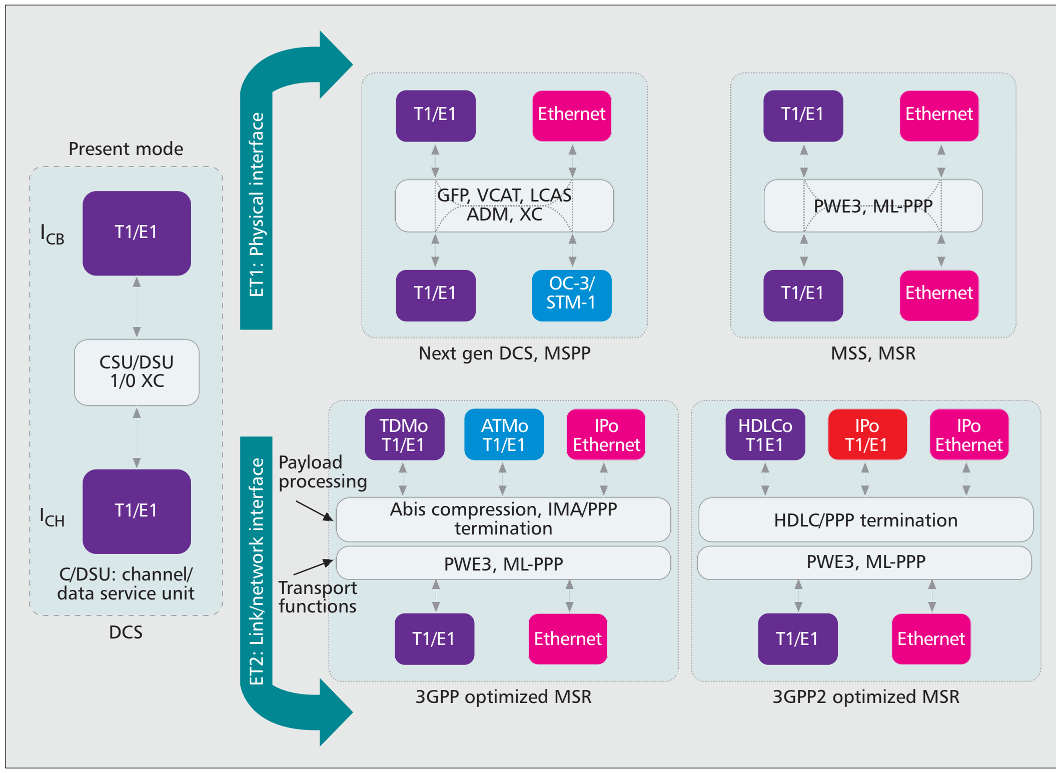


Figure 4. *Evolution of cell site devices based on physical and link/network interface requirements at BTS/Node Bs.*

devices to automate various operational and maintenance procedures discussed earlier.

MSO Site Device

The backhaul transport links originating from BTS/Node Bs terminate at one or multiple BSC/RNCs located typically at the mobile switching offices. As shown in the generic backhaul reference architecture (Fig. 1), in almost all of the cases an additional device in front of BSC/RNC is required to terminate the aforementioned RAN backhaul transport links. In this context, the functions performed by the MSO site device are:

- Mapping of the transport network interface I_{MH} to an appropriate physical interface of BSC/RNC.
- Reversal of the processing performed at the cell/hub site to recover the original Abis or I_{ub} payload from each BTS/Node B.
- For 3G traffic, aggregation of ATM or IP traffic from individual Node Bs, carried over independent VCs or PPP links, to generate a concatenated payload for RNC.

The evolution of a MSO site device is driven by similar trends as discussed earlier, i.e., the migration of interfaces from TDM to Ethernet, and the convergence of the network layer to IP. At the device implementation level, a similar trend is also observed: add drop multiplexer (ADM) and DCS deployed for present 2G/2.5G networks are being upgraded to MSPPs and MSRs for handling 3G traffic. Even though the underlying trends are similar, a slightly different implementation scheme for MSO site devices is shown in Fig. 5. The main difference is that the "evolved" MSO site device is shown in the form of two *logically or physically separated* devices to terminate 2G and 3G traffic in front of BSC and RNC, respectively. This difference can be mainly attributed to a much higher capacity requirement for a MSO site device, which is supposed to terminate traffic from hundreds of BTS/Node Bs. Hence a separate device each optimized for its respective payload is justifiable, which may not be the case for a cell site and even hub site device.

Having discussed the most common evolution trends of the RAN backhaul design, in the next section we will identify several issues that are being investigated within the vendor and operator communities for their commercial and technical feasibility.

Open Issues Related to RAN Backhaul Evolution

MPLS at the Cell Site

Readers familiar with the debate around the pros and cons of MPLS at the customer premises will be able to relate to the issues surrounding the choice of enabling MPLS at the cell site. The thought process that typically leads to justification of MPLS at the cell site is as follows:

- Bandwidth requirement at the cell site is expected to grow significantly due to adoption of mobile data applications.
- To effectively manage this projected bandwidth growth, fiber based transport services are required at the cell site.
- Native Ethernet, or even Carrier Ethernet, is not enough to provide the required QoS, traffic engineering and resiliency to transport mixed voice and data traffic within the

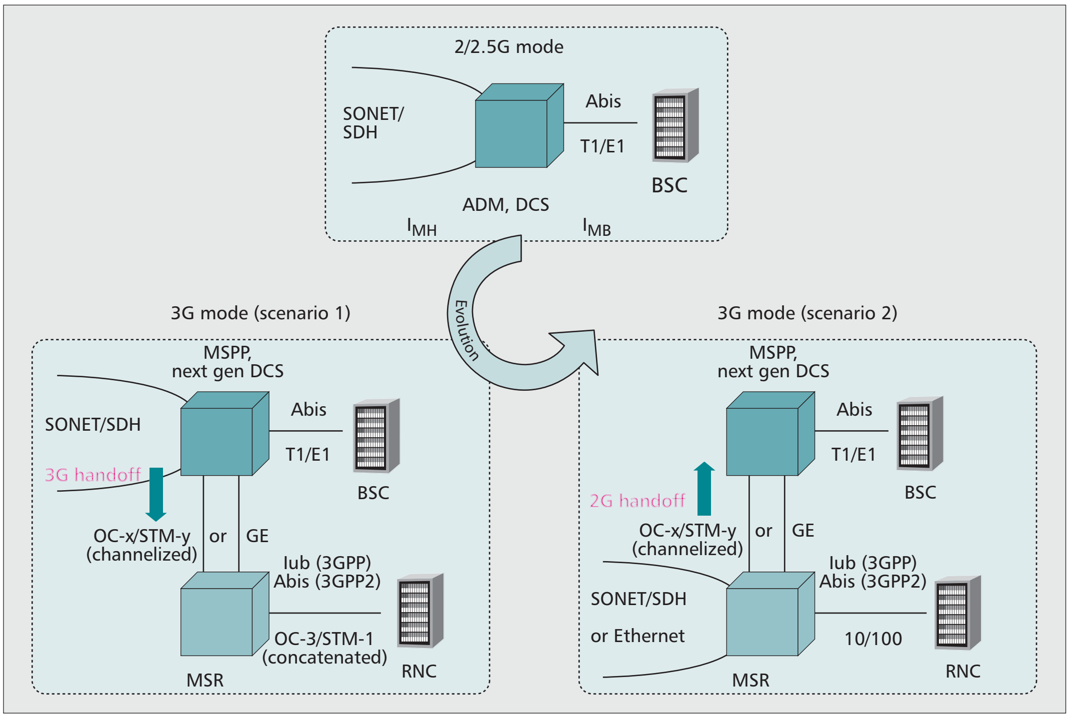


Figure 5. *Evolution at MSO site to accommodate RAN backhaul requirements for 2G and 3G traffic.*

In the case of UTRAN, MPLS provides the additional advantage of enabling migration from TDM to ATM to IP-based RAN. Therefore as MPLS has proven itself as the technology of choice for consolidating disparate wireline networks, similar advantages will drive its adoption in UTRAN backhaul.

RAN. In the absence of sophisticated traffic engineering capabilities, over-provisioning of Ethernet will be required, which may be even more expensive.

Based on the above arguments, proponents of MPLS at the cell site believe that the implementation schemes for transport of Ethernet over MPLS, such as RFC 4448 within PWE3 WG [8], provide the bandwidth scalability while maintaining the required QoS. For handling QoS, MPLS differentiated service classes, "Expedited Forwarding (EF)" and "Assured Forwarding (AFxy)," allow easy mapping and packet forwarding treatment of various bearer services such as "conversational," "streaming," and "interactive" for UTRAN as well as standardized traffic classes signaled through QoS Class Identifier (QCI) values within Evolved-UTRAN (E-UTRAN) bearer.

Furthermore, in the case of UTRAN, MPLS provides the additional advantage of enabling migration from TDM to ATM to IP-based RAN. Therefore as MPLS has proven itself as the technology of choice for consolidating disparate wireline networks, similar advantages will drive its adoption in UTRAN backhaul.

The typical arguments quoted against the idea of MPLS at the cell site are the added capital and operating expenses. The common perception is that MPLS enabled devices are more expensive than, for example, pure Ethernet devices, due to the additional hardware and/or software processing required for the associated set of protocols. Similarly, to gain the traffic engineering and resiliency benefits, the access devices require participation in full layer 3 routing topology, resulting in a non-trivial increase in operation and management expenses of these devices.

Based on the reasons quoted above, it is expected that wide scale adoption of MPLS at the cell site may take its due course, as is typical for adoption of new technologies.

RAN Backhaul Through xDSL

Another approach to deal with a projected steep growth in bandwidth requirements at the cell site is to utilize the commonly available xDSL access. The proponents of this approach argue that xDSL technology has advanced far enough to be considered a reliable alternative to the approach of extending fiber to the cell site, discussed in a previous section. The main idea behind this approach is to partition the traffic from Node B into two different streams, a "voice only" stream carrying circuit switched voice and a "data only" stream carrying traffic generated through typical data applications. Since the data applications can withstand higher latency and packet loss, due to inherent mechanisms within these applications, "data offload" through xDSL is perceived as a logical choice to deal with the anticipated growth in the bandwidth requirements.

A generic implementation of RAN backhaul through xDSL is shown in Fig. 6. The obvious advantages associated with this approach are:

Cost: In most areas, the cost of xDSL access at the cell site is much cheaper than the conventional T1/E1 or even Ethernet private line services.

Speed of deployment: Since xDSL access is available through the existing copper plant, backhaul links can be provisioned without having to wait for the availability of fiber at the cell site.

The arguments that are typically used to oppose this approach are:

Higher blocking probability leading to increased latency and/or packet loss: To optimally utilize the available spectrum, the common traffic engineering practice is to share the available codes for HSPA for a large number of

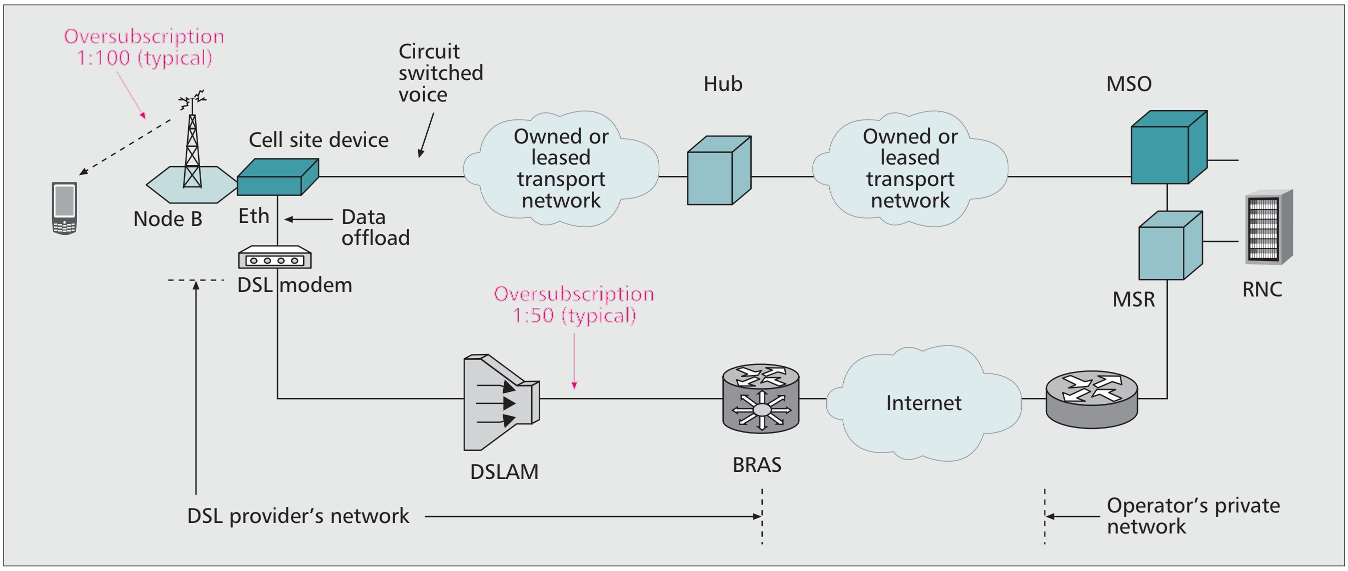


Figure 6. *A generic implementation scheme of RAN backhaul through xDSL links.*

active users. Therefore, during peak utilization the airlink oversubscription ratios of 1:100 are quite likely. However, if the same highly oversubscribed link is backhauled through xDSL, an additional level of blocking, between DSL access multiplexer (DSLAM) and broadband remote access server (BRAS) links, is introduced. Even in a liberally engineered xDSL network, an oversubscription ratio of 1:50 is typical. Therefore in the worst case, backhaul through xDSL can experience very high blocking probability during peak hours, leading to excessive delay and/or packet loss, which is not desirable.

Lack of proper fault management tools: Since xDSL networks are typically not designed with high availability in mind, the fault detection and isolation tools are fairly trivial, especially when compared to those used for the leased lines or private networks. To make matters worse, these links have to traverse the Internet, where the peer-peer service level agreements (SLAs) are typically best effort with little or no standardized tools to guarantee inter-provider SLAs.

Bandwidth and/or reach limitations: Most of the currently deployed xDSL technologies do not support the projected bandwidth requirements from each cell site, as shown in Fig. 3. Even though some of the upcoming xDSL technologies may support higher rates, the reach may not be enough to guarantee consistent coverage within a RAN segment.

Based on the pros and cons quoted above, it is obvious that even though there are economic advantages to providing RAN backhaul through xDSL, a wide scale deployment is quite unlikely until the specified technical challenges are resolved.

In addition to its use in macro-cellular RAN backhaul, DSL technologies will also be used to provide backhaul from Home-NodeBs (HNB), also known as femtocells. Since some aspects of HNB architecture are still being discussed in 3GPP and other standards forums, a detailed discussion of DSL backhaul for femtocells will be presented in Part II of this article.

LTE: Flat IP Architecture

The current standardization within 3GPP for R7 architecture is being handled by two working groups: LTE, focused on specification of E-UTRAN [9]; and System Architecture Evolution (SAE), mandated to evolve 3GPP packet core to scale with the requirements of E-UTRAN [10]. Specifically, the requirements imposed on E-UTRAN are to provide downlink peak data rates of 100 Mb/s within a 20 MHz band, and the user-plane latency of less than 5 ms. Similarly, the SAE proposal is for an all-IP packet switched (PS) domain to support both voice and data services, as well as interfaces to multiple radio access technologies (RATs), such as GERAN, UTRAN, and WLAN.

