IntechOpen

Broadband Communications Networks

Recent Advances and Lessons from Practice

Edited by Abdelfatteh Haidine and Abdelhak Aqqal

BROADBAND COMMUNICATIONS NETWORKS - RECENT ADVANCES AND LESSONS FROM PRACTICE

Edited by **Abdelfatteh Haidine**
and **Abdelhak Aqqal**

Broadband Communications Networks - Recent Advances and Lessons from Practice
http://dx.doi.org/10.5772/intechopen.69590
Edited by Abdelfatteh Haidine and Abdelhak Aqqal

Contributors

Ulrich H.P. Fischer-Hirchert, Matthias Haupt, Mladen Joncic, Stefanie Haupt, Sebastian Höll, Pavel Loskot, Salman Al-Shehri, Tolga Numanoğlu, Mehmet Mert, Vjaceslavs Bobrovs, Guntis Ancans, Arturo Peralta, Roberto Hincapie, Sheng Xu, Milan Aleksandar Ivanović, Nguyen Xuan Tien, Jong Myung Rhee, Milda Tamošiūnaitė, Stasys Tamošiūnas, Mindaugas Žilinskas, Francesco Marcuzzi, Andrea Tonello, Vipin Balyan, Ben Groenewald, Mario Ligwa, Raad Alhumaima, Hamzeh Khalili, David Rincon, Sebastia Sallent, Jose Ramon, Sonia Gul, Jairo Gutierrez, Krzysztof Borzycki, Eneko Atxutegi, Jose Oscar Fajardo, Fidel Liberal, Md Maruf Ahamed, Saleh Faruque, Ales Svigelj, Melisa Junuzovic, Hui Yang, Liping Ai, Hermann J. Helgert, Hyun Jung Lee, Myungho Kim, Abdelfatteh Haidine, Abdelhak Aqqal

First published in London, United Kingdom, 2018 by IntechOpen
eBook (PDF) Published by IntechOpen, 2019
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, The Shard, 25th floor, 32 London Bridge Street
London, SE19SG – United Kingdom
Printed in Croatia

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Broadband Communications Networks - Recent Advances and Lessons from Practice
Edited by Abdelfatteh Haidine and Abdelhak Aqqal

p. cm.

Print ISBN 978-1-78923-742-9

Online ISBN 978-1-78923-743-6

eBook (PDF) ISBN 978-1-83881-399-4

Meet the editors

Dr.–Ing. Abdelfatteh Haidine received his PhD from the *Technische Universität Dresden* in Germany with a focus on the planning and optimisation of telecommunications networks. He has worked as a consultant and manager of big companies (KEMA Consulting GmbH, Accenture PLC) for the deployment of smart metering systems and smart grid applications. Currently, he is an assistant professor with "Laboratory of Information Technologies" at the National School of Applied Sciences, Morocco.

His research interests include different issues related to machine-to-machine and Internet-of-Things communications, networking technologies for smart city and smart grid applications, etc. This covers LPWA networks and their techno-economical aspects. In recent years, he has been dealing with application of combinatorial optimization as well as the game theory paradigm in network planning/migration and resource allocation in broadband mobile networks.

Dr. Abdelhak Aqqal received a joint PhD degree in Sciences and Technologies of Information and Communication (STIC) from Darmstadt University of Technology (Germany) and Chouaib Doukkali University (Morocco) under the DAAD Sandwich Program. In 2010, he joined the Department of Telecommunications, Networks and Computer Science at the National School of Applied Sciences of El Jadida (Morocco), teaching engineering students in the area of telecommunications and networking with a focus on cloud computing and IT for business.

His research interests include ICT for smart cities and innovative approaches using green and smart technologies. He also has extensive experience and research in E-learning and technology-enhanced teaching in higher education. He has been involved in various research projects and is the author of several research studies published in national and international journals, conference proceedings and book chapters.

Contents

Preface

Nowadays, the Internet plays a vital role in our lives. It is always changing, and so is the world. As the base of the digital transformation that has made our world a global village, the Internet is currently one of the most effective media that is shifting to reach into all areas in today's society, including transportation, medicine and healthcare, industries, education, business administration and finance, among others. It has a distinguished record for bringing together many emerging technologies from a variety of different disciplines and facilitating innovations that are practically changing our daily lifestyle. Topics about web services and IT operations that use the phrase 'Internet technology' become front page news. As we move into the next decade, the future holds many fast-moving technologies-from Internet of things (IoT), cloud solutions, automation and artificial intelligence (AI), big data, machine learning, virtual reality, 5G and mobile technologies to connected vehicles, connected homes, and smart cities. All of these technologies are highly dependent on Internet connectivity and broadband communications. Consequently, recent surveys and research looking at digital mutations of our society and Internet-driven business trends by 2020 and beyond have revealed an exponential growth in information exchange, feature-rich communications, and large/integrated traffic over a huge number of fixed and mobile end-devices. The demand for mobile and faster Internet connectivity is on the rise as the voice, video, and data continue to converge to speed up business operations and to improve every aspect of human life. As a result, the broadband communication networks, which connect everything on the Internet, are considered now as a complete ecosystem routing all Internet traffic. In fact, just few years ago, they were well-known bottlenecks when it comes to fixed Internet connections. But in the age of mobile broadband, the information superhighway delivers Internet data faster and more flexibly than ever before.

With all this in mind, drawing on research experiences and lessons from over the globe, this book explores the latest research and developments in the broadband communication networks associated with multiservice modes and architectures in support of many emerging paradigms/applications of global Internet from the traditional architecture to the incorporation of smart applications. Consequently, this book may be used as a reference book on broadband communication networks as well as on the practical uses of wired/wireless broadband communications. Committed to bridging the gap between theory and practice, this book is also a concise guide for students and readers interested in studying Internet connectivity, mobile/optical broadband networks, and concepts/applications of telecommunications engineering.

Overall, this book is comprised of 21 chapters authored by specialists in the field. After an introductory chapter on the background of broadband communication networks, the remaining chapters are organized into three main parts: "Wireless/Mobile Broadband Network", "Optical Networks", and "Practical Aspects of Broadband Networking". The first part discusses some key challenges, uses cases from the field of wireless/mobile broadband networks and focuses on the following emerging areas/questions: the fundamental aspects, the evolution and techniques that increase the overall performance of the broadband communications over wireless access networks (Chapter 2 and Chapters 7 to 10). Moreover, this first part also includes a discussion on requirements, challenges, and emerging technologies of the future 5th generation (5G) cellular networks as the most recent technologies of the mobile broadband network (Chapters 3, 4, and 6). The fifth chapter addresses how power line communications (PLC) and small-cell network technologies can be brought together in a unified model to foster future small-cell technology.

The second part provides a review and technical information about optical broadband interconnection and the design of optical networks as one of the high-speed and fast network technologies used to adapt the needs of large-scale data in Internet usage, with the advantages of large capacity, high bandwidth, and high efficiency (Chapters 11 to 14). Finally, the third part provides some of the latest research and practical aspects of broadband networking in the context of the digital economies (Chapter 15), with respect to economic interests and social problems in realization of broadband network (Chapter 18), in the areas of smart connected city (Chapter 16), and FiWi Networks (Chapter 17). The third part highlights innovative strategies and techniques to enhance the performance of cutting-edge technologies in mobile broadband networks for the case of fast moving trains (Chapter 20), in HSR networks (Chapter 19) and in future cloud mobile broadband networks (Chapter 21).

To conclude, the editors are grateful to all colleagues who authored the chapters of this book and contributed with valuable references and interesting results related to their current research and applications. We are also grateful to the members of the support team at IntechOpen for their help and professionalism. Special thanks are due to the reviewers for their willingness to review the chapters and provide useful feedback to the authors. We would particularly like to thank our colleague Asmaa El Hannani for her thorough feedback in order to improve the quality of the publication.

Abdelfatteh Haidine and Abdelhak Aqqal
Laboratory of Information Technologies
National School of Applied Sciences
Chouaib Doukkali University, Morocco

Chapter 1

Introductory Chapter: Next Generation of Broadband Networks as Core for the Future Internet Societies

Abdelfatteh Haidine and Abdelhak Aqqal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.80508

1. Introduction: evolution of the needs for "broadband"

The Internet traffic is an ongoing explosive increasing from year to year, so that the annual global IP traffic surpassed the zettabytes threshold in 2016. Furthermore, it is predicted that the overall IP traffic will grow at a compound annual growth rate (CAGR) of 24% from 2016 to 2021 [1]. Different factors are alimenting this growth, such as the increasing number of connected devices from different types, as illustrated by **Figure 1**. This continuous growth has a big impact on different level of networking, such as the wide area network, metro (metropolitan) network, access networks and the home (in-house/in-home) networks. Along the evolution of telecom networks, the access networks were always the "weak point" of the infrastructure and therefore referred to as "the bottleneck". Consequently, one of the first challenges in the era of Internet is the realisation of high-speed "broadband access networks". Basically, the concept and the term "Broadband Communications Networks" refers to any type of networks/access technologies used by Internet Service Providers (ISP) to provide a broadband Internet access for a multimedia content delivery/distribution according to technical considerations and requirements such as guaranteed Quality of Service (QoS). So many technologies were/are developed to support Broadband Communications in different connection forms such as Dial-up, Digital Subscriber Line (DSL), Optical fibre, cable, Broadband over Powerline, Mobile and wireless Internet access and satellite Internet. There are also quite a few other broadband options available for the Internet connection. Both wired and wireless broadband solutions exist, but none is universally considered optimal for all use cases and products configuration. In fact, the quantification of the meaning of "broadband access" in Mbps is evolving with the time depending on user-experiences in using or consuming the offered data services. With the apparition of the notion "broadband" access, systems had to guarantee at least a capacity of 2 Mbps. This was achieved in a first stage through the successful rollout of Asymmetric Digital

Subscriber Line (ADSL). With the increasing data traffic, the fastest version of DSL, Very high bit rate DSL—VDSL, has partially fulfilled the requirements of intensive data traffic by offering 25 Mbps, but it is limited by the weak coverage of its signal transmission, which does not go over 300 m. Currently, it is expected that the speed of broadband access will merely double by 2021, so that the global fixed broadband speed will reach 53 Mbps, up from 27.5 Mbps in 2016 [1].