From the backhaul perspective, LTE is a significant departure from the conventional GERAN and UTRAN. The main differences are the requirements for full logical mesh connectivity between enhanced-Node Bs (eNBs) (X2 interface) and direct links from each eNB to at least two mobility management entity (MME)/user plane entity (UPE) (S1 interface). The separation of X2 and S1 interfaces allows an opportunity to optimally route user and control plane traffic based on a specific mobility scenario. For example, to handle inter-eNB handoffs, user plane traffic may traverse a X2 interface with the corresponding mobility management (control plane) messages being intercepted by MME across a S1 interface. On the other hand, to handle mobility between two distinct MME domains, the user plane as well as the control plane traffic may traverse a S1 interface.

To highlight the differences between the conventional and LTE backhaul architectures, Fig. 7 attempts to model two sub-segments of a RAN, each with a group of Node Bs interconnected through appropriate network elements dictated by a specific RAN architecture. In the case of UTRAN (Fig. 7a) the backhaul network model resembles a hub-spoke logical graph, with RNC as a hub and Node Bs as spokes. The two RNCs are then interconnected through their respective serving GPRS support nodes (SGSNs), representing a simplified PS domain. The same network migrated over to E-UTRAN architecture is shown in Fig. 7b. By comparing Figs. 7a and 7b, it is quite easy to identify the main differences between the backhaul requirements of these architectures.

There are multiple methods to design a backhaul scheme that can accommodate the requirements imposed by LTE/SAE and at the same time is scalable enough to reduce the cost of implementation. One such scheme, based on IP-virtual private networks (IP-VPNs) implemented through MSRs, is shown in Fig. 7c. Some of the obvious benefits associated with this scheme are:

Reduction in complexity: Changes in the network topology and associated connectivity requires $O(n)$ operations, as opposed to $O(n^2)$ if IP-VPNs are not used.

Optimal use of resources: Achieved by sharing transport links while maintaining full virtual separation of X2 and S1 interfaces.

In addition to solving scalability issues, VPNs also provide network domain security, especially in cases where leased backhaul on shared infrastructure is used. The need for network domain security arises due to open IP architecture adopted within a LTE/SAE framework. It is assumed that migration to IP based protocols and interfaces will result in increased probability of security threats from within and outside mobile networks. Some of these security related issues and their mitigation within backhaul networks will be addressed in Part II of this article.

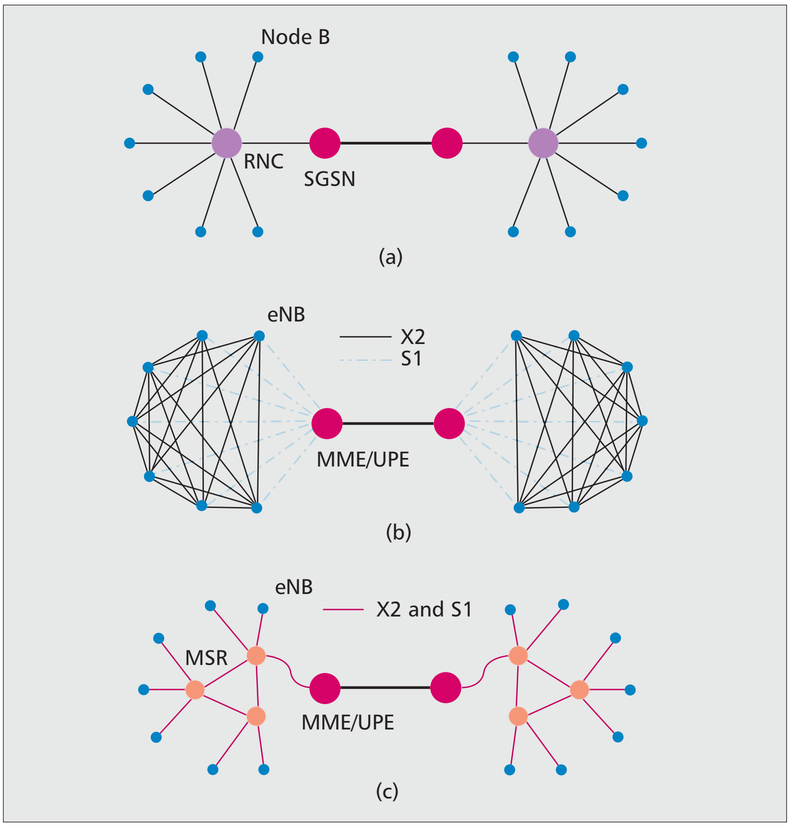


Figure 7. *a) Logical graph for UTRAN backhaul; b) logical graph for E-UTRAN backhaul, Note: For clarity sake, only one set of S1 links is shown; and c) implementation scheme for LTE backhaul through MSR enabled IP-VPNs.*

Conclusions

Due to the spatial nature of cellular (mobile) radio access networks, some form of wireline or wireless backhaul will always be needed. As mobile networks are migrating from a "voice/circuit-switched" architecture toward a "data/packet-switched" architecture, the adoption of Ethernet as the physical interface and IP as the network layer is accelerating. Since mobile networks are seldom deployed in a pure Greenfield scenario, the backhaul technologies have to often deal with a range of physical and network layer interfaces. In this article we have surveyed the evolution of RAN backhaul technologies to accommodate hybrid (TDM and packet) radio interfaces. We have presented several implementation options for the cell, hub, and MSO site devices that rely on multi-service extensions of SONET or MPLS standards. We have also identified some open issues that are currently being debated:

- How far should MPLS be extended?
- Will xDSL technologies satisfy the RAN backhaul requirements?
- Should LTE requirements be imposed on the 3G backhaul architecture?

The confidence gained through commercial deployments will shortly settle these debates.

Acknowledgment

The author would like to recognize Ed Evangelista for his feedback and support to enhance the concepts presented in this article.

References

[1] "Wireless Backhaul: Bandwidth Explosion and Emerging Alternatives," *ABI Research*, June 2005.

[2] "Mobile Backhaul Equipment, Installed Base and Services Market Outlook," *Infonetics Research*, Oct. 2006.

[3] "3rd Generation Partnership Project," http://www.3gpp.org/about/about.htm.

[4] "Mobile Station — Base Station Compatibility Standard for Dual Mode Wideband Spread Spectrum Cellular System," TIA/EIA/IS-95-A, May 1995.

[5] "3rd Generation Partnership Project 2," http://www.3gpp2.org/Public_html/Misc/AboutHome.cfm

[6] "Generic Framing Procedure," ITU-T Recommendation G.7041/Y.1303, Feb. 2003.

[7] "PWE3 Architecture," IETF RFC-3985, Mar. 2005.

[8] "Encapsulation Methods for Transport of Ethernet over MPLS Networks," IETF RFC 4448, Apr. 2006.

[9] "Feasibility Study for Evolved Universal Terrestrial Radio Access and UTRAN; Release 7," 3GPP TR25.912, Sept. 2006.

[10] "3GPP System Architecture Evolution: Report on Technical Options and Conclusions (Release 7)," 3GPP TR23.882, Sept. 2006.

Biographies

HUMAIR RAZA (humair.raza@tellabs.com) received MS and Ph.D. degrees in Electrical Engineering from the University of Southern California in 1994 and 1997, respectively. From 1997 to 2002 he held various positions in GSM and Optical Networking Groups at Nortel Networks. From 2003 to 2005 he consulted as a system architect for several early stage startups focusing on disruptive networking technologies. Since 2005 he has been working with Tellabs as a member of the wireless technology and strategy group. His research interests include design and performance analysis of spatially distributed networks.

ACCEPTED FROM OPEN CALL

A Roadmap from UMTS Optimization to LTE Self-Optimization

Oriol Sallent, Jordi Pérez-Romero, Juan Sánchez-González, and Ramon Agustí, Universitat Politècnica de Catalunya (UPC)

Miguel Ángel Díaz-Guerra, David Henche, and Daniel Paul, Telefónica

ABSTRACT

Self-Organizing Networks (SON) are currently being introduced by the 3rd Generation Partnership Project (3GPP) as part of the Long Term Evolution (LTE) system as a key driver for improving the operation of wireless networks. Given that many challenges have been identified when moving from the SON concept to practical implementation, this article claims the suitability to gain insight into the problem by taking advantage of optimization mechanisms in live Universal Mobile Telecommunications System (UMTS) networks. From that perspective, this article outlines a roadmap from actual manual optimization toward the inclusion of SON concepts in future networks. In particular, an optimization framework is developed considering different stages that include the collection of inputs from different sources (network counters, measurement reports, drive tests), the tuning parameters to achieve optimization, and the optimization procedure itself. Even though the proposed framework is presented and evaluated for convenience in a UMTS context, it is rather generic and technology-agnostic and, therefore, it can be formulated as an initial reference for LTE self-optimization use cases. The proposed framework is applied to a practical coverage optimization use case supported by data extracted from a real UMTS network in a European city, illustrating the capability to automatically identify a cell with sub-optimal coverage and to provide a solution to the problem. The extension of the case study to LTE is also analyzed. Finally, as a result of the lessons learned, the article makes a projection to the LTE context by identifying the key points to be solved for the materialization of a self-optimization procedure.

INTRODUCTION

Third-generation (3G) mobile communication systems have seen widespread deployment around the world. The uptake in High Speed Packet Access (HSPA) subscriptions, numbering well over 85 million at the start of 2009 [1], indicates that the global thirst for mobile data has just begun: broadband subscriptions are expected to reach 3.4 billion by 2014, and about 80 percent of these consumers will use mobile broadband.

Operators are doing business in an increasingly competitive environment, competing not only with other operators, but also with new players and new business models. In this context, any vision for technology evolution should encompass an overall cost-efficient end-to-end low latency proposition. By the end of 2004, 3GPP started discussions on 3G evolution toward a mobile communication system that could take the telecom industry into the 2020s. Since then, the 3GPP has been working on two approaches for 3G evolution: LTE and HSPA+. The former is being designed without backward compatibility constraints, whereas the latter will be able to add new technical features preserving Wideband Code Division Multiple Access (WCDMA) and HSPA investments.

Besides the evolution in radio technologies and network architectures, the conception of how these networks should be operated and managed is evolving as well. The envisaged high density of small sites, e.g., with the introduction of femto-cells, and the pressure to reduce costs, clearly indicate that deploying and running networks needs to be more cost-effective. SON, aiming to configure and optimize the network automatically, is seen as one of the promising areas for an operator to save operational expenditures. Indeed, it can be stated that SON will be required to make the business case for LTE more attractive. Consequently, SON has received much attention in recent years. For example, in the framework of the Next Generation Mobile Networks (NGMN) Alliance, operator use cases have been formulated for different stages [2]: planning (e.g., eNode-B power settings), deployment (e.g., transport parameter set-up), optimization (e.g., radio parameter optimization that can be capacity, coverage or performance driven), and maintenance (e.g., software upgrade). Similarly, the ICT-FP7 SOCRATES project defined a detailed use case taxonomy in [3]. 3GPP has also acknowledged the importance of SON [4]. Progress is being reflected e.g., in [5], where use cases are being developed together with required functionalities, evaluation criteria and expected results, impacted specifications and interfaces, etc. Similarly, the ICT-FP7 SOCRATES project is developing in deeper detail different use cases, such as cell outage [6]. Paradigms related to SON are discussed in an early paper in this area [7]. Additionally, in [8] the self-configuration of newly added eNode-Bs in LTE networks is discussed, and a self-optimization load balancing mechanism is proposed as well.

Due to its ambitious objectives, the practical realization of SON is seen as challenging, and it can be anticipated that SON will continue as a hot research topic in coming years requiring further research efforts. Some of the aspects that must be properly covered include: to find an appropriate trade-off between the performance gains and the implementation complexity (e.g., required signaling and measurements, computing resources), to ensure the convergence to a stable solution within a given timing requirement or to ensure the robustness to missing, wrong or corrupted input measurements.

The radical changes needed from current deployed networks such as Second-generation (2G) and 3G HSPA, mainly managed by centralized remote operations and maintenance (O&M) applications with intensive human intervention, to future SONs (LTE, HSPA+) undoubtedly require the definition of a clear roadmap when moving from SON's concepts to their practical implementation. The impact will be on all aspects of an operator's radio engineering department (operational procedures, O&M software tools, radio engineer's skills, etc.). In this respect, early actions enabling a smooth introduction of the "SON culture" may be highly beneficial on a long-term perspective. This article outlines a roadmap from actual manual UMTS optimization toward the futuristic objective of a self-optimized LTE.

In particular, this article first develops a UMTS optimization framework, discussing from a practical perspective the inputs available for an automatic optimization process as well as the optimization procedure itself. Even though the proposed optimization framework is presented for convenience in a UMTS context, it is rather generic and technology-agnostic and, therefore, it can be formulated as the initial reference for LTE self-optimization use cases. Second, this article formulates a UMTS coverage optimization case study, supported by data extracted from a real UMTS network in a large European city. The main objective of the proposed case study is to gain insight into the problem from a practical point of view, in order to identify and draw conclusions on the key points to be solved for the materialization of a full self-optimization procedure. Given the differences in radio access technology, we also analyze how the coverage optimization use case could be extended to LTE. In turn, we collect the lessons learned from the UMTS case study and project them into the LTE context. This is accompanied by the up-to-date status of SON in research and standardization fora, so that the motivation of this article to contribute to on-going efforts from a complementary perspective is enforced. Finally, we summarize the conclusions reached.

UMTS Optimization Framework

UMTS network optimization turns out to be a tricky and complex task because the target objectives in terms of coverage, capacity, and quality tend to be contradictory (e.g., capacity may be increased at the expense of coverage or quality reduction) [9]. Additionally, the number of tunable parameters in a WCDMA network is significantly higher than that of a 2G Time Division Multiple Access (TDMA)-based network. In such complex scenarios, efficient network management is needed in order to guarantee the QoS requirements to the subscribed users. In recent years, an intensive effort has been made in the field of automated optimization [10–14]. Two different phases can be distinguished in the optimization of a 3G system [15]. The first phase is RF optimization, whose objective is to guarantee the required coverage, avoiding excessive pilot pollution, cell overlap or cell overshoot by optimizing the setting of RF parameters such as pilot power, antenna down-tilt, etc. The second phase is service parameter optimization. This includes the setting of admission and congestion control thresholds, maximum downlink power per connection, events to change to compressed mode, channel switching, etc.

Inputs for the Optimization Process

In order to perform the radio optimization process, the actual network status and performance needs to be captured. Clearly, the more accurate and complete this picture is the more efficient the optimization process can be expected. In practice, the network operator can collect diverse information about the network status from different sources:

Network counters: These are measurements taken at Node-Bs and/or Radio Network Controllers (RNC), such as load levels, number of users in soft/softer handover, etc. that are transferred to the O&M module (e.g., transmitted power level can be recorded at the Node-B and forwarded to the O&M module every 15 minutes in the form of average values or statistical distributions). Typically, the manufacturer provides the operator with the software tool to manage the collection, processing (e.g., measurement averaging), and transference from network elements to O&M. The type of measurements that a given network element must be able to perform is usually defined in the standards [16–18].

Measurement reports: These are measurements taken by the mobile terminals while in an active connection and transmitted live to the RNCs (e.g., received power of the pilot channel, pilot channel energy per chip over total received power density for the serving and neighboring cells, etc.). The primary usage of these reports is as input for Radio Resource Management (RRM) algorithms, such as a handover algorithm. Nevertheless, these measurements can also be used as an input for the optimization process. In such a case, measurement reports need to be managed by a software tool, able to store and forward them to the O&M.

Drive tests: These are measurements carried out by one or several specialized terminals able

In order to perform the radio optimization process, the actual network status and performance needs to be captured. Clearly, the more accurate and complete this picture is the more efficient the optimization process can be expected.

Drive test measurements mainly illustrate the status of the downlink for the particular trajectory followed by the testing vehicle. Therefore, it is most important to specify such trajectory so that it ensures the acquisition of measurements in the areas where the real traffic is generated.

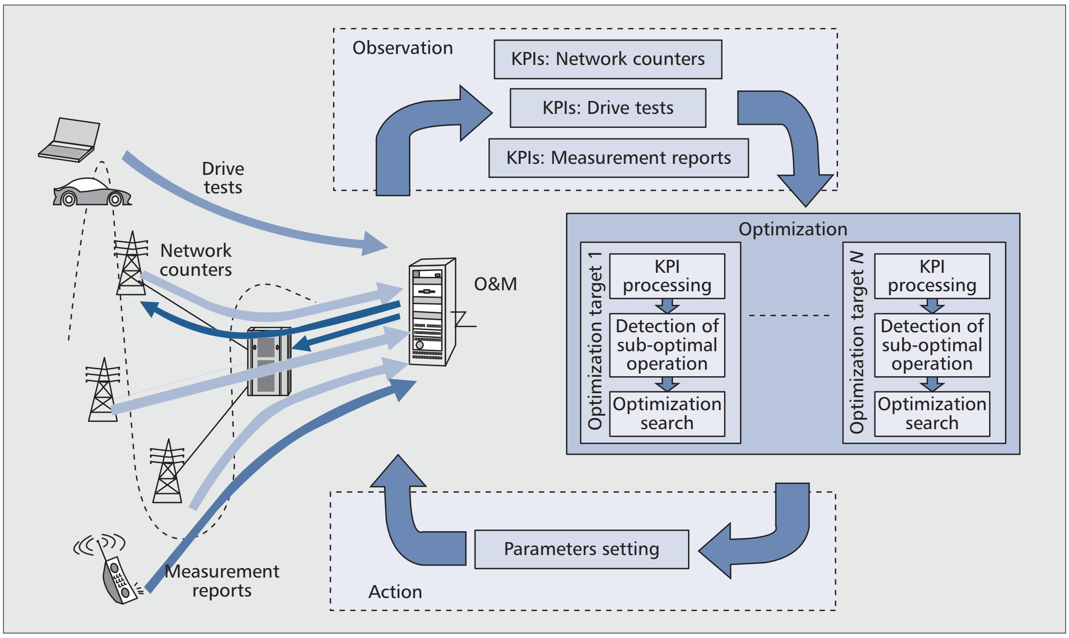


Figure 1. *Network optimization loop.*

to record a number of parameters while following a trajectory in the field. They are equipped with a Global Positioning System (GPS), so that the position where each measurement has been taken is also recorded. For example, when some misbehavior of the network is identified, radio optimization engineers can trigger a drive test over the affected area in order to obtain further information about the operating conditions. Nevertheless, drive tests can also be regularly performed as a basis of network quality control and as another input to the network optimization process. A drive test performed with a frequency scanner can also be very useful because it is able to capture true coverage predictions without the influence of the specific network configuration. This can be of particular interest in, for example, a 3G/2G heterogeneous scenario where quality problems in 3G could be automatically overcome by diverting traffic to the 2G network, thus hiding the 3G problems.

It is worth noting that network counters are able to capture mainly the behavior of the uplink, whereas measurement reports are mainly able to collect the status of the downlink. In turn, drive tests acquire similar information as measurement reports, with the additional advantage that the precise position where each measurement is taken is recorded. Drive test measurements mainly illustrate the status of the downlink for the particular trajectory followed by the testing vehicle. Therefore, it is most important to specify such trajectory so that it ensures the acquisition of measurements in the areas where the real traffic is generated.

Optimization Procedure

The overall network optimization procedure is illustrated in Fig. 1 and Fig. 2, providing the most generic view where possible inputs coming from network counters, measurement reports, and drive test are considered. The procedure acts as a loop that continuously interacts with the real network based on *observation* and *actions*.

At the observation stage, the optimization procedure collects information from the different available Key Performance Indicators (KPIs). Then the optimization procedure will contain different processes for each of the optimization targets specified based on operator policies (e.g., avoidance of coverage holes, interference reduction, reduction of overlapping between cells, etc.). For each optimization target a three-step process will be executed, consisting of:

- Processing the KPIs coming from the observation phase (e.g., combining several KPIs, obtaining statistical measurements such as averages or percentiles, etc.).
- Detecting the sub-optimal operation, i.e., that the optimization target is not properly fulfilled, based on some criteria dependent on the specific target.
- Carrying out an optimization search to find the proper parameter setting to solve the sub-optimal operation.

The final result of the optimization will be the adequate configuration of selected network parameters (e.g., antenna downtilts, pilot powers, RRM parameters configuration, etc.) affecting either the cell where the sub-optimal operation has been found and/or its neighboring cells. This change in the configuration parameter(s) turns out to be the *action* executed over the network.

The optimization process is formulated in the form of hypotheses tests against the sub-optimal operation for a number of established targets. Hypotheses are reinforced by a likelihood index that is increased every time a given condition evaluated over a certain KPI (or combination of KPIs) is met. Considering that a total of N_c conditions $c(i)$ $i = 1,\ldots N_c$ are evaluated, each condition will have an associated weight α_i depending on how relevant the condition is for the optimization target under consideration.

With the degrees of freedom provided by the weight factors α_i, the value of the likelihood index F is then simply defined as a linear combination of the evaluated conditions:

$$F = \frac{\sum_{i=1}^{N_c} \alpha_i U(c(i))}{\sum_{i=1}^{N_c} \alpha_i} \quad (1)$$

where $U(c(i))$ takes the value 1 if condition $c(i)$ is fulfilled, and 0 otherwise. As a result, F is a relative value between 0 and 1. The higher the F value is, the more likely that the target is not sufficiently optimized. The likelihood index F can be defined for every optimization target according to the specific conditions established, and the weight associated with every condition can be fixed according to the operator's strategy and policies.

Given that not necessarily all the required KPIs to detect a given sub-optimal operation may be available (e.g., measurement reports from the mobile may not be available at O&M, drive tests may not have been run for a given area, etc.), a reliability index R in the detection should be defined in accordance with how many conditions are actually evaluated with respect to the total number of potential conditions N_{tot} that would be evaluated if all KPIs were available:

$$R = \frac{\sum_{i=1}^{N_c} \alpha_i}{\sum_{i=1}^{N_{tot}} \alpha_i} \quad (2)$$

As a final result for the detection of suboptimal operation with respect to a given target, the cell under study will be categorized as one of the following:
- If $F < F_{\min}$, it is likely that the target is optimized. Reliability of this hypothesis will be associated with the obtained value of R.
- If $F \geq F_{\min}$, the most likely hypothesis is that the target is not sufficiently optimized. In this case:
  - –If the reliability index $R \geq R_{\min}$, then the optimization search stage is triggered. In this case, if drive test measurements are available, it may be possible to identify specific geographical areas (denoted here as clusters) where the sub-optimal behavior is detected. This is done by grouping the drive test positions exhibiting the particular sub-optimal behavior based on their geographical distance.
  - –If the reliability index $R < R_{\min}$, then it is decided that there is not sufficient information about the network status to perform an automated optimization search procedure. In this case, an alarm will be triggered indicating that human intervention is required for the optimization of this target in this cell.

It is worth noting that the thresholds $F_{\min}$ and $R_{\min}$ will be set according to the operator's strategy and policies. A high value for $F_{\min}$ will tend to trigger optimization search procedures only for cases where the likelihood that sub-optimal operation exists is high. A high value for $R_{\min}$ will indicate that the operator relies on the automatic operation when many inputs from the network status are available; otherwise the operator prefers analyzing the situation with the assistance of a radio engineer expert.

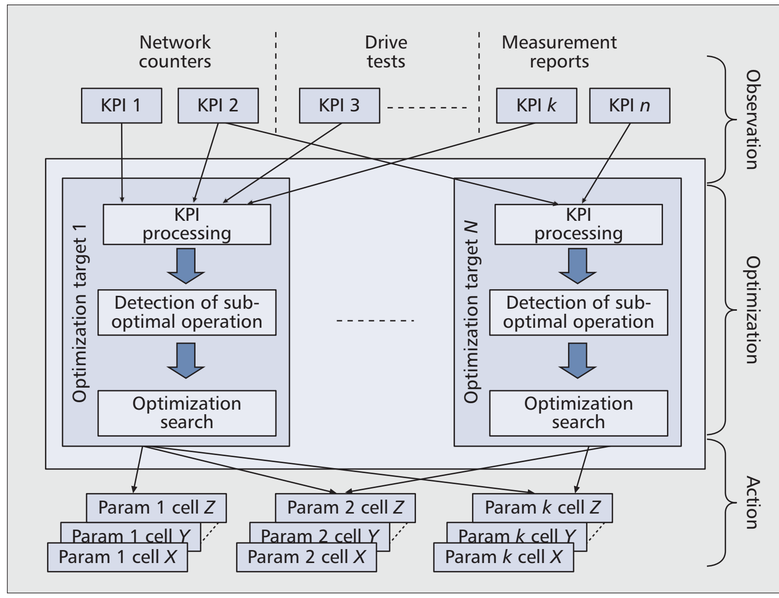


Figure 2. *Processes in the network optimization algorithm.*

In case the suboptimal operation has been detected, the role of the optimization search is to propose certain changes in the configuration parameters of one or several cells (e.g., the cell where the suboptimal operation was detected and/or its neighbors) in order to provide an optimized operation. Each target will have a number of associated parameters, which are related to the resulting performance for such target. Since a given parameter may be related to several optimization targets fixed by the operator and may also have a direct influence on other parameters, the choice of parameter and its new value must be carefully studied, requiring some degree of coordination between the optimization searches associated with different targets.

Given that parameter change on the live network is a very sensitive issue, optimization algorithms will usually require the support of a module devoted to estimate the expected impact of a given parameter change (e.g., a planning tool, a system level simulator, a software tool fed with real data measurements, etc.) prior to the modification in the live network. In practice, since the automatization of this stage may involve the integration of very diverse software tools, some human intervention may be required in the meantime.

The optimization search strategy can be based on different approaches depending on the specific target and the parameter(s) to optimize. Classical optimization methods, such as e.g., genetic algorithms, linear programming, particle swarm algorithms, etc., or other heuristic approaches, can be considered. Such methods could be adequately extended to address the problem from the multi-objective optimization perspective, jointly taking into account the dif-

Considered KPIs	Criteria/Condition	Source
RSCP (Received Signal Code Power from the CPICH channel)	$Prob(RSCP<RSCP^*)>Th_{RSCP}(\%)$	Drive Test/Measurement Report
Uplink transmission power P_T	$Prob(P_T>P_T^*)>Th_{PT}(\%)$	Drive Test/Measurement Report
Active Set and Monitored Set	The Active Set is not full ($AS < AS_{max}$) and there exists a cell in the Monitored Set that should be in Active Set.	Drive Test/Measurement Report
Number of handover attempts to GSM relative to cell load	$\frac{NumberHO\_GSM}{Cell\_throughput} > Th_{GSM}$	Network counters
Uplink RSSI (Received Signal Strength Indicator at the Node-B)	$Prob(RSSI<RSSI^*)>Th_{RSSI}(\%)$	Network counters
Statistical distribution of the propagation delay between mobiles and base station.	$Prob(Propdelay<P^*)>Th_{propdelay}(\%)$	Network counters
Network configuration parameters/Tunable parameters		
Positions of the different nodes B in the network.		
Antenna azimuth of the different cells.		
Neighbor lists.		
CPICH transmission power in the different cells.		
Antenna downtilt of the different cells		

Table 1. *Considered KPIs and tunable parameters.*

ferent optimization targets [19, 20]. Furthermore, learning mechanisms can also be applied at this stage.

Case Study: UMTS Coverage Optimization

In order to illustrate the framework presented in the previous section, the procedure for the optimization of UMTS coverage is presented. The considered scenario is shown in Fig. 6. It consists of six tri-sectorial cells in an urban area of a major European city. Each cell is identified as *Cell_x_y* where *x* corresponds to the Node-B identifier and *y* is the sector of this Node-B.

Table 1 shows the KPIs that can be useful for the optimization of the UMTS coverage in a given cell, the corresponding conditions to detect suboptimal behavior, and the source from which each KPI can be obtained. As shown, these conditions are based on statistics related to power measurements, handovers, or propagation delay measurements. Some hints about the rationale to consider these specific KPIs are given in the following:

• In general, poor coverage will be associated with low values of Received Signal Code Power (RSCP) of the Common Pilot Channel (CPICH) and Received Signal Strength Indicator (RSSI) at the Node-B.

• High uplink transmitted power may indicate excessive propagation losses and, consequently, poor coverage.

• A high ratio of handovers from UMTS to Global System for Mobile communications (GSM) cells may indicate UMTS coverage problems, provided that the applied network selection strategy is such that it favors continuity over UMTS once attached to this network. Besides, coverage problems may prevent the Active Set (AS) and Monitored Set (MS) lists from being adequately updated (e.g., due to problems in the signaling procedures for adding/removing cells from the Active Set, derived from the fact that some messages can be lost due to coverage problems).

• Finally, the statistical distribution of the propagation delay of specific signals transmitted from mobiles to cell is considered to be a valuable input because it reflects the statistical distribution of mobiles-to-cell distances when connected to the cell. More specifically, in case that users tend to connect only at distances much below the planned cell coverage radius, this might reflect suboptimal coverage at the cell edge.

Table 1 also shows the configuration parameters that are needed in the suboptimal operation detection process, which in turn will usually be the tunable parameters to be considered in the optimization search stage.

The algorithm for the detection of sub-optimal coverage operation is presented in Fig. 3, making use of the conditions presented in Table 1. It is worth noting that the RSCP condition is evaluated not only for the cell under study but

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

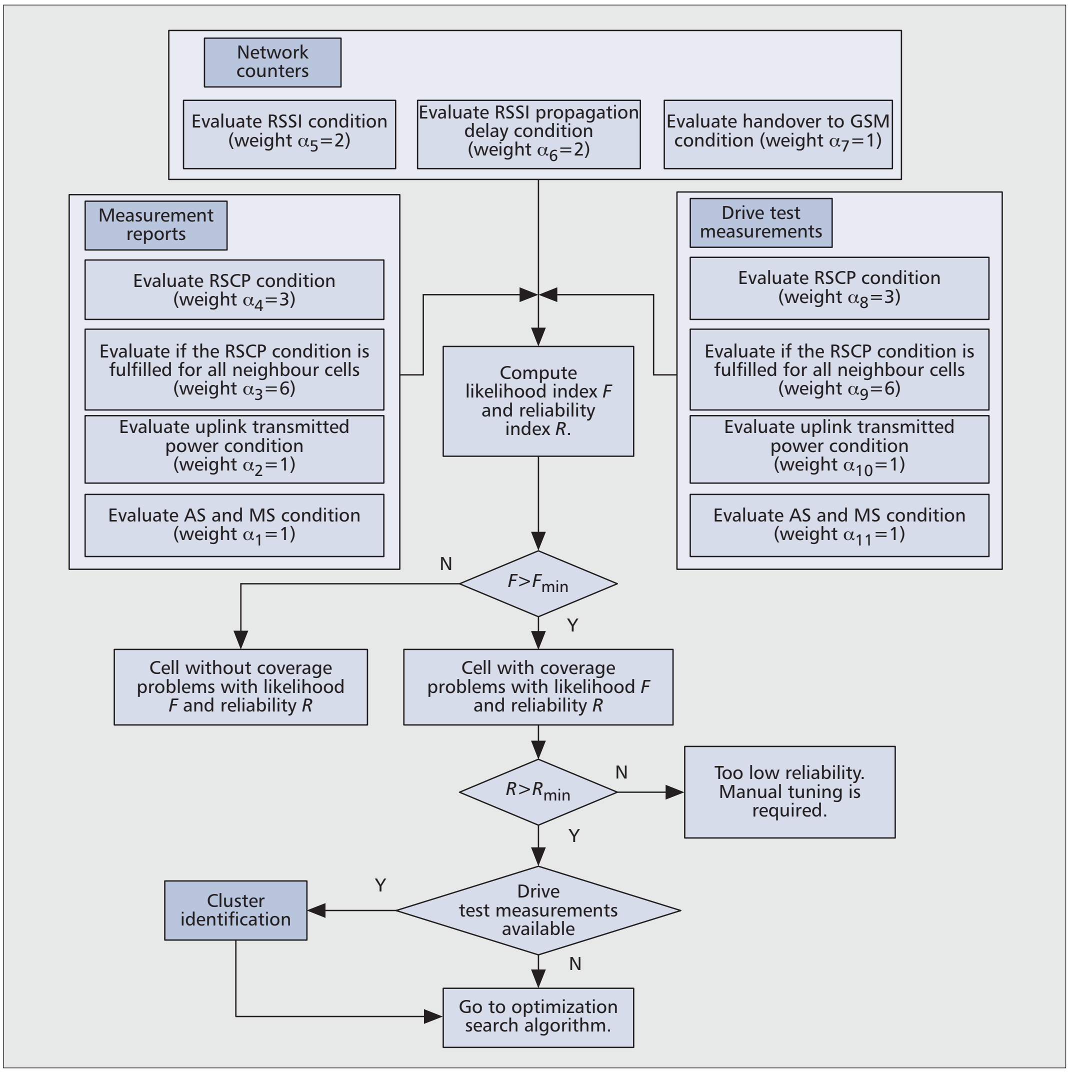


Figure 3. *Detection of sub-optimal operation algorithm.*

The statistical distribution of the propagation delay of specific signals transmitted from mobiles to cell is considered to be a valuable input because it reflects the statistical distribution of mobiles-to-cell distances when connected to the cell.

also for the neighboring cells, so as to increase the likelihood of sub-optimal coverage in case neither the cell under study nor any of its neighboring cells are detected with a high enough RSCP. According to these conditions and its corresponding weights α_i shown in Fig. 3, the algorithm determines the reliability index R and the likelihood index F in the detection of coverage problems. If the likelihood and reliability index are higher than certain thresholds (e.g., $R_{min} = 0.4$ and $F_{min} = 0.6$ are chosen in this case study), coverage problems in the cell under study are assumed and the optimization search is activated.