The classical paradigm requiring the high speed for downlink connection, which was the reason for the success of ADSL, is not valid anymore for the current broadband access networks. For the cloud services, the end-users also need high-speed uplink to be able to upload their data to the cloud server(s). Furthermore, services and businesses based on big data are nourished by huge data volumes, which are collected in different forms (video, location information, sensing information, software logs, etc.). These two aspects concern both wired as well wireless network access.

The realisation of broadband for downlink and uplink can always be achieved by using optical fibres in the access domain guaranteeing very high speeds. However, it remains in most case economically unfeasible/unprofitable. Thus, the deployment of fibre in access networks, either as fibre-to-the-home (FTTH) or through fibre-to-the-building (FTTB) remains low and extremely different from one country to another, even in the industrial western countries. For example, according to most recent statistics from FTTH Council Europe, fibre access penetration in France does not go over 14.9% of households (with 3% through FTTH and remaining 11.9% through FTTB); while in Germany, this rate reaches 2.3% of households (with 1% FTTH plus 1.3% using FTTB) [2].

Beside the high speeds, the mobility is the second major key requirements of today's society, which makes from the "Broadband Mobile Internet" the headache for mobile operators. The age of Mobile Internet has started with the Universal Mobile Telecommunications System (UMTS), i.e. the third generation of mobile communications—3G. However, this start did not reach the expected success. Among the main causes for this start failure, two facts can be cited:

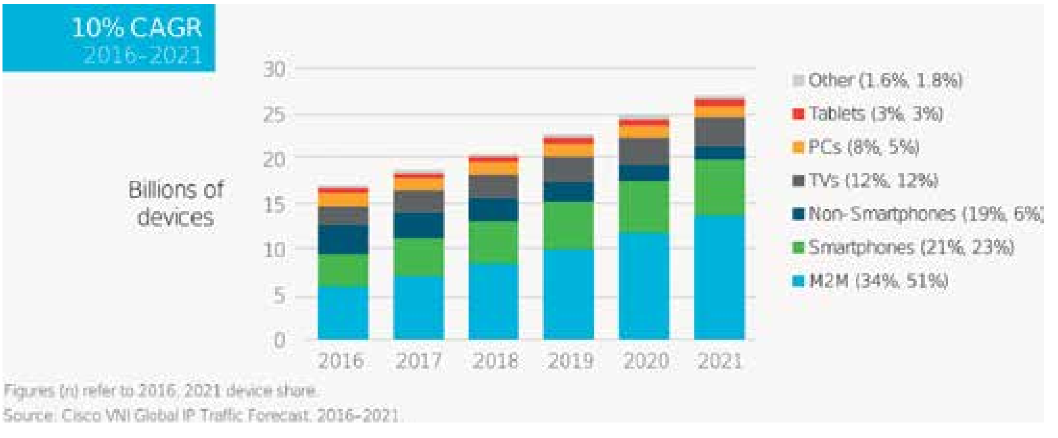

Figure 1. Global devices and connections [1].

Technically, UMTS was designed to offer up to 2 Mbps for the end-users; however, in the field only about 300 kbps were possible. From the economical aspect, the fees of spectrum licences have reached some astronomical levels that bring strong imbalance in the business model.

The failed targets of 3G were partially corrected through the new versions of UMTS, like the High-Speed packet Access (HSAP or 3G+, some references use 3.5G). However, the real breakthrough of mobile broadband has been brought by the fourth generation of mobile technology based on 3GPP Long Term Evolution (LTE) that allows capacities up to 300 Mbps. This speed increased significantly with the extension to LTE-Advanced and LTE-Advanced Pro (referred to as 4.5G). With the successful rollout of 4G around the world (except in some developing countries), the mobile broadband data traffic grew 70% between Q1 2016 and Q1 2017 and a further stronger increase in number of mobile/wireless connected device at their generated traffic is foreseen for the next years [3].

Mobile communications are experiencing a major revolution catalysed by the change in the way our today's society creates, shares and consumes information. While the preparation for massive deployment of 5G by 2020 is still ongoing, researchers are already talking about the "Beyond 5G" (B5G) mobile communications era. It is widely agreed that B5G network should achieve greater system capacity (<1000 times) in terms of data rate (terabits per second) and user density (the Internet-of-things and nano-things) [4]. Accordingly, three ways are considered to realise several orders of increase in throughput gain: the extreme densification of infrastructure, large quantities of new bandwidth and a large number of antennas, allowing a throughput gain in the spatial dimension.

2. Changing applications landscape

The telecommunications operators, especially the mobile services providers, have experienced one of the main mutation in the telecom markets, as the voice-dominated services are no longer making the main revenue for their business. In fact, in the period between 2006 and 2008 the operators' revenues were data dominated. At that period, the data traffic started its exponential growth, while the price stagnated accompanied with the economic recession. To balance their business model, operators started to converge their infrastructure to all-over-IP services, by the elimination of the circuit-switched infrastructure, which requires high OPEX and a wasting resources/bandwidth per excellence [5]. This was triggered by the adoption of 3GPP LTE as fourth generation mobile technology that transmits the voice service over IP packets (VoIP). The rollout of 4G has solved the main challenges that were facing the operators, but in the after-4G era, new challenges and requirements must be met such as more bandwidth, shorter latency and ultra-high reliable (UHR) communications. This is resulting either from new services/businesses or caused by a change in user or societal behaviour. As major pillars in new services or applications, we can cite the video, smart cities, big data and Mobile Big data (MBD), Internet-of-things (IoT), Car-to-X communications or Internet-of-vehicles (IoV), etc.

According to the mobile traffic analysis by application, the increased viewing of video on mobile devices, embedded video and emerging video formats will extremely drive data consumption; as stated in the recent Mobility Report [3]. As stated in this report, mobile video traffic is forecast to grow by around 50% annually through 2023 to account for 75% of all mobile data traffic. Social networking is also expected to grow—increasing by 34% annually over the next 6 years. However, its relative share of traffic will decline from 12% in 2017 to around 8% in 2023, as a result of the stronger growth of video traffic. The position of video in mobile data consumption is illustrated in **Figure 2**. Furthermore, streaming videos are available in different resolutions to increase the user experience and satisfaction by using more high-resolution videos, which will certainly affect data traffic consumption to a high degree. Accordingly, watching HD video (1080p) rather than video at a standard resolution (480p) typically increases the data traffic volume by around 4 times. An emerging trend with increased streaming of immersive video formats, such as 360-degree video, would also impact data traffic consumption. For example, a YouTube 360° video consumes 4–5 times as much bandwidth as a normal YouTube video at the same resolution [3]. In addition, the emergence of new applications and changes in consumer behaviour can shift the forecast relative traffic volumes.