In order to gain some insight into the network status in the considered scenario, some examples of measurements used for the detection of sub-optimal cell coverage are presented and discussed in the following. On the one hand, Fig. 4 shows the Cumulative Distribution Function (CDF) of the RSCP for different cells when each considered cell is detected either in the active set if the mobile is in connected mode or as best server if the mobile is in idle mode. While the RSCP in *Cell_6_2* is below –85dBm for around 90 percent of the samples available in the cell, *Cell_6_1* exhibits only 10 percent below that threshold and *Cell_6_3* exhibits that almost all samples are above the threshold. Therefore, the observation of this plot may point out coverage problems within *Cell_6_2*. On the other hand, the correlation between the uplink transmitted power and the RSCP is presented in Fig. 5. A cell with coverage problems may have a high percentage of measurements with a high uplink transmitted power and a low RSCP (e.g., *Cell_6_2*). In terms of the algorithm for the detection of sub-optimal coverage, this linkage has been captured by including one condition on RSCP and another condition on the uplink transmitted power. If both are fulfilled (i.e., low RSCP in downlink and high transmitted power in uplink are observed), the likelihood index F that the cell is affected by coverage problems increases. On the contrary, cells like *Cell_6_3*, with low values of uplink transmitted power and high RSCP values, are not expected to have coverage problems.

In order to assess the performance of the automated coverage optimization process in this real scenario, the algorithm presented in Fig. 3 is executed. Considering the weights for each con-

Communications Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page qMags

dition indicated in Fig. 3 and the definitions of likelihood index and reliability index indicated in Eqs. 1 and 2, the result obtained with the available data for *Cell_6_2* is $F = 0.93$ and $R = 0.48$. Therefore, since $F \geq F_{min} = 0.6$, the most likely hypothesis is that the coverage is not sufficiently optimized. Besides, since $R \geq R_{min} = 0.4$, the optimization search stage is triggered. Furthermore, since drive test measurements are available in this case, the algorithm is able to identify three clusters where the sub-optimal coverage is localized (the identified clusters are shown by circles in Fig. 6).

The optimization search considered here is based on readjusting the antenna downtilt of the cell under study (or some neighbors) so as to improve the coverage in the cluster locations while maintaining proper coverage in the rest of the cell. From a practical perspective, electrical tilt would be preferred, while mechanical tilt should be limited to a certain maximum in order to avoid deformation on the main radiation lobe.

The procedure makes use of the knowledge about the antenna radiation pattern of the different cells together with the measurements available from the drive tests and the information about the positions of the different clusters identified in the detection step. The procedure first considers downtilt changes for the cell where the problem was detected (*Cell 6_2* in this example). An iterative search is carried out by shifting the downtilt in steps of Δ degrees up to a maximum Δ_{max}. For each iteration and for each sample corresponding to each identified cluster an estimation of the resulting RSCP value is computed using the antenna radiation pattern. As a result, an estimation of the percentage of samples with expected RSCP above threshold RSCP* is obtained for each considered downtilt. If a downtilt providing sufficient improvement is found, then a recommendation for parameter change is drawn. Otherwise, the algorithm concludes that it is not possible to optimize the coverage of this cell through downtilting on the same cell and a solution search over neighboring cells is triggered. In that case, a similar procedure is executed by considering different downtilts in neighboring cells. The algorithm will be able to provide a recommendation about downtilt change in a neighboring cell provided that the neighbor can significantly improve the coverage over the affected clusters whilw still maintaining proper RSCP levels over its own coverage area.

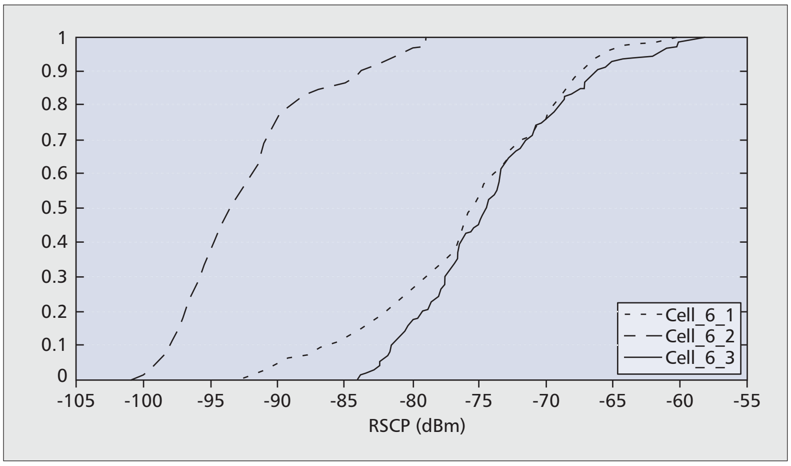


Figure 4. *Cumulative distribution function of the RSCP.*

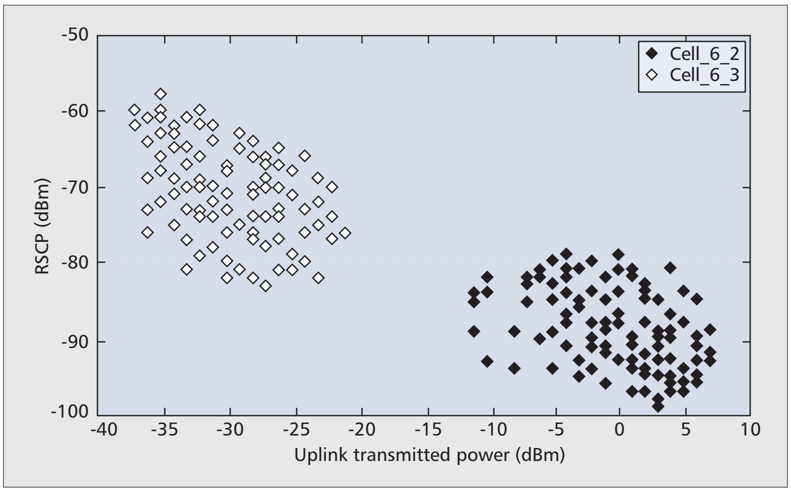


Figure 5. *Correlation between RSCP and uplink transmit power.*

Applying the optimization search algorithm in the analyzed scenario, it is found that no downtilt change on *Cell_6_2* would provide sufficient RSCP improvement over the cell. Therefore, the solution search was extended to neighboring cells. In this case, it was found that a downtilt change from 11° to 4° in *Cell_4_1* would improve the overall situation.

In order to understand the outcome of the optimization algorithm, Fig. 7 shows, on the one hand, the CDF for the RSCP from *Cell 6_2* corresponding to the samples taken along the drive test segment marked in Fig. 6. It can be clearly observed that poor coverage is provided over *Cell 6_2*, since almost 85 percent of the samples are below –88 dBm. In a closer view, measured data over the drive test (not shown in the figures for the sake of brevity) reveal that *Cell 6_2* is not detected at all in Cluster 1, while it is detected with an average value of –92 dBm in Cluster 2 and –93 dBm in Cluster 3. In turn, with respect to *Cell 4_1*, this becomes the best server along Clusters 1, 2 and 3 (which are within the theoretical coverage area from *Cell 6_2*), although it is received with very low values of RSCP, in most of the cases below -88 dBm.

Figure 7 shows, on the other hand, the CDF for the RSCP received from *Cell 4_1* along the drive test for downtilt values of 4° and 11°. Thanks to the downtilt change from 11° to 4° the probability of low RSCP values is reduced, which illustrates that RSCPs from *Cell 4_1* improve at Clusters 1, 2 and 3. This is at the expense of reducing the probability of high RSCP, although this is not impacting negatively on the coverage provided by *Cell 4_1* along its own theoretical coverage area. After the downtilt adjustment in *Cell_4_1*, the likelihood index and the reliability index for *Cell_6_2* are calculated again for the new situation. $F = 0.46$ and $R = 0.48$ are obtained. F is reduced because the RSCP condition for neighbor cells (Fig. 3) is not fulfilled in the resulting scenario. Given that $F < F_{min}$, the output of the algorithm is that coverage is sufficiently optimized and, therefore, the coverage problem has been solved with this action.

In conclusion, this example has shown a practical case where a cell provides poor coverage, which is eventually overcome by increasing the overlapping from a neighboring cell. Indeed, in order to verify the conclusions reached in this exercise, a closer analysis in-the-field was conducted and it was found that there were some high buildings obstructing the antenna of *Cell_6_2* in the direction of the clusters, so it was not possible to improve the coverage in those areas by acting on the downtilt of that particular cell.

EXTENSION TO LTE

In order to illustrate that the actual work on UMTS optimization can also provide valuable background for LTE, this section discusses how the coverage optimization case study presented for UMTS could be extended to the LTE case. In this way, the usefulness of the presented framework as a basis for LTE is emphasized, even though the type of optimization targets to be considered and the specific KPIs can be different as they are technology-specific. Similarly, there would be differences in terms of system architecture, so that while in UMTS many of the KPIs can be available at the RNC, in LTE most of the KPIs will be available at the eNode-Bs.

Table 2 presents the extension of the KPIs and tunable parameters from the UMTS case to the LTE case. As can be observed, the conditions to detect the sub-optimal coverage are also based on statistics related to power measurements, handovers, and propagation delay. Nevertheless, there is some specificity as indicated in the following:

- The corresponding measurement that can be equivalent to the RSCP in UMTS is the Reference Signal Received Power (RSRP), which corresponds to the measured power by the mobile terminal at the specific reference signals transmitted by the eNode-B [21].
- As for the uplink case, given that in LTE there also exists a power control scheme, the uplink transmitted power in either Physical Uplink Control Channel (PUCCH) or Physical Uplink Shared Channel (PUSCH) that can be collected through drive tests or measurement reports, can also be used to identify poor coverage situations whenever the transmit power is above some threshold, reflecting excessive propagation losses. Note also here that the RSSI in the uplink that was considered in UMTS is no longer considered in the LTE case, given that this measurement is actually not collected by the eNode-B [21].
- The condition regarding the number of handover attempts to GSM has been reformulated to also include the handovers to UMTS, to reflect that a large number of handovers to these technologies could be an indication of LTE coverage problems.
- The condition regarding the AS and MS configuration used in UMTS is no longer considered in LTE, given that no soft handover is used. Instead, the considered condition corresponds to the monitoring of the RSRP in the neighboring cells. In particular, if the best cell is not actually the serving cell, this could reflect problems in the signaling associated with the handover procedure that can be due to coverage problems.
- As for the condition related to the propagation delay, it can also be used in the context of LTE. While in UMTS the estimation of this metric is implementation-dependent and could be obtained from the delays in the random access, the availability of time advance measurements in LTE makes even more straightforward the estimation of the propagation delay, with a resolution of 0.52 μs [22].

Based on the above conditions, the algorithm for the detection of sub-optimal coverage in LTE would be very similar to the one presented in Fig. 3 for UMTS with the corresponding modifications in accordance with the KPIs and condi-

Figure 6. *Scenario and location of the identified clusters.*

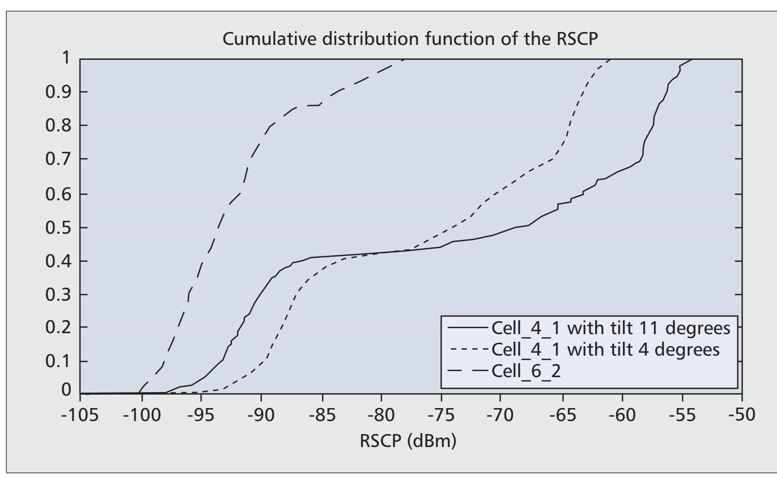


Figure 7. *CPICH RSCP before and after the downtilt correction.*

Considered KPIs	Criteria/Condition	Source
RSRP (Reference Signal Received Power) of serving cell	$Prob(RSRP<RSRP^*)>Th_{RSRP}(\%)$	Drive Test/Measurement Report
Uplink transmission power P_T	$Prob(P_T>P_T^*)>Th_{PT}(\%)$	Drive Test/Measurement Report
RSRP of neighboring cells	The best cell is not the serving cell	Drive Test/Measurement Report
Number of handover attempts to GSM/UMTS relative to cell load	$\frac{NumberHO\_GSM + NumberHO\_UMTS}{Cell\_throughput} > Th_{HO}$	Network counters
Statistical distribution of the propagation delay between mobiles and base station.	$Prob(Propdelay<P^*)>Th_{propdelay}(\%)$	Network counters
Network configuration parameters/Tunable parameters		
Positions of the different e-nodes B in the network.		
Antenna azimuth of the different cells.		
Neighbor lists.		
Transmit power in the different cells.		
Antenna downtilt of the different cells		

Table 2. *Considered KPIs and tunable parameters for the LTE case.*

tions from Table 2. Similarly, the concepts of reliability R and likelihood index F would also be applicable, although perhaps with different values depending on the weights to be set for each condition.

Insights for LTE Self-Optimization

SON is introduced as part of the 3GPP LTE as a key driver for improving O&M. SON concepts are introduced in LTE starting from the first release of the technology (Release 8), and expanding in scope with subsequent releases. The progressive inclusion of standardized SON features certainly reflects the expected LTE network evolution stages as a function of time. For example, Release 8 includes functions covering different aspects of the eNode-B self-configuration (automatic inventory, automatic software download, automatic neighbor relation, automatic physical Cell Identifier (Cell ID) assignment [23]). The objective of self-configuration is to provide "plug and play" functionality in the planning, integration, and configuration of new eNode-Bs. In turn, Release 9 will provide SON functionalities addressing optimization. This phased approach, targeting automated optimization for more maturing network stages, seems fairly reasonable given the ambitious objectives of a full SON. Some of the considered use cases in Release 9 are:

• Mobility robustness optimization, encompassing the automated optimization of parameters affecting active mode and idle mode handovers to ensure good end-user quality and performance.

• Load balancing optimization, in order to intelligently spread user traffic across the system's radio resources as necessary in order to provide quality end-user experience and performance, while simultaneously optimizing system capacity.

• Coverage and capacity optimization, which is related to the case study presented earlier. While the goal is the same regardless of radio technology, the specific algorithms and parameters vary with technology. This has been illustrated earlier, which has outlined a possible solution.

Release 9 also includes elements to increase energy savings, UE reporting functionality to minimize the amount of drive tests, etc. Nevertheless, it is worth emphasizing that SON-related functionalities will continue to expand through the subsequent releases of the LTE standard.

In this context, where a full LTE-SON can be set as the skyline ultimate objective, it is important to devise a roadmap starting from actual commonly-used manual-based UMTS optimization. The practical experience gained so far, as reflected in this article, has brought to some observations, sometimes over-sought when tackling the problem from a theoretical/simulation perspective, as well as to the identification of some key open points to be tackled in this evolutionary process toward LTE-SON. Similarly, lessons learned can also be of help and provide a complementary and reinforcing approach, particularly at the time that LTE-SON is still taking full shape and while it cannot benefit from live network experiences. In this respect, it is worth mentioning:

- Given that the practical implementation of SON concepts puts strong requirements on the capabilities offered by O&M tools and impacts all aspects of an operator's radio engineering department, early actions enabling a smooth introduction of the "SON culture" may be highly beneficial on a long-term perspective.
- The optimization framework presented earlier can be considered as a reference starting point for LTE, given that the main principles highlighted are technology-independent. In particular, the different sources to acquire the network status, the network optimization loop and the process in the network optimization algorithm are, indeed, of rather general applicability.
- Related to SON architectures, guiding principles and operational procedures, again the framework proposed in this article, which has been derived from a practical perspective, can be useful for the consolidation of SON in the framework of LTE [24].
- Given the uncertainty about the network status, outputs from SON algorithms require associated likelihood/reliability indicators. Therefore, this approach as presented in Section II with *F* and *R* parameters tightly coupled with the detection of sub-optimal behavior could also be applicable in the LTE context.
- Given that changing the network configuration/parameterization is a critical action from the operator's perspective, the automated optimization procedure is likely to be introduced in a step-by-step approach. While automatic sub-optimal behavior detection mechanisms can be considered mature enough (as reflected in the presented UMTS case study), the automatization of optimization search and parameter change is seen at this stage, and even for UMTS networks, far from mature. The confidence in such automatic procedures as required by the operator needs to be significantly enhanced, as discussed in the points below.
- Optimization search procedures need to be specifically developed for each possible parameter. Criteria to anticipate the most suitable parameter to cope with a given sub-optimal operation need to be established.
- Supporting tools and/or models for the optimization search stage require further research. The estimation of the impact of a given parameter change on the live network as part of the optimization procedure is a critical point.
- The ultimate view of SON, where the joint self-optimization of several targets is performed, may raise complex interactions. The influence of the execution of a stand-alone optimization target on the rest of the targets needs to be investigated.
- In the best case, SON will be able to automatically detect a sub-optimal operation and provide a corrective action. However, in some cases the SON algorithm may fail in detecting the problem or may be unable to provide a recommendation for a corrective action.
- Besides, there are incipient market solutions for UMTS optimization that allow KPIs' delocalization. This promising feature, which is expected to mature in a LTE context, may reduce the need for drive tests in the future and pave the way for a full self-optimization framework.

Conclusion and Future Work

SON is introduced as part of the 3GPP LTE as a key driver for improving O&M. However, given that many challenges are identified when moving from the SON concept to practical implementation, this article has claimed the suitability to gain insight into the problem by taking advantage of optimization mechanisms in live UMTS networks.

The article has developed an optimization framework considering different stages: inputs from different possible sources (network counters, measurement reports, drive tests), tuning parameters to achieve optimization and the optimization procedure itself (including KPIs processing, sub-optimal operation detection for a given target and optimization search). The optimization process has been formulated in the form of hypotheses tests against the sub-optimal operation for a number of established targets. Hypotheses are reinforced by likelihood and reliability indexes, so that the number and relevance of diverse KPIs with respect to the considered target can be weighted.

The optimization framework has been applied to a UMTS coverage optimization use case in a large European city. The algorithm has allowed the automated identification of a cell with sub-optimal coverage and provided a solution to better optimize its operation. The extension to LTE has been analyzed, identifying a number of specificities to be considered due to technology change.

Finally, this article has provided some insight for LTE self-optimization. On one side, it has been discussed that the presented optimization framework could also be taken as a reference in the LTE context. On the other side, lessons learned from the UMTS case study have enabled the identification of a number of key issues to be solved in the evolutionary path toward LTE-SON. Besides, it has been highlighted that a step-by-step introduction of SON can be sound in practice, implementing self-detection with manual self-optimization in a first stage and progressively moving toward automatization of the optimization search.

Acknowledgments

This work has been supported by Telefónica. This article reflects the authors' views and does not necessarily represent the view of Telefónica.

References

[1] UMTS Forum, Annual Report 2008 and Directions for 2009

[2] NGMN Alliance Deliverable, "NGMN Use Cases Related to Self Organizing Network, Overall Description," May 2007.

[3] INFSO-ICT-216284 SOCRATES project, "D2.1: Use Cases for Self-Organizing Networks," March 2008.

[4] 3GPP TR 32.816, "Study on Management of Evolved Universal Terrestrial Radio Access Network (E-UTRAN) and Evolved Packet Core (EPC)"

[5] 3GPP TR 36.902, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-Configuring and Self-Optimizing Network Use Cases and Solutions."

[6] M. Amirijoo *et al.*, "Cell Outage Management in LTE Networks," *6th Int'l. Symp. Wireless Commun. Systems (ISWCS 2009)*, Italy, Sept. 2009, pp. 600–04.

[7] C. Prehofer and C. Bettstetter, "Self-Organization in Communication Networks: Principles and Design Paradigms," *IEEE Commun. Mag.*, vol. 43, no. 7, July, 2005, pp. 78–85.

There are incipient market solutions for UMTS optimization that allow KPIs' delocalization. This promising feature, which is expected to mature in a LTE context, may reduce the need for drive tests in the future and pave the way for a full self-optimization framework.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

A step-by-step introduction of SON can be sound in practice, implementing self-detection with manual self-optimization in a first stage and progressively moving toward automatization of the optimization search.

[8] H. Hu *et al.*, "Self-Configuration and Self-Optimization for LTE Networks," *IEEE Commun. Mag.*, vol. 48, no. 2, Feb., 2010, pp. 94–100.
[9] J. Pérez-Romero *et al.*, *Radio Resource Management Strategies in UMTS*, John Wiley & Sons, 2005.
[10] I. Siomina, P. Värbrand, and D. Yuan, "Automated Optimization of Service Coverage and Base Station Antenna Configuration in UMTS Networks," *IEEE Wireless Commun.*, vol. 13, no. 6, Dec. 2006, pp. 16–25.
[11] J. Zhang *et al.*, "Mathematical Modeling and Comparisons of Four Heuristic Optimization Algorithms for WCDMA Radio Network Planning," *Proc. 8th IEEE Int'l. Conf. Transparent Opt. Networks*, June 2006, pp. 253–57.
[12] S. Ben Jamaa *et al.*, "Steered Optimization Strategy for Automatic Cell Planning of UMTS Networks," *IEEE 61st Vehic. Tech. Conf. (VTC 2005 Spring)*, June, 2005, pp. 359–62.
[13] S. Allen *et al.*, "Smart Cell Planning and Optimization for UMTS," *5th IEE Int'l. Conf. 3G Mobile Commun. Technologies (3G 2004)*, 2004, pp. 34–38.
[14] J. Laiho *et al.*, Eds., *Radio Network Planning and Optimization for UMTS, 2nd Edition*, John Wiley & Sons.
[15] C. Chevalier *et al.*, *WCDMA Deployment Handbook: Planning and Optimization Aspects*, John Wiley & Sons, 2006.
[16] J. Sánchez-González *et al.*, "A New RF Methodology for Failure Detection in UMTS Networks," *IEEE/IFIP Network Operations and Management Symp.*, Salvador da Bahia, Brazil, April, 2008, pp. 718–21.
[17] 3GPP TS 25.215 "Physical Layer Measurements (FDD) (Release 5)."
[18] 3GPP TS 25.331 "Radio Resource Control (RRC) (Release 5)."
[19] D.E. Goldberg, *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison Wesley, 1989.
[20] C.M. Fonseca and P. J. Fleming "Multiobjective Optimization and Multiple Constraint Handling with Evolutionary Algorithms — Part I: A Unified Formulation," *IEEE Trans. Systems, Man and Cybernetics — Part A: System and Humans*, vol. 28, no. 1, Jan. 1998, pp. 26–37.
[21] 3GPP TS 25.214 v8.7.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer — Measurements (Release 8)"
[22] 3GPP TS 25.213 v8.1.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer — Procedures (Release 8)"
[23] 3G Americas White Paper, The benefits of SON in LTE, December 2009
[24] T. Jansen *et al.*, "Embedding Multiple Self-Organization Functionalities in Future Radio Access Networks," *69th IEEE Vehic. Tech. Conf. (VTC 2009 Spring)*, Barcelona, Apr. 2009.

Biographies

ORIOL SALLENT (sallent@tsc.upc.edu) is Associate Professor at the Universitat Politècnica de Catalunya. His research interests are in the field of mobile communication systems, especially radio resource and spectrum management for cognitive heterogeneous wireless networks. He has published more than 150 papers in international journals and conferences. He has participated in several research projects of the 5th 6th and 7th Framework Program of the European Commission and served as a consultant for a number of private companies.

JORDI PÉREZ-ROMERO is currently associate professor at the Dept. of Signal Theory and Communications of the Universitat Politècnica de Catalunya (UPC) in Barcelona, Spain. He received the Telecommunications Engineering degree and the Ph.D. from the same university in 1997 and 2001. His research interests are in the field of mobile communication systems, especially packet radio techniques, radio resource and QoS management, heterogeneous wireless networks and cognitive networks. He has been involved in different European Projects as well as in projects for private companies. He has published papers in international journals and conferences and has co-authored one book on mobile communications. He is associate editor of IEEE Vehicular Technology Magazine and Eurasip Journal on Wireless Communications Networks.

JUAN SANCHEZ-GONZALEZ received the Engineer of Telecommunications and Ph.D. degrees from the Universitat Politècnica de Catalunya (UPC), Barcelona, in 2002 and 2007, respectively. He is currently assistant professor in the department of Signal Theory and Communications in UPC. He has participated in projects funded by the European Union as well as in projects for private companies. His research interests include radio resource and QoS management for wireless communication networks, optimization of cellular systems and cognitive networks.

RAMON AGUSTI received the Engineer of Telecommunications degree from the Universidad Politécnica de Madrid, Spain, in 1973, and the Ph.D. degree from the Universitat Politècnica de Catalunya (UPC), Spain, 1978. He became Full Professor of the Department of Signal Theory and Communications (UPC) in 1987. After graduation he was working in the field of digital communications with particular emphasis on transmission and development aspects in fixed digital radio, both radio relay and mobile communications. For the last fifteen years he has been mainly concerned with aspects related to radio resource management in mobile communications. He has published about two hundred papers in these areas and co-authored three books. He participated in the European program COST 231 and in the COST 259 as Spanish representative delegate. He has also participated in the RACE, ACTS and IST European research programs as well as in many private and public funded projects. He received the Catalonia Engineer of the year prize in 1998 and the Narcis Monturiol Medal issued by the Government of Catalonia in 2002 for his research contributions to the mobile communications field. He is a Member of the Spanish Engineering Academy.

MIGUEL ÁNGEL DIAZ-GUERRA received the Telecommunications Engineering degree from the Universidad Politécnica de Madrid, Spain, in 1992. In 1994 he joined Telefónica. He has been working in Radio Planning and Radio Optimization during 16 years. He has collaborated in different research projects with University. He has participated in international projects inside Telefónica Group in Latin-American and Europe. He currently is Access Network Engineering Manager.

DAVID HENCHE received the Telecommunications Engineering degree from the Universidad Politécnica de Cataluña, Spain, in 1993. In 1995 he joined Telefónica, where he currently is Radio Access Network Engineer working in the planning and deploying area. His main interests are in the field of internal and external interference detection and removal to increase network performance, and optimal deployment of mobile networks for current and new future data services.

DANIEL PAUL received the Telecommunications Engineering degree from the Universidad Politécnica de Madrid, Spain, in 1989. In 2001 he joined Telefónica, where he currently is Access Network Engineering Director. Prior to that, he was with Motorola, where he has been working from 1990 participating in the initial development of the GSM system and participating in the GSM deployment in a number of operators in Europe. His main interests are in the field of new techniques of radio access network optimization and the optimal deployment of MBB networks adjusted for the new future data services.

Communications | Previous Page | Contents | Zoom in | Zoom out | Front Cover | Search Issue | Next Page | qMags

ACCEPTED FROM OPEN CALL

Medium Access Control for Power Line Communications: An Overview of the IEEE 1901 and ITU-T G.hn Standards

Md. Mustafizur Rahman, Choong Seon Hong, and Sungwon Lee, Kyung Hee University
Jaejo Lee, Korea Electrotechnology Research Institute
Md. Abdur Razzaque and Jin Hyuk Kim, Kyung Hee University

ABSTRACT

Power lines are becoming increasingly popular for high-speed communications, while at the same time reliable communications over them poses several unique challenges. Recently, two standardization bodies, the Institute of Electrical and Electronics Engineers (IEEE) and the International Telecommunication Union (ITU), addressed in their IEEE 1901 and ITU-T G.hn standards, respectively, both the medium access control and the physical layer specifications of power line communications. In this article, we provide an overview of these standards from the media access control technology perspective, and the similarities and dissimilarities between the IEEE 1901 and ITU-T G.hn standards are discussed, as well as potential performance issues. We use a top-down approach to analyze the features and technologies of the standards.

INTRODUCTION

Home networking is a term that encompasses numerous important applications, such as indoor wired LAN services with broadband Internet access for both residential and commercial areas. These applications include smart home appliances and security, in-home audio/video streaming, and other applications such as traffic and street light controlling [1, 2].

Power line communication (PLC) is a good candidate for home networking because of the availability of power lines in almost every existing infrastructure. However, achieving good performance over the power lines poses unique challenges to modem designers [3]. Several technical problems are faced when using power line networks as the medium for high-speed data communication, because these networks were initially designed to distribute electric power at very low frequencies (50 Hz or 60 Hz) [1]. Power line cables are usually found to be unshielded and therefore generate and are subject to electromagnetic interference. However, recent developments in data-modulation and transmission techniques have made it possible to use power lines for high-speed communications, so that products are today available on the market for various PLC applications. In 2005, the IEEE Communications Society formed the 1901 working group (P1901 WG), which was tasked with creating an international technical standard for high-speed (>100 Mb/s at the physical layer) communication through AC electric power lines in order to bring PLC products into a common network framework.