The world knows currently a strong urbanisation of today's societies, where more people are living in cities than in rural areas. This generates a pressure on different resources, which are always available or generated in limited volumes and/or capacity, such as energy, water, road, spaces, transport means, hospitals, etc. Therefore, the decision-makers developed roadmaps for building smart cities, with the goal of an optimal generation and utilisation/consumption of these resources. The roadmaps differ from one country to another, but they all agree that the Information and Communications Technologies (ICTs) platforms will constitute the core of these smart cities. In some visions, smart cities consists in developing different smart domains, such as smart grid, smart parking, smart building/homes, smart education,

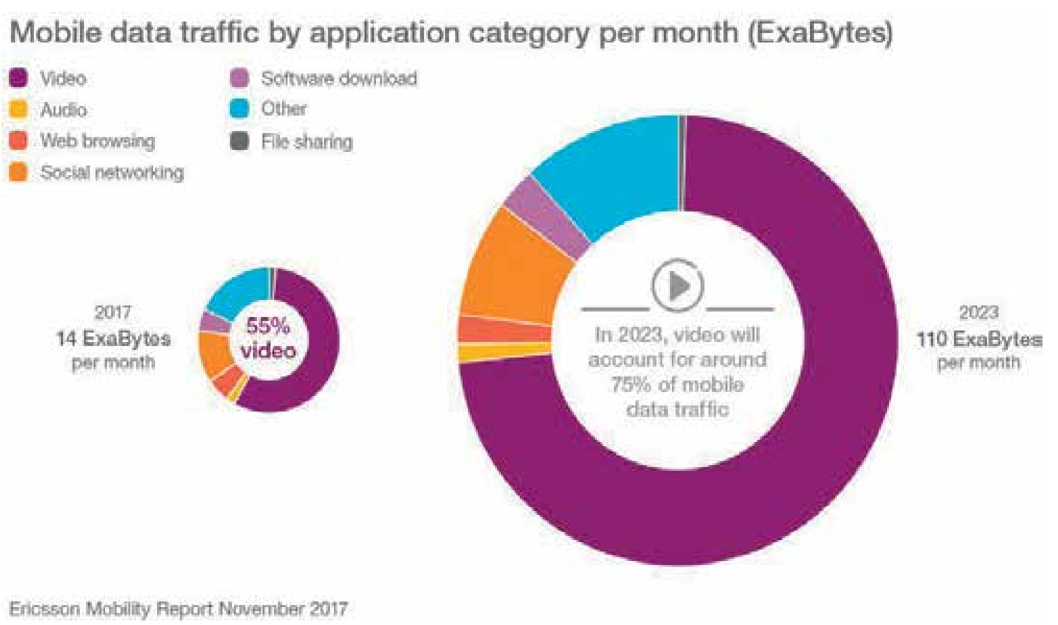

Figure 2. Increasing dominance of video content in mobile data [3].

http://dx.doi.org/10.5772/intechopen.80508

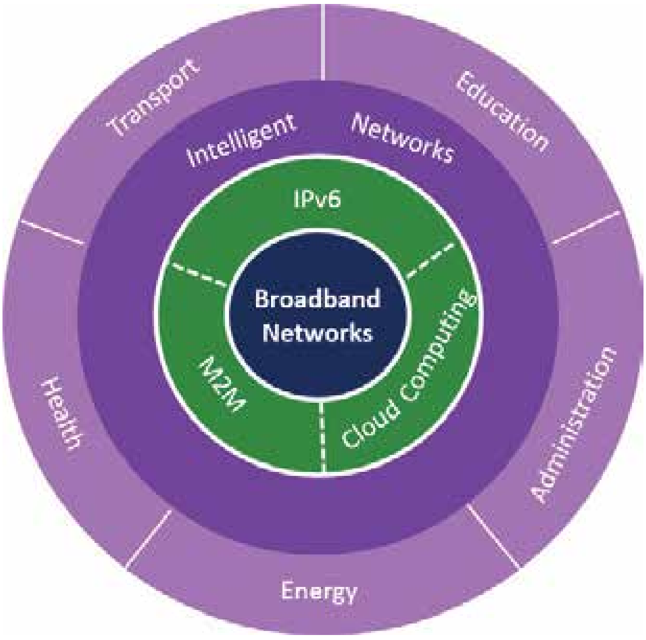

Figure 3. Broadband networks as core of future smart cities – German government's point of view [6].

e-administrations, etc. One of the visions is depicted in **Figure 3**, representing an early version of the smart cities from the German government's point of view, where broadband networks are the cornerstone in the ICT infrastructure [6]. A more complex and detailed recent version of this structure, which includes security and big data, can be found in [7].

3. Technologies to build the next generation of broadband networks

In spite of the above-cited challenges, the future mobile generation (5G) has found new emerging technologies to overcome all the boundaries, as explained in different chapters of this book. As key technologies of 5G, we can cite massive multiple-input multiple-output (massive-MIMO), network densification (or Ultra Network Densification), Cloud-based Radio Access Network (C-RAN), virtualisation, improved energy efficiency by energy-aware communication and energy harvesting, etc. However, mobile/wireless communications alone cannot be successful without the support from optical fibre, especially for the backhauling. This later is one of major parts of 5G as well as new spectrum parts, among others. The issues related to backhauling in 5G are discussed in Chapter 4 "5G Backhaul: Requirements, Challenges, and Emerging Technologies" and Chapter 5 "Radio Access Network Backhauling Using Power Line Communications", while spectrum issues are discussed in Chapter 3 "Spectrum usage for 5G mobile communication systems and electromagnetic compatibility with existent technologies".

One of the key success factors of next mobile broadband networks is the finding of new locations in the spectrum (Unlicensed, mmWave and THz). In a first step, operators have started to restructure their network to offload their traffic in the unlicensed bands. The 3GPP new technologies of Licenced Assisted Access (LAA) and LTE in unlicensed band (LTE-U) employ an unlicensed radio interface that operates over the 5 GHz unlicensed band to leverage the radio resources for operators' transmission [8, 9]. In a second step, to overcome the increase demand for wireless communication and scarcity of the spectrum bands, new land or parts of the spectrum are currently under exploration. Specifically, millimetre-wave (mmWave) communications systems (30–300 GHz) have been officially adopted in the fifth generation (5G) cellular systems, and several mmWave sub-bands have been allocated for licenced communications. However, the total consecutive available bandwidth for mmWave systems is still less than 10 GHz, which makes it difficult to go to the next step of the evolution and support data rates of the terabit per second. This pushes the researchers' community to explore the terahertz band (0.1–10 THz) communication, which is now envisioned as a key wireless technology to fulfil the future demands within 5G and beyond. Detailed discussion of this topic is given in Chapter 9 "Atmospheric attenuation of the terahertz wireless networks".

Optical fibre is forcing its way to go beyond the backhauling domain and FTTB, since so many countries have recognised the importance of high-speed broadband networks. The passive version of optical access network had a big part in lowering the deployment costs, besides the utilisation of software defined network (SN) technology. Discussion of the aspects related to "how to force the way for fibre" near to the end-user can be found in Part 2 of the book.

4. Conclusion

Broadband communication networks are currently a real need for today's society, and not more just a trend or luxury. In general, there is a lot of conferences and documentation in the literature about Broadband Communications Networks, but only few are interested in making it very transparent and accessible to others according to a broad perspective, cutting across wired/wireless technologies and Internet sectors. For example, what kind of Internet access do we need to have when moving to new Internet-driven applications of business and/or for specific computing contexts/smart environments? High Speed? Broadband? Wireless connection? Satellite? Optical Fibre? Mobile networks? What are the lessons we need to know about the practice in order to get fresh innovative ideas to speed up business operations and to improve every aspect of the human life? What are recent advances and notable emerging technologies, which will make us look beyond the horizon? Providing answers to these questions, based on the latest research and developments of the broadband communications technologies, is the focus of the following chapters.

With all this in mind, drawing on research experiences and lessons from over the globe, this book explores the latest research and developments of the broadband communications technologies associated with broadband communications network architectures in support of many emerging paradigms/applications of the global Internet from the traditional architecture to the incorporation of smart applications.

Author details

Abdelfatteh Haidine* and Abdelhak Aqqal

*Address all correspondence to: haidine.a@ucd.ac.ma

Laboratory of Information Technologies, National School of Applied Sciences, Chouaib Doukkali University, Morocco

References

[1] Cisco. The Zettabyte Era: Trends and Analysis. Cisco's White Paper—Part of Part of the Cisco Visual Networking. June 2017. Available from: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.pdf [Accessed: July 25, 2018]

[2] FTTH Council Europe. FTTH/B European Ranking—Sep 2017 [Internet]. February 2018. Available from: http://www.ftthcouncil.eu/documents/FTTH%20GR%2020180212_FINAL.2.pdf [Accessed: July 22, 2018]

[3] Ericsson AB. Ericsson Mobility Report. Ericsson's Reports. November 2017. Available from: https://www.ericsson.com/en/mobility-report/reports/november-2017 [Accessed: July 25, 2018]

[4] Saidu-Huq KM, Jornet J, Gerstacker W, Al-Dulaimi A, Zhou Z, Aulin J. THz communications for mobile heterogeneous networks. IEEE Communications Magazine. 2018;**56**(6): 94, 95

[5] Nutaq. Towards 5G—Business Model Innovations [Internet]. Available from: https://www.nutaq.com/towards-5g-business- [Accessed: July 21, 2018]

[6] Nationaler IT Gipfel-AG2. Empfehlungen für eine nationale Strategie Intelligente Netze. Report (in German) Arbeitsgruppe 2-Digitale Infrastrukturen als Enabler für innovative Anwendungen. Essen; 2012

[7] Arbeitsgruppe 2 des Nationalen IT-Gipfels. Digitale Infastrukturen - Schwerpunkte und Zielbilder für die Digitale Agenda Deutschlands. Arbeitsgruppe 2, Jahrbuch 2013/2014 (in German); 2014

[8] Third Generation Partnership Project (3GPP). Technical Specifications Group radio Network; Study on Licensed-Assisted Access to Unlicensed. Technical report 3GPP-36889-d00; 2015

[9] Zhou Z, Mumtaz S, Huq KMS, Al-Dulaimi A, Chandra K, Rodriquez J. Cloud miracles: Heterogeneous cloud RAN for fair coexistence of LTE-U and Wi-Fi in ultra dense 5G networks. IEEE Communications Magazine. 2018;**56**(6):64-71

Section 1

Wireless/Mobile Broadband Network

Chapter 2

Evolution of Broadband Communication Networks: Architecture and Applications

Sonia Gul and Jairo Gutierrez

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.73590

Abstract

With the rapid increase in users' demand for flexibility and scalability of communication services, broadband communication networks are facing an ongoing challenge of providing various broadband services using a single communication architecture. This leads to the evolution of a challenging field of multiservice broadband network architectures. This chapter discusses the basic concepts associated with broadband communication network architectures with emphasis on provision of multiservice, and it also focuses on the evolution of broadband communication networks from the traditional architecture to the incorporation of virtualization services, that is, cloud computing. Another important aspect, which relates to the multiservice broadband network, is the "applications" which, as this chapter highlights, are a key-driving factor for the evolution of broadband communication networks. Moreover, this chapter also includes a discussion on New Zealand's government initiatives to provide improved network coverage within the country.

Keywords: broadband, communication network, evolution, applications, architecture

1. Introduction

Broadband communication networks have become increasingly popular. These are the networks, which give telecommunications a new perspective by supporting traffic of multiple types, that is voice, video and data (also known as multimedia), but also communicating these to the end user using a single packet. Some of the networks which started with the provision of multimedia capabilities include asynchronous transfer mode (ATM) and Frame Relay [1].

Broadband communication networks are regarded as aggregated networks (providing voice, video and data) over the wired network including Ethernet and fibre. The multiservice support of these networks not only opened a new era in telecommunications but rather also has challenged the network operator's abilities and capacities.

In this chapter, a tutorial background on broadband communication networks is provided. The challenges faced and the techniques adopted by the service providers are described with the insight into the evolution of broadband communication networks starting from ATM to the virtualization of broadband services.

2. Broadband communication networks

During the 1980s, the telecommunication industry started to work towards the concept of providing any type of information to anyone, anywhere. This concept pushed the evolution of wireless networks including both cellular networks and wireless local area networks (WLANs). The WLANs have then evolved supporting faster data rates, higher throughput and better roaming capabilities, hence a more efficient communication network for offices, home and other purposes. Similar to WLANs, cellular networks have also gone into the transition from 2G to the current 5G efforts mainly targeting the communication needs of mobile users. All these advancements brought up a range of technologies including ATM, International Mobile Telecommunication (ITM)-2000 systems [2], wireless IP networks, WLANs, and 4G and 5G networks (ITM-2020).

2.1. History: standardization point of view

The evolution of broadband communication networks can be seen as a competition between two technological domains, namely WLAN and cellular. The international standardization organizations have been working hard to progress each of these technological domains that are taking over the broadband communication era [3]. A number of standards under both domains are being set, published and provided to the market to be used commercially. In **Tables 1** and **2**, we summarize the standards to date for both technological domains.

2.2. Technological view point

The terminology "band" has been used for a long time by engineers, and it started with the definition referring to a set of channels and/or frequencies. Later, this term has been used with other adjectives to make it sound more understandable, for example, baseband, passband, etc. In the 1980s, the term wideband started to be used very frequently when referring to a number of channels [4].

In telecommunications, broadband communications are achieved using a wideband of channels. This allows channels to have more capacity, and hence they can support communications from

Year	WLAN standards	Year	WLAN standards
1997	IEEE 802.11-1997	2009	IEEE 802.11 n, w
1999	IEEE 802.11b	2010	IEEE 802.11z
2001	IEEE 802.11 a, c, d	2011	IEEE 802.11 s, u, v
2003	IEEE 802.11F, g	2012	IEEE 802.11 aa, ad, ae, 802.11-2012
2004	IEEE 802.11h, i, j	2013	IEEE 802.11 ac
2005	IEEE 802.11e	2014	IEEE 802.11af
2007	IEEE 802.11-2007	2016	IEEE 802.11 ai, ah, 802.11-2016
2008	IEEE 802.11k, r, y	2017 (in progress)	IEEE 802.11 aj, aq, ak, ax, ay, az, ba

Table 1. IEEE WLAN standard summary [5].

the applications that require high data rates. As per the standard (802.16-2004) broadband refers to "having instantaneous bandwidths greater than 1 MHz and supporting data rates greater than about 1.5 Mbit/s" [6]. Lately, the Federal Communications Commission (FCC) has increased the download and upload speeds to at least 25 and 3 Mbits/s, respectively [7].

Broadband communication networks are still considered as one of the vital factors which may help businesses throughout the world to achieve their goals for sustainable development, by 2030 [8]. A number of broadband technologies are required to work together to meet this goal. These technologies include, but are not limited to, mobile and fixed broadband, backhaul satellite networks, Wi-Fi (unlicensed) technologies and cellular (licensed) networks.

With the advancements in technology, wireless networks are able to support almost the same data rates as those of some wired networks (including cable modems and asymmetric digital subscriber line (ADSL). Therefore, both fixed and wireless broadband networks have grown tremendously in previous years [9]. In **Table 3**, the broadband technologies to date are summarized for easy reference.

Year	Cellular network standards	Year	Cellular network standards
1985	AMPS (TIA, EIA), N-AMPS, TACS, ETACS	2003	UMTS, W-CDMA
1992	GSM, CSD, HSCSD, CDMA, D-AMPS	2005	HSPA, HSPA+, LTE, WiMAX, Flash-OFDM
2000	GPRS, EDGE, CDMA2000	2011–2013	LTE Advanced/Pro, WiMAX (IEEE 802.16m), WiMAX 2.1

Table 2. Popular cellular network standard summary [10].

Year	Broadband technologies	Data rate (max)
Fixed line		
1990	Hybrid fibre coaxial (HFC)	400 Mbits/s
1990s	Broadband over power lines (BPL)	3 Mbits/s [11]
1998	ADSL	12.0–1.8 Mbits/s
2003	ADSL2+	24.0–3.3 Mbits/s
2008	FTTH-FTTX	2.5 Gbits/s–622 Mbits/s [12]
Wireless		
1947	Microwave	800 Mbits/s
1998	LMDS	64 kbits/s–155 Mbits/s [13]
1998	MMDS	27–38 Mbits/s
2000	W-CDMA, CDMA2000	153 kbits/s
2008	FSO	10 Gbits/s
2008	LTE (standard)	144 Mbits/s
2011	WiMAX	30 Mbits/s–1 Gbits/s
2011	LTE Advanced	200–300 Mbits/s
2015	LTE Advanced Pro	1 Gbits/s

*Satellites are also being used for broadband communications.

Table 3. Popular broadband technology summary.

3. Multiservice-broadband network architecture evolution

The evolution of multiservice broadband network architectures extends from the digital subscriber line (DSL) architecture [14, 15]. The network architecture has evolved to fulfill the increasing demands of service, not only for residential users but also for wholesale markets and businesses. Based on these the focus is not only of the provision of quality of service (QoS) but also on availability and reliability. This leads to a whole list of new motivations about the services which were expected from the multiservice broadband network architecture including a simpler network design architecture, unified connectivity, enhancements and improvements in operations, independent provision of services, improved availability, improved scalability and efficient support for multi-edge services [15, 16].

Multiservice broadband architectures evolve to address the needs of triple play and converged broadband networks. This evolution provides a pathway for service providers to face the upcoming broadband network challenges in an effective and efficient manner. Some of these challenges are listed below [17]:

- Easy and quick provision of services for both business and residential users
- Support of IP-based applications

- High level of quality of service (QoS), security, reliability, scalability and availability
- Effective management of services
- Support for virtualization services

3.1. Triple play services: aggregated services using multicast

In the early 2000s, there was an increasing demand for voice, data and video (triple play) services especially for residential purposes. Based on the demand, the network architecture started to evolve towards providing these triple play services using Ethernet technology [18–21]. One major requirement which turns up for the provision of the said service aggregation is multicast forwarding which is designed to minimize the number of network links used for media streams.

On one side, the use of multicast enhances the performance and reduces the network load, while, on the other hand, using it with the traditional Point-to-Point Protocol over Ethernet (PPPoE) sessions makes the big picture more complex [22, 23]. To address this issue, the use of a more dynamic protocol, that is, DHCP is introduced instead of PPP for IP [24]. In broadband networks the sessions using DHCP started to prove their worth by proving simple and always-on connections for residential users.

This aggregated multicast networks proved so successful for residential clients that soon service providers started to offer these for businesses. This led to the next step of its evolution where the same network architecture is used to transport both mobile and fixed traffic. This was enabled with the introduction of backhauling for mobile traffic of 2G/3G networks, a capability that was later moved to work with 4G/LTE networks [25, 26].

3.2. IP to MPLS-based routing

Internet Protocol/Multiprotocol Label Switching (IP/MPLS) is a well-adopted technology [27, 28]. However, based on its flexible nature, it was initially used for backbone and core networks. Later on, this technology also started to spread in the aggregated service network domain and started to support the service access layer which connects IP-based service nodes and customer premises equipment (CPEs).