The HomeGrid Forum [4], a global and nonprofit trade group, and its members have been supporting and contributing to G.hn specification of the International Telecommunication Union's Telecommunication standardization sector (ITU-T). The goal of G.hn is to unify connectivity of digital content and media devices by providing a wired home network over telephone, coaxial, and data-grade cable networks, as well as residential power line wiring to supply data at rates of up to 1 Gb/s. Thus, it addresses key concerns of service providers, electronic manufacturers, and consumers alike [2]. The goal of the ITU-T G.hn standard was to define the physical (PHY, layer 1) and link (layer 2) layers for home-wired networks; this work culminated in the ITU-T Recommendations G.9960 and G.9961 specifying the physical and data link layers of G.hn, respectively [5, 6]. The unified technology is desirable to reduce the cost and complexity of installation and to allow cross-device communication and functionalities. With the G.hn standard, ITU enables service providers to deploy new offerings, including Internet Protocol television (IPTV) more cost effectively. It also allows manufacturers in the consumer electronics market to provide powerful devices for connecting a variety of entertainment, home automation, and security products throughout the house, thus simplifying consumer purchasing and installation processes. In this article, the description of G.hn relates to power line operations only.

Both home networking standards discussed

0163-6804/11/$25.00 © 2011 IEEE

Feature and Technology	IEEE 1901		ITU–T G.hn
	FFT-PHY	Wavelet-PHY	
Channel Access			
Fundamental Technology	CSMA/CA	CSMA/CA	TDMA, CSMA/CA
Contention-based Scheme	CSMA/CA	CSMA/CA	CSMA/CA
RTS/CTS Reservation	Optional	Optional	Optional
Access Priorities	4	8	4
Virtual Carrier Sensing	Yes	Yes	Yes
Contention-free Scheme	TDMA	TDMA	TDMA
Persistent Access	Yes	Yes	Yes
Access Administration	Beacon Based	Beacon Based	MAP Based
Quality of Service	Supported	Supported	Supported
Security			
Security Framework	DSNA/RSNA	PSNA/RSNA	AKM
Encryption Protocol	CCMP	CCMP	CCMP
Burst Mode Operation	Uni-/bi-directional	Not supported	Bi-directional
Addressing Scheme			
Modes	Uni-, Multi-, and Broadcast	Uni-, Multi-, and Broadcast	Uni-, Multi-, and Broadcast
Space (per domain)	8-bit	8-bit	8-bit
Framing			
Aggregation Fragmentation and Reassembly	Supported	Supported	Supported

Table 1. *Features and technologies of the current PLC MAC standards at a glance.*

	ITU-T	IEEE
High-rate Broadband Access	No	1901
High-rate Broadband In-home	G.hn (50/100 MHz)	1901
Low-rate Broadband In-home	G.hn (25 MHz)	No
Low-rate, Low-frequency Narrowband	G.hnem (500 kHz)	1901 (500kHz)

Table 2. *Application areas of PLC MAC standards.*

above, although designed from different perspectives, share some common core technologies and functionalities, because their design goals are based on current market demand for standards to support several high-definition television (HDTV) and standard definition television (SDTV) streams, voice over IP (VoIP), and data traffic (i.e., Internet services, gaming etc.). To enable handling of such different types of traffic, both standards provide contention-free and contention-based channel access, Quality of Service (QoS) provisioning, and security, among other features. Some of the features offered are based on the same core technologies, while others use different technologies. The medium access control (MAC) features supported by IEEE 1901 and ITU-T G.hn are summarized in Table 1.

The fundamental difference between the two technologies is their application domains. Table 2 describes possible applications for ITU-T G.hn and IEEE 1901. ITU-T G.hn targets to high-rate broadband in-home, low-rate broadband in-home, and low-rate low-frequency narrowband. And IEEE targets high-rate broadband access and high-rate in-home applications. However, the New Standards Committee (NesCom) of the IEEE-SA Standards Board approved a revision to the scope of the Project Authorization Request (PAR) that clarifies the usability of the P1901 standard for Smart Grid applications, for transportation platforms (vehicle) applications, and for broadband over power line (BPL) devices operating on DC lines.

The low rate profiles of ITU-T G.hn enable scalable design supporting multiple bandwidths, and this technology supports seamless communications between low-end and high-end devices. A new project called "Home Networking Aspects of Energy Management" was started by ITU-T in association with the International Electrotechnical Commission (IEC) and Joint Coordination Activity on Home Networking (JCA-HN) to produce G.hnem in January 2010. The main goal of G.hnem is to define low complexity home networking devices for home automation, home control, electrical vehicles, and Smart Energy applications. The G.hnem standard was expected to be completed by February 2011.

The organization of standard documents in the two protocols is fundamentally different, as shown in Table 3. While ITU-T PLC targets coaxial cables, phone lines, and power lines as communications media, IEEE PLC only targets power lines.

Channel Access

In accessing channels, both IEEE 1901 and ITU-T G.hn provide non-prioritized and prioritized contention-based services for best effort and QoS-required traffic, while contention-free access is provided for QoS-guaranteed traffic as well. The fundamental access technology differs between the protocols; both IEEE 1901 (using CSMA/CA) and IUT-T G.hn (using CSMA/CA and TDMA) adopt synchronized access by beacon/MAP.

MAC Channel

PLC MAC protocols arrange access to the media by continuous MAC cycles. The MAC cycle is usually synchronized with the AC line cycle, but it can be synchronized with any external source. The MAC cycle in IEEE 1901 has three regions for three types of channel access, as shown in Fig. 1. The local administrator device (BSS Manager or BM) starts a new MAC cycle by transmitting beacon(s) during the beacon period. The beacons contain the start time of the other two periods in the cycle. In the contention period (CP), a station accesses the channel after a contention through CSMA/CA and backoff procedures. The TDMA-based contention-free access period starts after the CP. Multiple BSSs can be located in the same network. In that time, BSSs recognize each other and they share the medium. In this case, special periods called Reserved and Stayout are used to coordinate the access between multiple BSSs at the same location.

The management of MAC cycles in G.hn, however, is slightly different from that in IEEE 1901. First, there is no beaconing in the G.hn standard. Instead, a Medium Access Plan (MAP) is used to describe a MAC cycle. The local administrator, namely the Domain Manager or DM, transmits at least one MAP in a MAC cycle that defines the transmission opportunities (TXOP) of the following (one or several) MAC cycles. Second, the MAP in G.hn schedules TXOPs in a TDMA fashion, as shown in Fig. 1; however, it also enables CSMA/CA-driven contention-based access in one (or several) TXOP(s). The DM reserves TXOP(s) of the subsequent MAC cycles to broadcast the MAP for future cycles. Note here that ITU-T G.hn and IEEE 1901 differ in how they use information and which devices access the media. In G.hn, the beacon is used for the present cycle, while in MAP, it is used for the next cycle.

	ITU-T PLC	IEEE PLC
PHY Layer	Single (OFDM)	Dual (OFDM/ Wavelet)
MAC Layer	Single (OFDM)	Dual (OFDM/ Wavelet)
Target Medium	Coax, Phone line, Power line	Power line
Standard Docs.	G.hn (MAC-G.9961) G.hn (PHY-G.9960) G.cx (Coexistence-G.9972) G.hnem (Narrowband)	IEEE 1901 (MAC/PHY/Coexistence)

Table 3. *Organization of standard documents.*

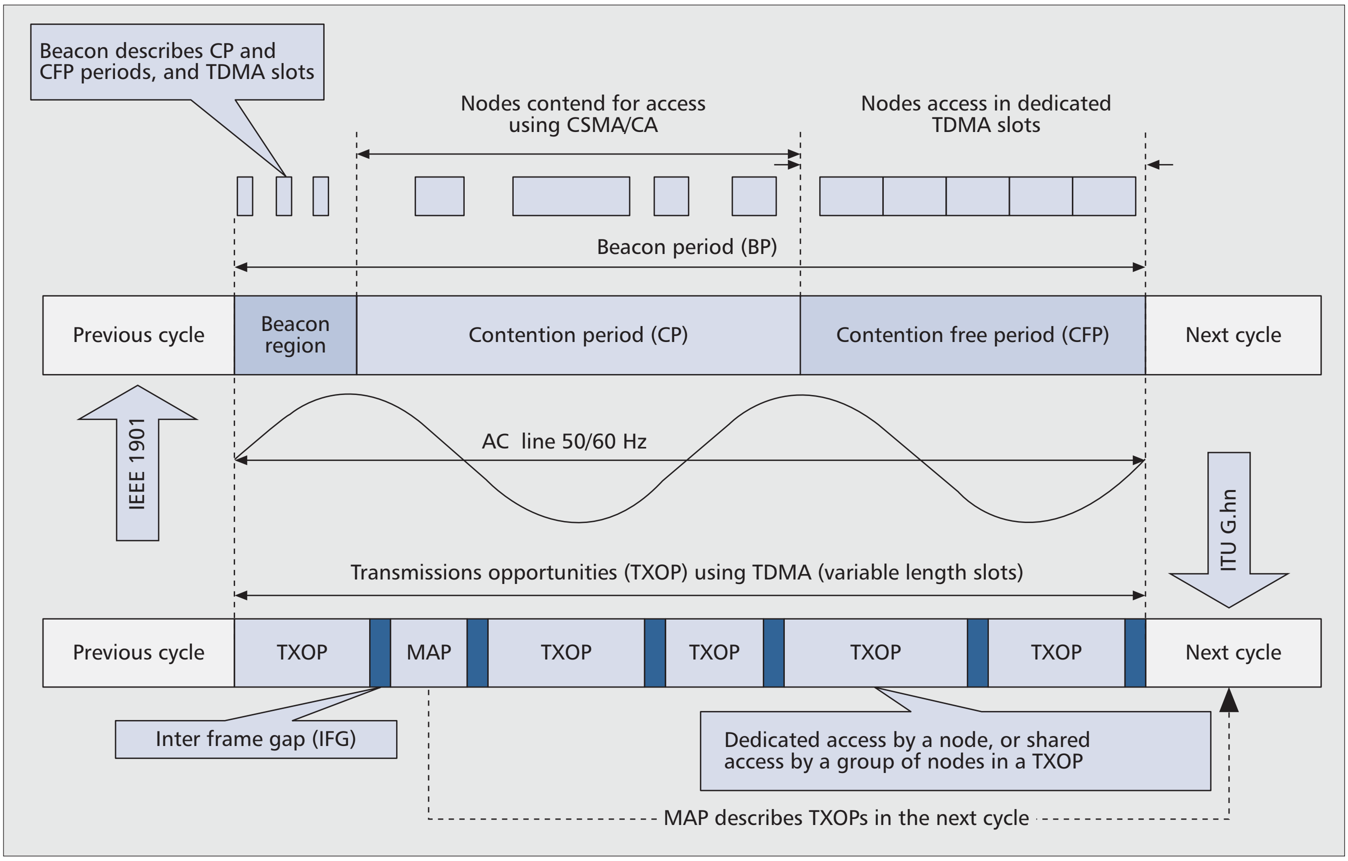


Figure 1. *The MAC cycle in IEEE 1901 and ITU-T G.hn.*

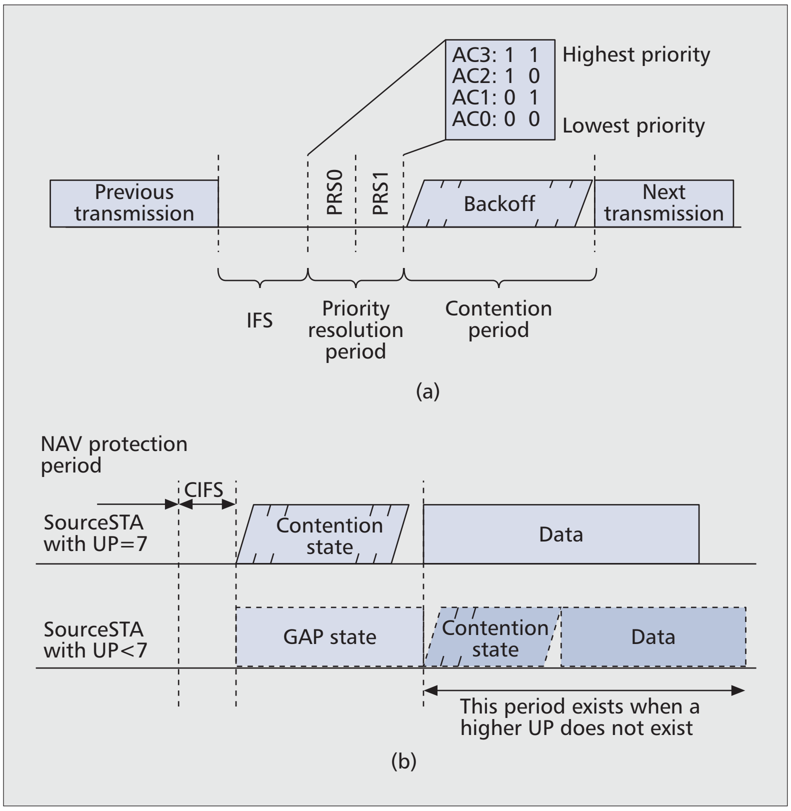


Figure 2. *Typical contention-based access in IEEE 1901: a) For FFT OFDM PHY; b) for wavelet OFDM PHY.*

Contention-Based Access

Both standards provide contention-based access for best effort and QoS-required services with the help of CSMA/CA. However, because of the different cycle structures in the standards, the channel access procedures of the two protocols are fundamentally different from each other. In IEEE 1901, there is at least one CP in a MAC cycle after the beacon period. In G.hn, multiple contention-based time slots can exist in a single MAC cycle. The location and duration of these time slots are defined by DM. Between the contention-based time slots, contention free-time slots and MAP exist.

Contention Based Access in IEEE 1901 — In the contention-based access for this protocol, an IEEE 1901 station checks for a carrier in the channel before transmitting a data frame. A station transmits data immediately when it wins the priority contention. If the station fails to win the priority contention, then it follows a random backoff procedure. After deferral, or prior to attempting to transmit again immediately after a successful transmission, the station initiates a random backoff procedure to resolve medium contention conflicts. There is nevertheless still a probability of collision between contending stations when two or more stations defer for the same period and transmit simultaneously. A successful transmission is one in which an Ack frame is received from the receiving station. A station can make use of RTS/CTS handshaking prior to data frame transmission to further minimize collisions under various circumstances.

IEEE 1901 manages access to the medium differently for the two different PHY technologies. In FFT OFDM PHY-based MAC, stations that want to transmit send their own priority resolution period, as shown in Fig. 2a. Stations that do not have a higher priority than other stations join the random backoff procedure, and the winner station accesses the channel.

The MAC for Wavelet OFDM PHY, however, uses a gap state before the backoff procedure, as shown in Fig. 2, where the duration of the gap state is set according to the priority level of the transmission. The duration of the gap state for a station with the highest user priority traffic (UP = 7) is zero; hence, the station starts the backoff procedure immediately after the CIFS period (see the top sequence in Fig. 2b). The gap state duration of a station with lower priority traffic is greater than the backoff duration of the higher priority levels. Therefore, a station starts its backoff countdown only when it detects no high priority transmission in the channel, as shown in the bottom sequence in Fig. 2b.