The success of IP/MPLS in the access layer of multiservice broadband networks opened new horizons for building a unified network architecture using commonly used technologies. This allowed service providers to use the same technology throughout their network including at the access, aggregation and core levels, which results in [29, 30]:

- Better scalability of the network
- Flexibility in the placement of service nodes
- Simple provision of aggregated services

Based on the need and functionality of access node (AN), two implementation models for IP/MPLS have been mainly used [17]:

3.2.1. Seamless MPLS model

Seamless MPLS model extended only limited functionality to the AN. This model used the simplest form of IP routing, that is, static routes between aggregation nodes and the access nodes. To get better scalability, the label distribution feature of MPLS is used.

3.2.2. Full MPLS model

On the other hand, full MPLS model extended the complete functionality of Layer 3, that is, dynamic routing to AN. This makes access nodes and aggregation nodes functionally equivalent.

The choice of model depends on the specific requirements and the current network structure.

3.3. Architecture-supported functions

Based on the challenges mentioned in the previous section, the functions of the multiservice broadband architecture are defined. These functions are defined keeping in mind the main objective, that is, provision of all these services using a common network infrastructure [31]:

- Layering functions include forwarding (relaying layer information) and termination/adaptation functions (mapping user information to a specific layer).
- Control functions include session control and resource control.
- Filtering and scheduling involve filtering of data (e.g. using ACL-access control lists) and scheduling considering priorities, policies, etc.
- Synchronization functions deal with frequency, phase and time synchronization.

3.4. Service layers

The unique multiservice nature allows the provision of services at various layers. This yield to the design of various service layers as follows:

- IP-service layer: These include the IP-layer services which can be seen directly by the end user. Such services include VPNs, Internet access for business and residential purposes, etc.
- Ethernet-service layer: These are the services which provide transport capabilities based on the service, for example, service aware, such as Ethernet access services (some are defined by Metro Ethernet Forum (MEF)), etc. [32, 33]. This is mainly achieved by using the concept of Infrastructure Virtual Circuit (IVC).
- Support-aggregation layer: These services support to map Ethernet services on top of other technologies, for example, IP/MLPS, etc.

Cloud computing and virtualization services do not get across to all the above-mentioned layers. **Figure 1** depicts the view of the discussion above.

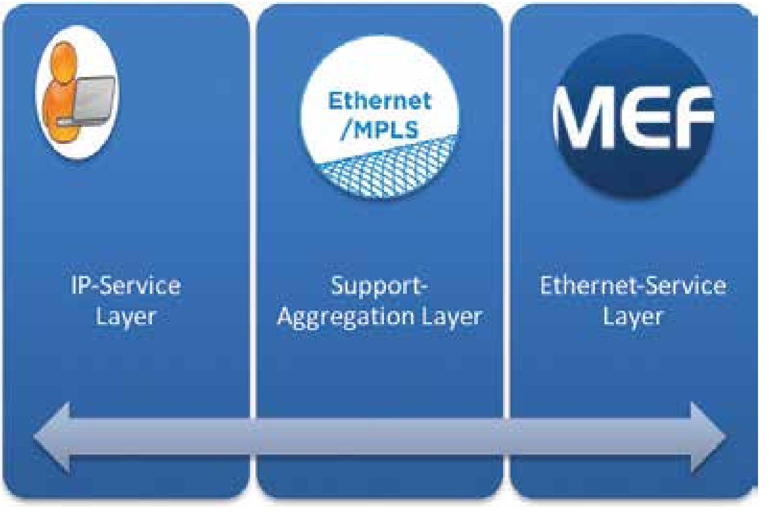

Figure 1. Multiservice broadband network's service layers.

3.5. Trend of fixed mobile convergence

The need for ubiquitous service delivery between fixed and mobile networks has emerged with the advancement in technology, specifically the availability of IP-based mobile handsets/devices. The number of mobile devices' users continues to grow, and the demand for service availability regardless of the type (i.e. fixed or wireless) of the access network has increased. This laid the basis for a new trend in technology, that is, fixed mobile network convergence [34].

Several aspects have been considered for the internetworking of fixed and mobile network architectures, but mainly the standardization bodies have focused on the following:

- Convergence of business needs and services
- Convergence of network technologies and infrastructure
- Convergence of end user devices and management

The broadband forum has produced a technical report considering all the challenges and aspects of the internetworking of fixed and mobile networks [35]. The main aim is to provide a converged network architecture that will support the provision of any service, anywhere to anyone regardless of type of the access network.

For the convergence of network technologies and infrastructure, one of the emerging solutions is the use of fibre wireless (FiWi) networks. Wired networks based on fibre optics are considered as having potential to deliver huge bandwidth to the end users; however, the technology has limitations in terms of supporting end user roaming requirements. While the networks using wireless-access technologies are supporting easy roaming and mobility, they are not supporting high-bandwidth and long-distance solutions. Fibre wireless (FiWi) has introduced to the best features of both wired-fibre networks and wireless-access networks [36]. FiWi technology allows to use wireless technologies for access, while the rest of the network is mainly fibre.

A conceptual view of the fixed mobile convergence is illustrated in **Figure 2**.

The above-illustrated conceptual view incorporates various technologies which support high-bandwidth requirements with mobility. Such technologies include access networks using mm-wave and radio-over-fibre (RoF), micro−/millimetre wave-based relay and RoF and digital baseband-based core/backhaul networks. Another important aspect included in the above conceptual view is the support of cloud computing. More details on cloud computing and virtualization support are provided in the following section.

3.6. Cloud computing and virtualization support

To keep pace with the increasing demands of the applications over the network, service providers have started to embrace cloud computing. Cloud computing has given freedom to the service providers which enabled them to serve the user requests with the use of virtualization services in a cost-effective and time-saving manner. Virtualization enabled them to achieve ubiquitous and on-demand access to network services.

Cloud computing is one of the vital parts of the multiservice broadband architecture. Virtualization services have been added in the said architecture in many ways; however, one of the popular techniques is to incorporate virtual services as one of the network functions [17]. Virtualization services should be providing the following features:

- The resources are accessible immediately as per the request, and the allocation can be terminated when the job is done.
- The resources should be available whenever requested, that is, resource scarcity should not occur.

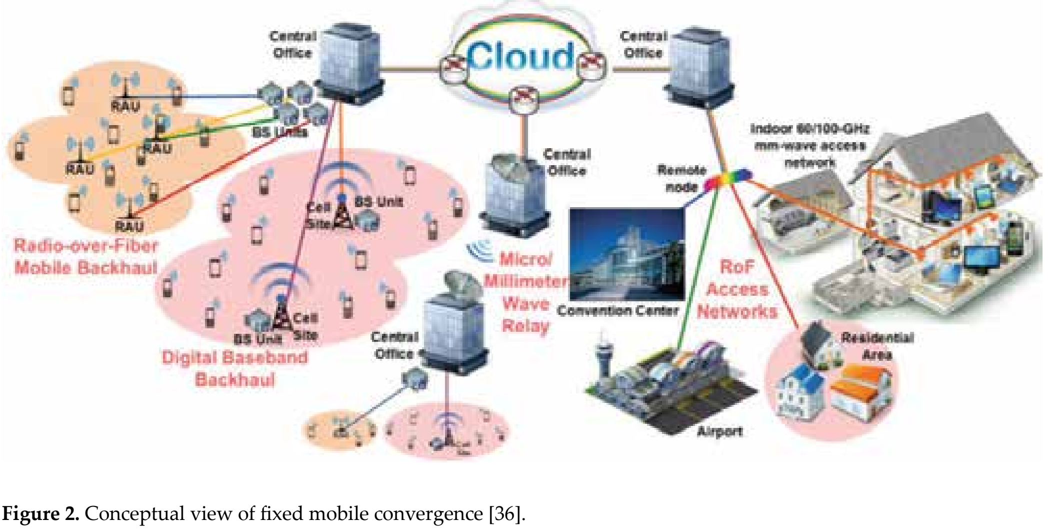

Figure 2. Conceptual view of fixed mobile convergence [36].

- Supports resource scalability according to the demand.
- Ubiquitous access of resources is provided.
- Resources are provided using a single infrastructure for multiple users.
- Accountability of resources usage is done, and customer is billed accordingly.

A number of frameworks have been proposed to fulfill the virtualization of service requirements. These cloud services have taken communication networks towards an entire new dimension. Currently, researchers have stated working towards a multi-access edge computing (MEC) platform. The MEC aims to converge services from IT and telecommunications and provide these at the edge of a radio network, that is, access layer. It also proposes to use cloud computing to provide an efficient access to requested services [37–39]. Another key tool used to achieve the said virtualization is network function virtualization (NFV) [40]. This is considered as an embedded part of the network, which ensures tight integration among various network nodes in order to provide a better communication service.

Cloudlets [41] are also introduced to address the challenge of quick response between the mobile devices and the associated cloud. Cloudlets are there to support the applications with high user interaction and require real-time responses. Cloudlets are small-scale cloud-based data centres which are located towards the edge of Internet to provide quick responses to mobile user requests. One of the other candidate technologies for supporting virtualization is fog computing [42]. It is considered as the combination of edge and cloud computing.

4. Application-driven network evolution

The evolution of multiservice broadband networks was not only led by the advancement in technology; another rather important factor was the "Applications". Applications are what the end users' experience, and network architectures have to enhance their offerings to provide end users with the flawless experience they are looking for. The following are some of the application trends, which contributed to network architecture advancements:

- IP television supporting through broadcasting
- Video on demand (VoD)
- Internet TV
- Video playing functions: play, pause, rewind, forward, etc.
- Distant learning
- Advertising: embedded inside the video
- Internet-based business-supported applications including HTTPS, Email, FTP and VPNs
- Gaming: multiplayer Internet-based gaming
- High-definition TV
- IP telephony, etc.

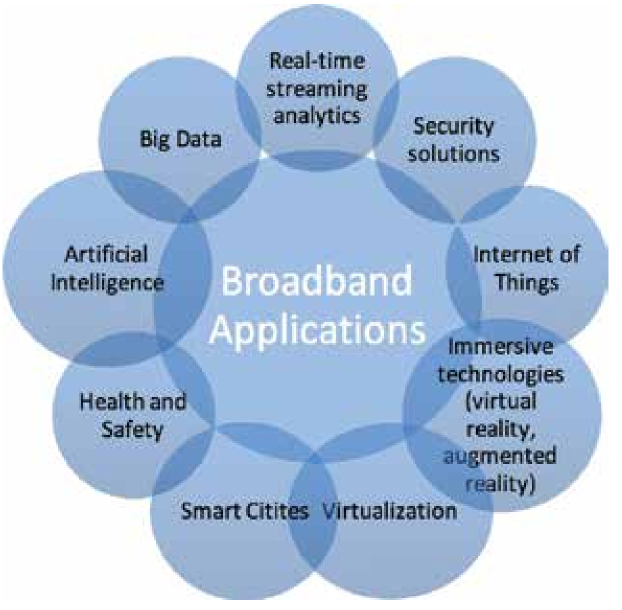

Figure 3. Future broadband application domains.

4.1. Latest trends in broadband applications

In 2017, the growing demand for broadband communications still persists with the increasing need for mobility, machine-to-machine communications, big data, all-purpose sensors and the Internet of things (IoT) [43].

The need for hybrid networks (fiber and wireless) is growing to address the challenging goals of upcoming application domains. The latest application domains for broadband networks are highlighted in **Figure 3**. With the ongoing growth of broadband users for both business and residential [44] in conjunction with increasing demands for large volumes of data, the applications are getting more and more data hungry.

5. Broadband network: a New Zealand perspective

New Zealand has taken a number of steps to embrace the fibre broadband in previous years. Some major initiatives are the Ultra-Fast Broadband (UFB), Rural Broadband Initiative (RBI) and Mobile Black Spot Fund (MBSF) [45].

5.1. Ultra-fast broadband (UFB)

The UFB project aims to deploy optical fibre cables to provide fibre to the premises (FTTP) to as many New Zealanders as possible. Previously copper lines were laid in the whole country, and they served as the main communication medium. With the advancement in technologies and the need for more speed and data (more specifically broadband technologies), the copper has to be replaced by optical fibres. With the UFB New Zealanders will be able to access data and applications at the speed of 1000 Mbits/s approx.

In New Zealand, UFB is considered one of the biggest infrastructure-based projects. Around 85% of the population will have access to fibre to the premises (FTTP) towards the end of 2024 [46]. The NZ government, to make the UFB accessible by as many people as possible, is investing $1.8 billion. **Figure 4** shows the progress of the UFB deployment till June 2017.

5.2. Rural broadband initiative (RBI)

The Ultra-Fast broadband project mainly focuses on the provision of fast Internet services to urban areas. However, there are various rural and coastal areas in the country, which also need access to these fast network connections. To ensure every New Zealander can access and experience the improved Internet access, the NZ government has started another project named Rural Broadband Initiative (RBI). Funds worth 430 million NZD (approx.) has been allocated for this initiative.

The RBI project is divided in multiple phases. The first phase of the RBI project has already been completed by mid-2016. In this phase fast broadband connections are being provided to rural areas using the combination of upgrading existing fixed lines and installing new wireless fixed coverage solutions. **Figure 5** shows the highlights of the improved connectivity in rural areas after the completion of RBI phase 1.

The RBI (phase 2) project aims to provide fast broadband connection to more than 70,000 businesses and households in remote and rural areas. For the second phase, the NZ government is encouraging local network operators to propose some innovative ideas/solutions rather than specifying any particular technology use.

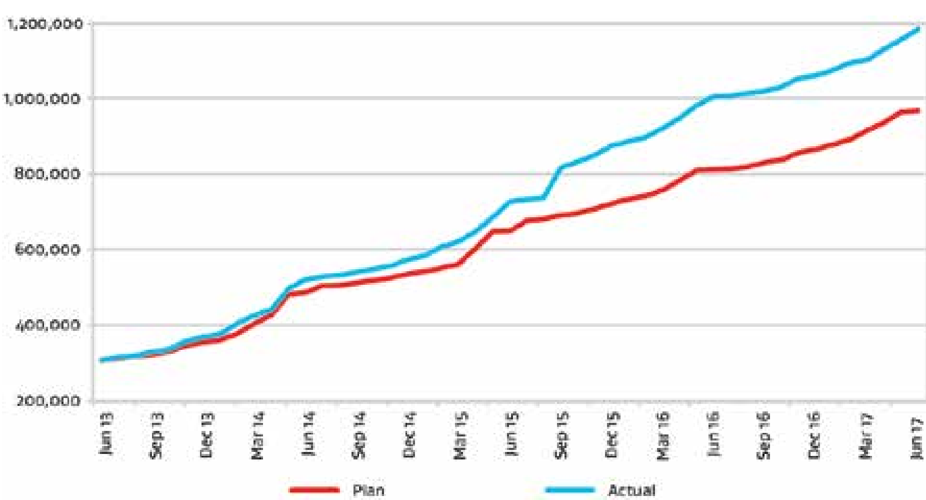

Figure 4. User (business and household) connections, UFB-1 [46].

Figure 5. Summary of improved connectivity in rural areas after RBI phase 1 [47].

5.3. Mobile black spot fund (MBSF)

Another important step taken by the NZ government is the creation of MBSF. The purpose of MBSF is to provide improved network coverage to areas which are of tourists' interest and also to better cover the country's state highways. The government is looking forward to achieving improvements in the fields of public safety and in the tourism industry. The project will explicitly target two state highways (6 and 94) in Southland, covering in total of 11 tourism areas [48].

6. Conclusion

This chapter discusses the evolution of broadband communications with a focus on the development and adoption of multiservice broadband network architectures with the support of cloud and virtualization services. The need for this evolution is also discussed with focus on triple play services, IP/MPLS and mobile-fixed network convergence. Applications are also recognized to play a vital role in this evolution of broadband networks with the latest trends in broadband applications being discussed. Lastly, an overview of New Zealand's government initiatives to improve the network coverage is also provided including a brief discussion about Ultra-Fast Broadband, the Rural Broadband Initiative and the Mobile Black Spot Fund.

Author details

Sonia Gul* and Jairo Gutierrez

*Address all correspondence to: sonia.gul@aut.ac.nz

Auckland University of Technology, Auckland, New Zealand

References

[1] Akujuobi CM, Sadiku MO. Introduction to Broadband Communication Systems. Boca Raton/[Raleigh, NC]: Chapman & Hall/CRC/SciTech Pub.; 2008

[2] Haas M. Management of innovation in network industries. Wiesbaden: Deutscher Universitäts-Verlag; 2006

[3] Adachi F. Evolution towards broadband wireless systems. In The 5th International Symposium on Wireless Personal Multimedia Communications, 2002; Vol. 1. IEEE; 2002 October, pp. 19-26

[4] Fontana RJ. A Brief History of UWB Communications. Multispectral.com, Multispectral Solutions, Inc.; 2002. www.multispectral.com/history.html

[5] IEEE 802.11. The Working Group Setting the Standards for Wireless LANs [online]. IEEE802.org. 2017. Available at: http://www.ieee802.org/11/ [Accessed: 25 October 2017]

[6] P802.16h/D6. IEEE Draft Amendment to IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems: Improved Coexistence Mechanisms for License-Exempt Operation (Amendment to IEEE 802.16-2004); May 2008

[7] FCC Finds U.S. Broadband Deployment Not Keeping Pace | Federal Communications Commission. Fcc.gov. 2015-02-04. Retrieved 2016-06-21

[8] Broadband Annual Report 2016; Broadband Commission. URL: www.broadbandcommission.org/Documents/reports/bb-annualreport2016.pdf [Accessed: 24-Oct-2017]

[9] Bolcskel H, Paulraj AJ, Hari KV, Nabar RU, Lu WW. Fixed broadband wireless access: State of the art, challenges, and future directions. IEEE Communications Magazine. 2001; **39**(1):100-108

[10] All about the technology. Itu.int. (2017). [online] Available at: https://www.itu.int/osg/spu/ni/3G/technology/index.html [Accessed 25 October 2017]

[11] Held G. Understanding broadband over power line. Boca Raton, FL: Auerbach; 2006