Contention-Based Access in G.hn — In G.hn, contention-based access is offered within a TXOP, namely a shared TXOP (STXOP), as shown in Fig. 3. Unlike the contention in IEEE 1901, wherein every station can contend for channel access, the DM operating under the G.hn standard selects a single station or a group of stations to contend for channel access in a STXOP. Hence, the collision probability in G.hn is expected to be lower than that in IEEE 1901-based networks. An STXOP is divided into a grid of time slots (TS): a contention-free TS (CFTS), a contention-based TS (CBTS), and a registration CBTS (RCBTS), which is used for registration only. An STXOP can include only CFTS, only CBTS, or both. An STXOP that is composed only of CBTSs is denoted as CBTXOP. Nodes may also register in other CBTSs than RCBTS. The DM assigns a TS to one or a group of stations using a TS assignment rule. Stations sharing an STXOP track the passage of time slots on the line and transmit only within their assigned TS using carrier sensing. Because a DM may assign the same TS to a number of stations, a station performs a backoff procedure upon receiving a shared TS prior to channel access to avoid collisions. When a station obtains a TS designated for access, it transmits an INUSE signal in the TS and starts the priority resolution procedure. The G.hn adopts the PRS concept of the FFT type in IEEE 1901; i.e., a station resolves the priority level before the backoff state. Other stations that are not designated to access the slot stop TS counting as soon as they receive the INUSE signal. After the priority resolution, stations contend for medium access using the backoff procedure.

Contention-Free Access

To provide QoS-guaranteed services such as high-resolution audio/video streams or VoIP, both standards offer TDMA-based contention-free channel access. IEEE 1901 uses a Beacon-based periodic channel access mechanism, whereas ITU-T G.hn uses MAP-based scheduling. IEEE 1901 stations (STAs) can use the power line communication

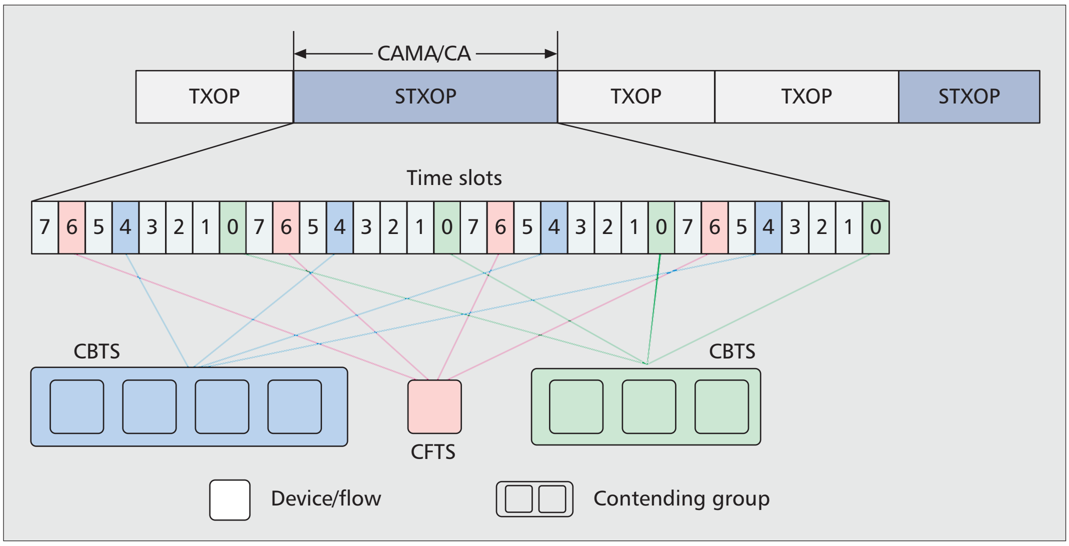


Figure 3. *Contention-based access in a STXOP in the G.hn standard.*

Persistent access is useful when the channel conditions of a power line are poor and a beacon/MAP is lost in one or more consecutive cycles. In this situation, persistent access is able to prevent an interruption in transmission.

medium exclusively by implementing a TDMA mechanism during the contention-free period (CFP). The BM assigns a bandwidth for each link, but not for the STA. STAs can guarantee the time for frame transmission for each link to use the TDMA mechanism, thus the TDMA is usually used for audio/video streams and VoIP streams.

The admission-control procedure plays a critical role in TDMA. Admission control deals with long-term allocation of network resources. All contention-free sessions go through an admission-control procedure during the session establishment. The scheduling procedure deals with the short-term (on a Beacon-period basis) allocation of network resources.

In *persistent access*, a flow gets channel access at the given reserved TDMA unit (or CFTXOP in G.hn) in subsequent MAC cycles. Persistent access is useful when the channel conditions of a power line are poor and a beacon/MAP is lost in one or more consecutive cycles. In this situation, persistent access is able to prevent an interruption in transmission.

QOS SUPPORT

PLC MAC schemes provide QoS traffic scheduling for the intra-BSS QoS frame transfers used by delay constraint services by assigning timeslots for transmissions using either CSMA-based or TDMA-based access protocols. A possible implementation is as follows. In contention-based channel access, user priority (UP) affects the AC/MA/gap state duration. In contention-free channel access, traffic specification (TSpec) affects bandwidth management.

IEEE 1901 provides two basic mechanisms to support applications with QoS requirements. The first mechanism delivers traffic based on differentiated UPs. This differentiation is achieved by varying the following for different UP values:

- Amount of time a station senses the channel to be idle before backoff or transmission.
- The length of the contention window to be used for the backoff.
- The duration during which a STA may transmit after it acquires the channel.

The second mechanism allows for the reservation of transmission slots (TXOPs) in the CFP. A non-BM station, based on the requirements placed by the upper layer, requests TXOPs, both for its own transmissions as well as for transmissions from the BM to itself. The BM either accepts or rejects the request, based on an admission control policy. When the request is accepted, the BM schedules TXOPs for both the BM and the non-BM station.

The QoS prioritization facility in IEEE 1901 supports eight UP levels with values identical to the IEEE 802.1D priority tags listed in Table 4. A station with low-delay/low-jitter traffic requirements is tagged as UP > 3 and is allowed exclusive use of the medium during the CFP. All streams requiring transmission in the CFP are managed by a QoS controller. A station with these priority levels can also gain access during a CP. The priority resolution during a CP in the IEEE 1901 Wavelet MAC provides direct support for the eight UP levels for both contention-based and contention-free access. The FFT-MAC in IEEE 1901 and G.hn, however, offer four priority levels in the CP because of the 2-bit priority resolving slot (PRS). These standards follow the access category levels used in IEEE 802.11 (listed in Table 5). However, some related works show that the priority based mechanism may not work well, and its contention window size value may be optimized based on the traffic load and the delay requirements [7].

SECURITY

According to the inherent characteristics of power lines, power line channels are considered to be shared media for communication. Thus, a PLC network may be affected by both external and internal attacks. An external threat implies an attacker capable of eavesdropping on transmissions and sending frames within the network, but with out-of-network access credentials. Internal threats refer to those from legitimate users

User Priority	Application Class
7	Network Control — to maintain the network infrastructure.
6	"Voice" — less than 10 ms delay, and hence maximum jitter
5	"Video" or "Audio" — less than 100 ms delay.
4	Controlled Load —for bandwidth reservation per flow
3	Excellent Effort — important best effort
0	Best Effort — LAN traffic as we know it today
1,2	Background — bulk transfers and other activities that are permitted on the network, but that should not impact the use of the network by other users and applications.

Table 4. *Application classes for user class mappings.*

Priority Resolution		FFT MAC	G.hn	Description
PRS0	PRS1			
0	0	AC0	MA0	Lower Priority
0	1	AC1	MA1	
1	0	AC2	MA2	
1	1	AC3	MA3	Higher Priority

Table 5. *Channel access priority in 1901 FFT MAC and ITU-T G.hn.*

of the network who have an illegitimate interest in the communications of another user or access to a specific network client. In the case of hidden stations, communications between two particular stations may pass through a relay station, causing a man-in-the-middle threat [1].

The PLC MAC standards do not provide any specialized security protocol for access control. The IEEE 1901 standard uses the security framework in IEEE 802.1X, along with the Cipher-block chaining Message authentication code Protocol (CCMP; counter mode with cipher-block chaining MAC). The IEEE 1901 WG was basically inspired by the 802.11i standard for security in wireless networks, which is based on the robust security network association (RSNA) concept found in 802.1X and CCMP. The RSNA defines a number of security features:

- Enhanced authentication mechanisms for stations.
- A set of key management algorithms.
- Cryptographic key establishment.
- An enhanced data cryptographic encapsulation mechanism called Counter mode with CCMP.

Security services in IEEE 1901 are provided by the authentication service and the CCMP mechanism. The scope of the security services provided is limited to station-to-station data exchanges. The data confidentiality service offered by an IEEE 1901 CCMP implementation is protection of the MSDU. CCMP combines counter mode (CTR) for data confidentiality, Cipher Block Chaining Message Authentication Code (CBC-MAC) for data integrity, and it also provides message integrity code (MIC) protection for both the frame body and almost the entire header in a MAC frame. The security services provided by CCMP in IEEE 1901 are as follows:

- Data Confidentiality.
- Authentication.
- Access control in conjunction with layer management.

For the purposes of the standard, CCMP is viewed as a logical service located within the MAC sublayer, as shown in the reference model in Fig. 4. The security mechanisms span the MAC layer through the station management services.

During the authentication exchange, both parties exchange authentication information. The MAC sublayer security services provided by CCMP rely on information from nonlayer-2 management or system entities. Management entities communicate information to CCMP through a set of MAC sublayer management entity (MLME) interfaces and management information base (MIB) attributes.

To deal with external threats, G.hn defines an authentication procedure based on the Diffie-Hellman algorithm and the Counter with Cipher Block Chaining-Message Authentication Code algorithm (CCM), which uses Advanced Encryption Standard (AES) –128 [8]. G.hn defines pairwise security against internal threats: a unique encryption key is assigned to each pair of communicating nodes and is unknown to all other nodes. Pairwise security maintains confidentiality between users within the network and builds another layer of protection against an intruder who has broken through the network admission control. The expected grade of security in G.hn is the same as or stronger than that defined in the most recent specification for WLAN IEEE 802.11n.

BURST MODE OPERATION

To operate effectively in an environment with periodic impulse noise, the power line MAC protocols allow a station to transmit multiple packets of the same priority level one after another in a single access. This feature is usually referred to as the burst transmission mode, and both IEEE 1901 and ITU-T G.hn support burst mode transmission in both CFP and CP. Figure 5 shows typical bursting modes that are supported by IEEE 1901. In general, the MAC header of each frame contains an indication whether it is the final frame of the burst transmission or not, and the receiver tunes itself based on the indication. Further, the first frame may also indicate the ACK policy for the bursting: immediate ACK or selective ACK (SACK). In the immediate ACK policy, the receiver replies with an ACK frame upon receiving every frame. However, in the SACK policy, when the destination station receives any frame that is not the final one, the receiver does not reply with an ACK frame and stores the corresponding ACK information. This process continues until the last frame is received. Upon receiving the last frame in the burst, the receiver transmits an ACK frame, acknowledging all frames belonging to that burst.

In case of burst transmission in a CFP, the bursting is allowed only until the end of the current contention-free allocation (or TXOP). In G.hn, contention-based access occurs within a STXOP, and hence, a station can use bursting until the end of STXOP. However, in the case of bursting during a CP in IEEE 1901, the standard assigns a maximum duration of a burst (for example, 5 ms for FFT MAC).

Bi-directional Bursting — Both IEEE 1901 (using FFT-PHY) and G.hn offer bi-directional bursting between two communicating stations of a flow. Using this mechanism, the destination may request a bi-directional flow in the burst through the ACK frame header to the source station. The source station may accept or reject the request. When accepted, the source station sends the grant information through the header in the immediate next frame and the receiver transmits in the reverse direction after transmitting the ACK frame. The source station adjusts the burst interframe space to accommodate the reverse flow.

Coexistence of IEEE 1901 and ITU-T G.hn Products

IEEE 1901 and ITU-T G.cx (or G.9972) specify the same coexistence mechanism — ISP (Inter System Protocol) — to accommodate up to four power line systems: IEEE 1901 in-home OFDM, IEEE 1901 in-home Wavelet, ITU-T G.hn, and the one access system [9]. In G.cx, ISP is a self-contained independent coexistence recommendation, whereas in IEEE 1901 ISP is contained mostly (not entirely) in one chapter [10]. The first coexistence mechanism developed was Inter-PHY Protocol (IPP) in IEEE 1901. This was necessary because 1901 lacks interoperability as it specifies two non-interoperable PHY/MACs. Recognizing the importance of coexisting with G.hn, IPP was modified to allow interoperability with G.hn, and IPP became ISP in IEEE 1901. To transfer the ISP to ITU, a new project called G.cx was created. G.cx gained consent in October 2009 and became recommendation G.9972. G.9972 is currently in the Last Call Comment (LCC) resolution period, and is expected to be approved by June 2010.

G.9972 will cover resource allocation for Frequency Domain Multiplex (FDM) and Time Domain Multiplex (TDM), start-up and re-synchronization procedures, power control, and management functionalities. Because G.9972 and IEEE 1901 have gone through two different comment resolution schemes, the two specifications are now diverging and may no longer be compatible. Now both G.cx and 1901 working groups are actively working on aligning the specifications during the comment resolution phase.

ITU-T G.hn supports not only power lines but also pre-existing wire lines such as coaxial

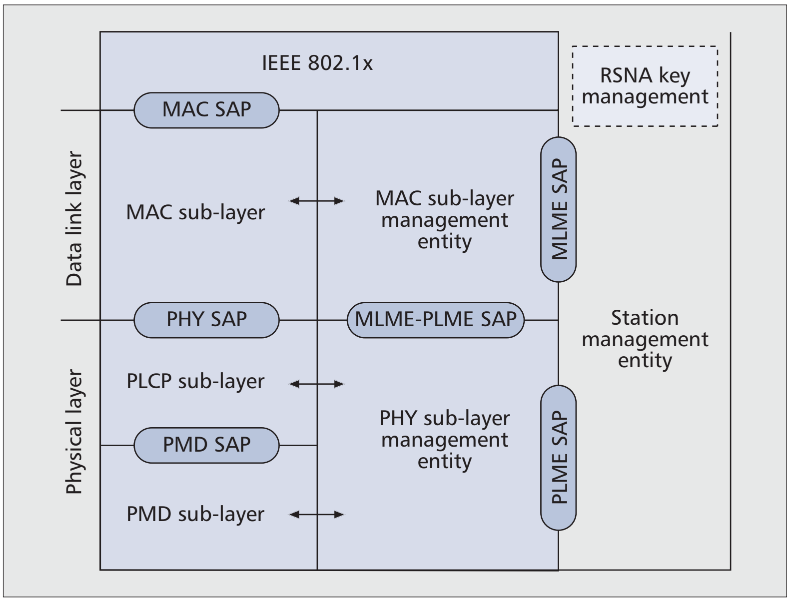


Figure 4. *Security reference model of IEEE 1901.*

cables and phone lines, and it extends its domain into the in-home Smart Grid area by enabling G.hnem. Thus, G.hn is designed for high/low-rate broadband/narrowband in-home services. IEEE 1901 is designed to support broadband access and high-rate broadband in-home services. Nevertheless, there is still no single standard to cover all application areas, and a single co-existence standard technology is a key focus in communications research. Several co-existence scenarios can be expected, such as between ITU-T G.hn and IEEE 1901 in the same in-home area, between IEEE 1901 broadband access and ITU-T G.hn broadband in-home, and between ITU-T G.hnem and IEEE 1901 broadband access in-home.

Summary

In this article, we tried to point out the available features and technologies in the PLC standards IEEE 1901 and ITU-T G.hn. As both of these standards are still in the development stage, it is difficult to make direct comparisons and comment on the compatibility issues. The IEEE 1901 working group has decided to make a standard based on the already existing technologies with additional improvement, while ITU-T G.hn decided to make a new standard in a clean-state approach. However, based on an analysis of the current stage and defined targets, it is clear that both of these standards offer similar features such as the master-slave architecture, contention-based and contention-free medium access schemes, transmission modes, QoS, and security services, among others. Therefore, many commonalities of the two standards will help facilitate dual mode implementation, which can help solve non-interoperability issues. However, an eventual solution to the problem is for the two standards bodies to agree on recommending a single coexistence standard. Thus, a single coexistence standard will solve non-interoperability issues between multiple technologies such as

An eventual solution to the problem is for the two standards bodies to agree on recommending a single coexistence standard. Thus, a single coexistence standard will solve non-interoperability issues between multiple technologies such as ITU-T G.hn, IEEE 1901, other forthcoming and legacy technologies.

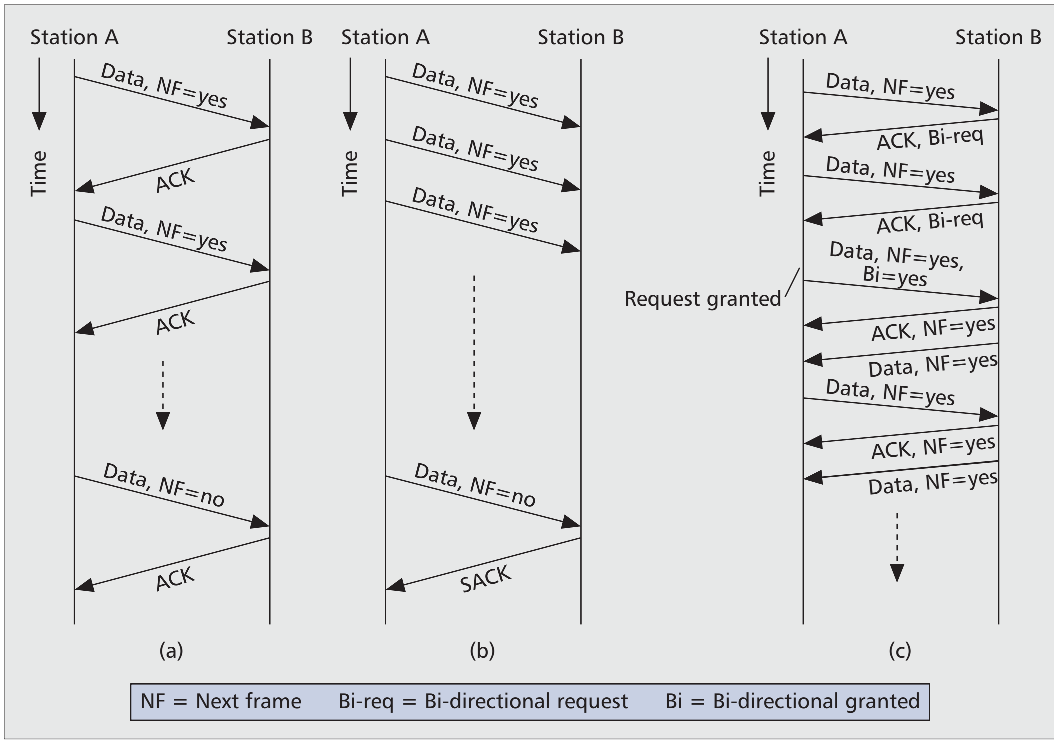


Figure 5. *Typical Burst Mode scenarios for IEEE 1901 devices: a) uni-directional bursting with an immediate ACK policy; b) uni-directional bursting with a selective ACK policy; and c) bi-directional bursting.*

ITU-T G.hn, IEEE 1901, other forthcoming and legacy technologies.

Acknowledgments

This work was supported by a grant from the Kyung Hee University in 2010 (KHU-2010372). Dr. C. S. Hong is the corresponding author.

References

[1] S. Galli and O. Logvinov, "Recent Developments in the Standardization of Power Line Communications within the IEEE," *IEEE Commun. Mag.*, vol. 46, no. 7, July 2008, pp. 64–71.

[2] V. Oksman and S. Galli, "G.hn: The New ITU-T Home Networking Standard," *IEEE Commun. Mag.*, vol. 47, no. 10, Oct. 2009, pp. 138–45.

[3] E. Biglieri, "Coding and Modulation for an Horrible Channel," *IEEE Commun. Mag.*, vol. 41, no. 5, May 2003, pp. 92–98.

[4] The HomeGrid Forum; http://www.homegridforum.org.

[5] ITU-T Recommendation G.9960, "Unified High-Speed Wire-Line based Home Networking Transceivers — Foundation," Jan. 2010.

[6] ITU-T Recommendation G.9961, "Data Link Layer (DLL) for Unified High-Speed Wire-Line based Home Networking Transceivers," Jan. 2010.

[7] M. Y. Chung *et al.*, "Performance Analysis of HomePlug 1.0 MAC With CSMA/CA," *IEEE JSAC*, vol. 24, no. 7, July 2006, pp. 1411–20.

[8] IEEE P1901/D2.01, "IEEE Draft Standard for Broadband over Power Line Networks: Medium Access Control and Physical Layer Specifications," Jan. 2010.

[9] S. Galli, A. Kurobe, and M. Ohura, "The Inter-PHY Protocol (IPP): A simple Co-existence Protocol," *IEEE ISPLC*, Dresden, Germany, Mar. 29–Apr. 1, 2009.

[10] ITU-T Recommendation G.9972, "Coexistence Mechanism for Wireline Home Networking Transceivers, " Jan. 2010.

Biographies

Mustafizur Rahman (mustafiz1952@yahoo.com) received the B.Sc. (Hons.) degree in applied physics and electronics and the M.Sc. degree in computer science from the University of Dhaka, Bangladesh, in 1997 and 1999, respectively. He joined the University of Dhaka in 1997, where currently he is an associate professor with the Department of Computer Science and Engineering. He received the Ph.D. degree in Computer Engineering in 2009 at Kyung Hee University, South Korea. His research interests include Multiple Access Protocols, Routing and Network Security in Wireless Mobile Ad Hoc and Sensor Networks.

Choong Seon Hong (cshong@khu.ac.kr) received his B.S. and M.S. degrees in electronic engineering from Kyung Hee University, Seoul, Korea, in 1983, 1985, respectively. In 1988 he joined KT, where he worked on Broadband Networks as a member of the technical staff. From Sept. 1993, he joined Keio University, Japan. He received the Ph.D. degree at Keio University in March 1997. He had worked for the Telecommunications Network Lab. KT as a senior member of technical staff and as a director of the networking research team until August 1999. Since September 1999, he has worked as a professor of the department of computer engineering, Kyung Hee University. He has served as a Program Committee Member and an Organizing Committee Member for International conferences such as NOMS, IM, APNOMS, E2EMON, CCNC, ADSN, ICPP, DIM, WISA, BcN, TINA, SAINT and ICOIN. His research interests include Ad hoc Networks, Power Line Communications, Network Management and Future Internet. He is a member of IEEE, ACM, IEICE. IPSJ, KIISE, KICS and KIPS.

Sungwon Lee (drsungwon@khu.ac.kr) received the Ph.D. degree from Kyung Hee University, Korea. He is a professor of the Computer Engineering Departments at Kyung Hee University, Korea. He was a senior engineer of Telecommunications and Networks Division at Samsung Electronics Inc. from 1999 to 2008. He is a editor of the Journal of Korean Institute of Information Scientists and Engineers: Computing Practices and Letters. His research interests include mobile broadband networks and communication protocols. He is a member of IEEE, KIISE, KICS and KIPS.

Jae-Jo Lee (jjlee@keri.re.kr) is a senior researcher in Power Telecommunication Network Research Group of KERI(Korea Electrotechnology Research Institute). He received his B.S. degree in physics and M.S. degree in electronic engineering from Kyung Hee University, Seoul, Korea, in 1990, 1992, respectively. In 1992 he joined KERI, where he worked on power telecommunication networks as a researcher. He received the Ph.D. degree in electronics engineering from Kyung Hee University in 2005. From 1999 to 2004, he had worked for the development of power line communication network technology as a senior member of technical staff. Since October 2005, he has worked as a project manager of the ubiquitous power line communication networks development project. His research interests include Power Line Communications, Home Networking and Network Management System. He is a member of Korean Institute of Electrical Engineers (KIEE) and Korean Institute of Communication Sciences (KICS).

ABDUR RAZZAQUE [M] (m_a_razzaque@yahoo.com) received the B.S. degree in Applied Physics and Electronics and M.S. degree in Computer Science from the University of Dhaka, Bangladesh in 1997 and 1999 respectively. He obtained his Ph.D. degree in Computer Engineering from Kyung Hee University, South Korea in August, 2009. He had worked as an Assistant Professor in the department of Computer Science and Information Technology of Islamic University of Technology (IUT), Gazipur, Bangladesh. He is an Assistant Professor (now on leave) in the Department of Computer Science and Engineering, University of Dhaka, Bangladesh. He is now working as a post doctoral fellow in the Institute of Multimedia Technology, Kyung Hee University, South Korea. His research interests include the area of modeling, analysis and optimization of communication protocols and architectures for wireless sensor networks, distributed systems, etc.

JIN HYUK KIM (mukjim@gmail.com) received the B.S. and M.S. degrees in computer engineering from Kyung Hee University, South Korea, in 2009 and 2011, respectively. His research interests include MAC protocol of Powerline Communication and Wireless Sensor Networks, especially on the coexistence of several PLC protocols.

ADVERTISERS' INDEX

ADVERTISING SALES OFFICES

Closing date for space reservation:
1st of the month prior to issue date

NATIONAL SALES OFFICE
Eric L. Levine
Advertising Sales Manager
IEEE Communications Magazine
3 Park Avenue, 17th Floor
New York, NY 10016
Tel: (212) 705-8920
Fax: (212) 705-8999
Email: e.levine@comsoc.org

SOUTHERN CALIFORNIA
Patrick Jagendorf
7202 S. Marina Pacifica Drive
Long Beach, CA 90803
Tel: (562) 795-9134
Fax: (562) 598-8242
Email: patrick@roymcdonald.com

NORTHERN CALIFORNIA
George Roman
Roy McDonald Assoc.
2336 Harrison Street
Oakland, CA 94612
Tel: (702) 515-7247
Fax: (702) 515-7248
Cell: (702) 280-1158
Email: george@roymcdonald.com

SOUTHEAST
Scott Rickles
560 Jacaranda Court
Alpharetta, GA 30022
Tel: (770) 664-4567
Fax: (770) 740-1399
Email: srickles@aol.com

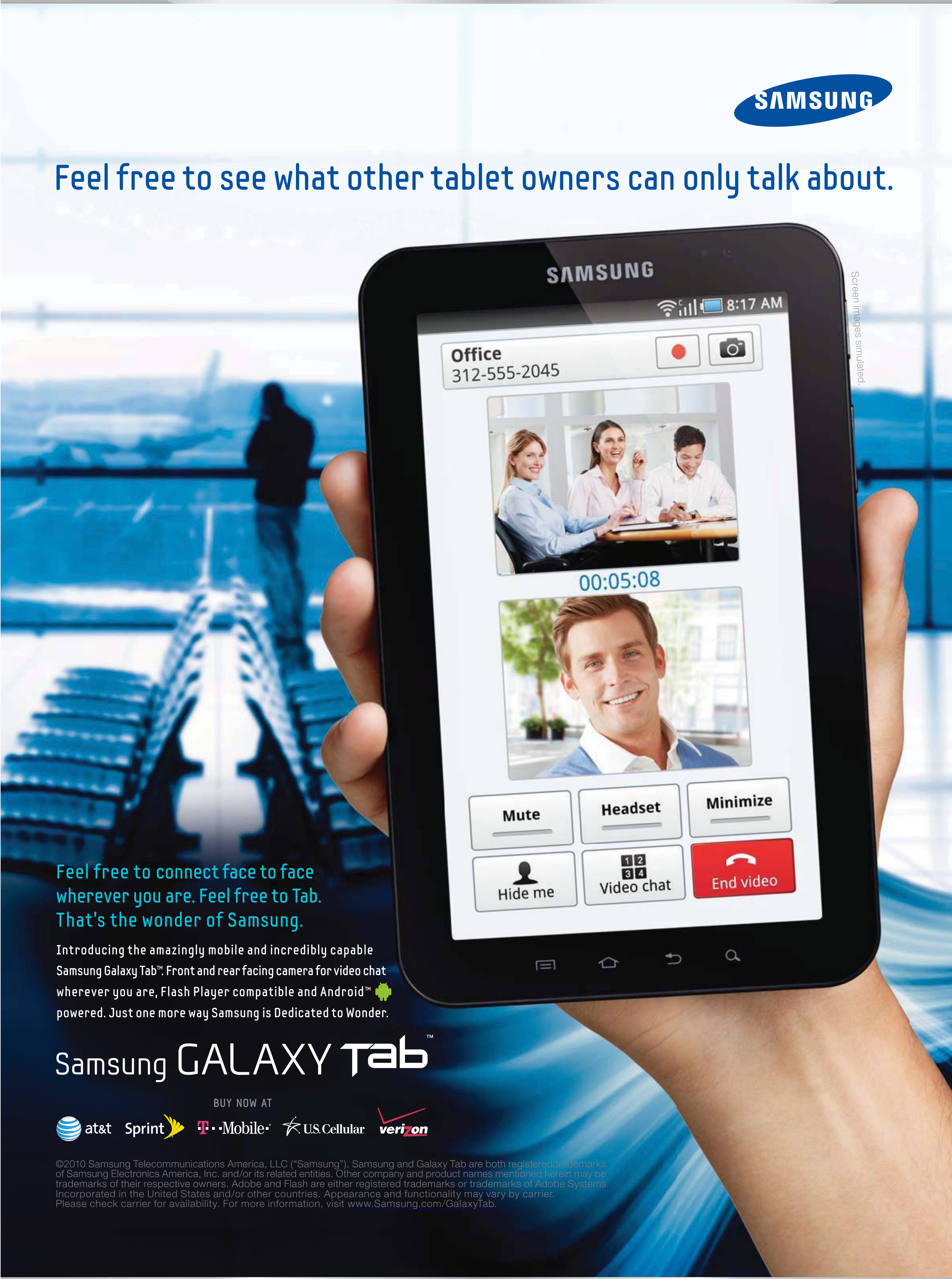
EUROPE
Rachel DiSanto
Huson International Media
Cambridge House, Gogmore Lane
Chertsey, Surrey, KT16 9AP
ENGLAND
Tel: +44 1428608150
Fax: +44 1 1932564998
Email: rachel.disanto@husonmedia.com

JULY 2011 EDITORIAL PREVIEW

Future of the Internet
ITU Standards in Communications
Network and Service Management

AUGUST 2011 EDITORIAL PREVIEW

Modeling & Analysis of Wireless Networks Utilizing Game Theory
Design and Implementation
Energy Efficiency in Communications


SAMSUNG
Feel free to see what other tablet owners can only talk about.
SAMSUNG
8:17 AM
Office
312-555-2045
00:05:08
Mute
Headset
Minimize
Hide me
Video chat
End video
Screen images simulated.
Feel free to connect face to face wherever you are. Feel free to Tab. That's the wonder of Samsung.
Introducing the amazingly mobile and incredibly capable Samsung Galaxy Tab™. Front and rear facing camera for video chat wherever you are, Flash Player compatible and Android™ powered. Just one more way Samsung is Dedicated to Wonder.
Samsung GALAXY Tab™
BUY NOW AT
at&t Sprint T··Mobile· U.S. Cellular verizon
©2010 Samsung Telecommunications America, LLC ("Samsung"). Samsung and Galaxy Tab are both registered trademarks of Samsung Electronics America, Inc. and/or its related entities. Other company and product names mentioned herein may be trademarks of their respective owners. Adobe and Flash are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States and/or other countries. Appearance and functionality may vary by carrier. Please check carrier for availability. For more information, visit www.Samsung.com/GalaxyTab.