[12] DeCusatis C, Kaminow I. The Optical Communications Reference. 1st ed. Burlington: Elsevier Science; 2009

[13] Ahamed SS. Review and Analysis of Local Multipoint Distribution System (LMDS) to Deliver Voice, Data, Internet, and Video Services. arXiv preprint arXiv:1001.4184; 2010

[14] Song KB, Chung ST, Ginis G, Cioffi JM. Dynamic spectrum management for next-generation DSL systems. IEEE Communications Magazine. 2002;**40**(10):101-109

[15] Shin DH. Determinants of customer acceptance of multi-service network: An implication for IP-based technologies. Information Management. 2009;**46**(1):16-22

[16] Roberts JW. Realizing quality of service guarantees in multiservice networks. In: Performance and Management of Complex Communication Networks. US: Springer; 1998. pp. 277-293

[17] Broadband Multi-Service Architecture & Framework Requirements. Technical Report (TR-144) Broadband Forum; Sept. 10, 2014

[18] Haque S, Jain A. U.S. Patent Application No. 11/293,120; 2005

[19] Kung FC, Russell JE, Sankalia A, Wang S. U.S. Patent No. 7,120,139. Washington, DC: U.S. Patent and Trademark Office; 2006

[20] Ghaibeh G. U.S. Patent No. 5,926,476. Washington, DC: U.S. Patent and Trademark Office; 1999

[21] Anschutz T. Migration to ethernet-based broadband aggregation broadband forum tr-101 issue 2; 2011

[22] Melsen T, Andersen OH. U.S. Patent No. 7,577,137. Washington, DC: U.S. Patent and Trademark Office; 2009

[23] Newell L, Sif M. Optimizing the broadband aggregation network for triple-play services. Annual Review of Broadband Communications. 2006;**15**

[24] Marc Bernstein, IPTV Solutions. Understanding PPPoE and DHCP (white paper). California, USA: Juniper Networks, Inc; 2006

[25] Dahlman E, Parkvall S, Sköld J. 4G: LTE/LTE-advanced for mobile broadband: Amsterdam: Elsevier; 2014;**2014**

[26] Hwang JS, Consulta RR, Yoon HY. 4G Mobile Networks–Technology beyond 2.5 G and 3G. PTC (Pacific Telecommunications Council) Proceedings, Hawaii; 2007

[27] Robertazzi TG. Multiprotocol label switching (MPLS). In: Introduction to Computer Networking. Springer International Publishing; 2017. pp. 61-65

[28] RFC 7900 on Extranet Multicast in BGP/IP MPLS VPNs [Internet]. Ietf.org. 2018. Available from:https://www.ietf.org/mail-archive/web/ietf-announce/current/msg15584.html [Accessed: 01-FEB-2018]

[29] Savi M, Rozic C, Matrakidis C, Klonidis D, Siracusa D, Tomkos I. Benefits of multi-layer application-aware resource allocation and optimization. In: 2017 European Conference on Networks and Communications (EuCNC); IEEE. 2017, June, pp. 1-5

[30] Wright P, Lord A, Velasco L. The network capacity benefits of Flexgrid. In: 2013 17th International Conference on Optical Network Design and Modeling (ONDM); IEEE. 2013, April. pp. 7-12

[31] Multi-service Broadband Network Functional Modules and Architecture; TR-145 Broadband forum; Nov. 2012

[32] Carrier Ethernet Network Architecture Framework Part 2: Ethernet Services Layer; MEF Technical Specification 12.2; May 2014

[33] Migration to Ethernet-Based DSL Aggregation; TR-101 Broadband forum; Apr. 2006

[34] Shen G, Tucker RS, Chae CJ. Fixed mobile convergence architectures for broadband access: Integration of epon and wimax [topics in optical communications]. IEEE Communications Magazine. 2007;**45**(8):44-50

[35] Internetworking between next generation mobile and 3GPP fixed wireless network. TR-203 Broadband Forum; Aug. 2012

[36] Tornatore M, Chang GK, Ellinas G, editors. Fiber-Wireless Convergence in Next-Generation Communication Networks: Systems, Architectures, and Management. Springer; 2017 Feb 10

[37] Taleb T, Samdanis K, Mada B, Flinck H, Dutta S, Sabella D. On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration. IEEE Communications Surveys & Tutorials. 2017;**19**(3):1657-1681

[38] Rimal B, Van D, Maier M. Mobile Edge Computing Empowered Fiber-Wireless Access Networks in the 5G Era. IEEE Communications Magazine. 2017;**55**(2):192-200

[39] Nunna S, Ganesan K. Mobile edge computing. In: Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare. Springer International Publishing; 2017. pp. 187-203

[40] Han B, Gopalakrishnan V, Ji L, Lee S. Network function virtualization: Challenges and opportunities for innovations. IEEE Communications Magazine. 2015;**53**(2):90-97

[41] Satyanarayanan M, Bahl P, Caceres R, Davies N. The case for vm-based cloudlets in mobile computing. IEEE Pervasive Computing. 2009 Oct;**8**(4)

[42] Bonomi F, Milito R, Zhu J, Addepalli S. Fog computing and its role in the internet of things. In: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing; ACM. 2012 Aug 17. pp. 13-16

[43] 2017 Trends in Broadband|Intelsat. [online]. Available at: http://www.intelsat.com/news/blog/2017-trends-in-broadband/ [Accessed: 27 Oct. 2017]

[44] Ericsson Mobility Report 2016; [online] Available at: https://www.ericsson.com/assets/local/mobility-report/documents/2016/Ericsson-mobility-report-june-2016.pdf [Accessed: 27 Oct. 2017]

[45] Broadband and Mobile Programmes; [online] Available at: http://www.mbie.govt.nz/info-services/sectors-industries/technology-communications/fast-broadband/broadband-and-mobile-programmes#rural [Accessed: 18 Nov. 2017]

[46] Broadband Deployment Update Report 2017; [online] Available at: http://www.mbie.govt.nz/info-services/sectors-industries/technology-communications/fast-broadband/documents-image-library/june-17-quarterly-broadband-update.pdf [Accessed: 18 Nov. 2017]

[47] Rural Broadband Initiative Phase 1 Report 2016; [online] Available at: http://www.mbie.govt.nz/info-services/sectors-industries/technology-communications/fast-broadband/documents-image-library/rural-broadband-initiative/rural-broadband-initiative-phase-1-august-2016.pdf [Accessed: 18 Nov. 2017]

[48] RBI and Mobile Black Spot Funding Decision Announced; [online] Available at: http://investors.sparknz.co.nz/DownloadFile.axd?file=Announcements/NZX/20170830/264825.pdf [Accessed: 18 Nov. 2017]

Chapter 3

Spectrum Usage for 5G Mobile Communication Systems and Electromagnetic Compatibility with Existent Technologies

Guntis Ancans and Vjaceslavs Bobrovs

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72431

Abstract

The increased demand of consumers on services in the mobile broadband environment with high data rate and developed mobile broadband communication systems will require more spectrum to be available in the future. New technologies as well as the existing services require frequencies for their development. In this chapter, we investigate the available and potential future mobile terrestrial radio frequency bands (5G)—worldwide and in Europe. An insight into the mobile spectrum estimate is provided. Characteristics and requirements of IMT-2020, future possible IMT frequency bands, and examples of 5G usage scenarios are also addressed in the chapter. Electromagnetic compatibility evaluation methods are provided mainly focusing on existent mobile technologies below 1 GHz where also 5G technologies will be developed in the future. It is stressed that the radio frequency spectrum is a limited national resource that will become increasingly precious in the future.

Keywords: 4G mobile communication, 5G, electromagnetic compatibility, frequency band, international mobile telecommunications (IMT), IoT, international telecommunication union (ITU), M2M, mobile service, radio wave propagation, spectrum planning, WRC-19

1. Introduction

5G referred to as IMT-2020 in ITU-R terms is the next generation of mobile communication technologies. IMT systems are now being evolved to provide diverse usage scenarios and applications such as enhanced mobile broadband (eMBB) communication, massive machine-type communication (mMTC), and ultrareliable and low-latency communication (URLLC) requiring larger contiguous blocks of spectrum than currently available bandwidth to realize those applications.

5G aims to provide high data rates, low latency, seamless coverage, low power, and highly reliable communications. Used cases under consideration include enhanced mobile broadband communications, but also machine-to-machine (M2M), Internet of Things (IoT), home and industrial automation and applications, etc. expected to respond to requirements from vertical sectors (e.g., utilities, automotive, railways, public protection). 5G is planned to be deployed around the world by 2020.

The potential usage scenarios shown in **Figure 1** have different operational and technological requirements.

Different players from various verticals, i.e., different industries, can be brought together using the 5G concept. The network capabilities are intended to match the requirements of the different vertical players.

The first 5G specification in 3GPP Release 15 is planned to be available by the end of 2018 and will address the more pressing commercial needs. The second release, 3GPP Release 16, planned for March 2020, will address all used cases and requirements. There is an ongoing work on development of new radio access technology targeted for completion by the end of 2017.

3GPP has been working to standardize the 5G-NR (New Radio) specification. In March 2017, 3GPP decided to accelerate the timescale in order to finalize the *non-stand-alone* mode by March 2018. This mode will operate in parallel with 4G long-term evolution (LTE) to provide boosted data rates. The *stand-alone* 5G mode is planned for completion in September 2018. This accelerated timescale will facilitate 5G network trials in the early 2018.

In general, 5G technologies will describe the following characteristics: high-frequency operation, very wide bandwidth, massive beam forming, and interworking with LTE. ITU will complete its work for standardization of IMT-2020 no later than the year 2020 [2].

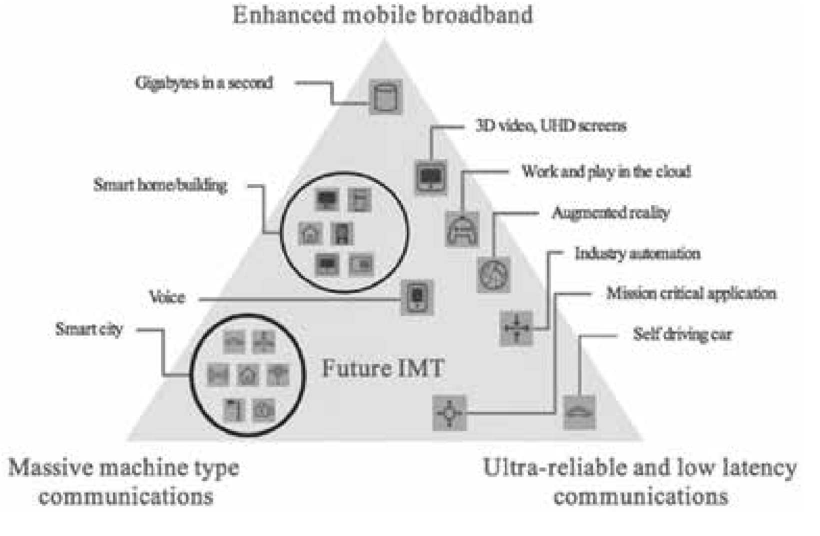

Figure 1. 5G usage scenarios according to ITU-R Recommendation M.2083-0 [1].

From a spectrum management's point of view, one of the main innovations brought by 5G is its capacity not only to handle broadband mobile communications as in the previous generations but also to cover the needs from a range of sectors, the so-called "verticals."

2. Mobile spectrum estimate for terrestrial IMT

The ITU terms for 3G and 4G are IMT-2000 and IMT-Advanced accordingly. The term IMT-2020 is adopted for 5G. Collectively, they are known as IMT [3]. IMT systems have contributed to global economic and social development.

With the mobile data traffic, increasing more spectrum resources will be necessary for the future mobile broadband communication systems. The Report ITU-R M.2290-0 provides a global perspective on spectrum requirement estimate for terrestrial IMT in the year 2020. The predicted total spectrum requirement for both low and high user density scenarios was calculated to be 1340 and 1960 MHz (including the spectrum already in use or planned to be used) at least by the year 2020 [4]. In some countries, national spectrum requirement can be lower than the estimate derived by lower user density settings, and in some other countries, national spectrum requirement can be higher than the estimate derived by higher user density settings. The mobile traffic forecast is presented in **Figure 2**.

It is assumed that for the year 2020, the median traffic growth will fall in between the lowest and highest growths, anticipating at least 25-fold traffic growth ratio in 2020 compared to 2010. Other estimates [1] anticipate that global IMT traffic will grow in the range of 10–100 times from 2020 to 2030.

An option for increasing data rates is the development of small cells and the combination of the capacity of unlicensed bands (e.g., 2.4 GHz, 5 GHz) with the capacity of a licensed

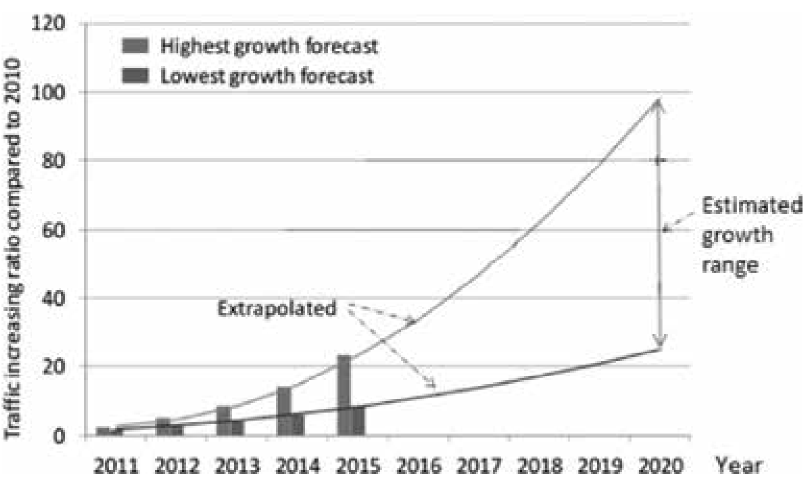

Figure 2. Mobile traffic forecasts toward 2020 by extrapolation according to the Report ITU-R M.2290-0.

frequency block [5]. Another option is carrier aggregation, which enables to increase data rates, but its complexity is exponential with the number of possible combinations of frequency bands used; spectrum sharing is also possible as a solution.

Another option is to develop and introduce the next generation of broadband communication technologies (5G) [6]. Authors presume that 5G base stations in the future will be connected by fiber optical lines or microwave backhaul links as an alternative solution. Huge investment in fiber is needed in order to realize the 5G vision.

3. Characteristics and requirements of IMT-2020

According to the Ericsson paper [7], LTE will evolve in a way that recognizes its role in providing ubiquitous wide area coverage for mobile users, and 5G networks will incorporate LTE radio access, based on orthogonal frequency division multiplexing (OFDM), along with new air interfaces.

Millimeter wave cells are very small, and they could be deployed mainly in dense urban areas or indoors delivering greater capacity. In the long term, it is expected that all devices that benefit from network connectivity eventually will become connected through M2M communications in the future.

According to Recommendation ITU-R M.2083-0, goals of future development of 5G capabilities are summarized in **Table 1**, which include IMT-2020 capability eight key parameters [1].

Performance requirements must be met but at the same time depend particularly on the used cases or scenario. The key capabilities of IMT-2020 are presented in **Figure 3** compared with those of IMT-Advanced.

In the enhanced mobile broadband scenario, peak data rate, user-experienced data rate, area traffic capacity, mobility, energy efficiency, and spectrum efficiency all have high importance in comparison to connection density and latency.

Parameter	**Key value for 5G**
Peak data rate	10–20 Gbit/s
User-experienced data rate	100 Mbit/s (for wide area coverage, e.g., in urban and suburban areas) and 1 Gbit/s (for hotspots, e.g., indoor)
Latency	1 ms
Mobility	500 km/h (e.g., for high-speed trains)
Connection density	106 devices/km^2
Energy efficiency	100x more than IMT-Advanced
Spectrum efficiency	3x more than IMT-Advanced
Area traffic capacity	10 Mbit/s/m^2

Table 1. 5G target capabilities.

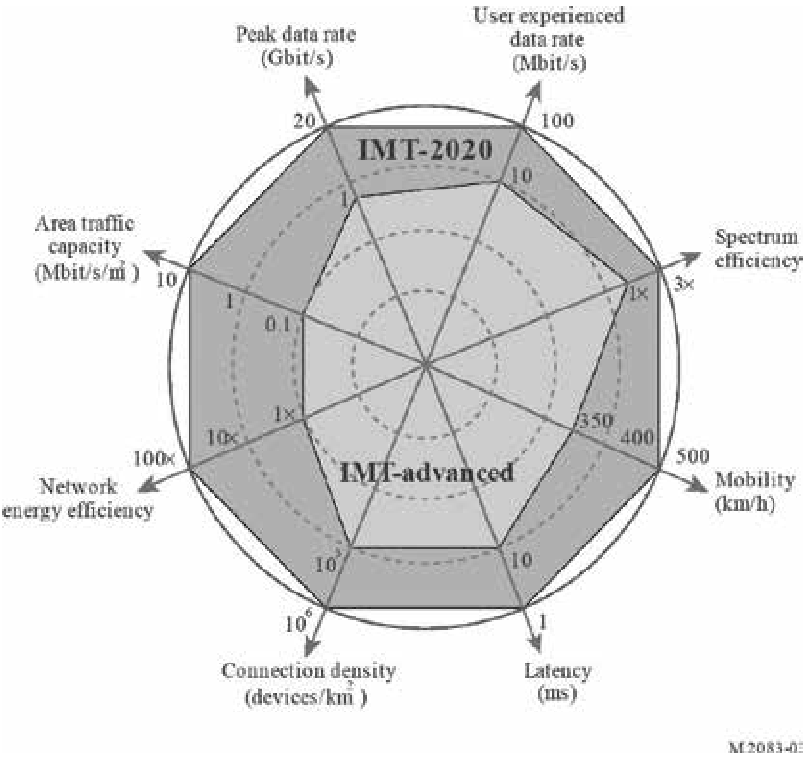

Figure 3. Enhancement of key capabilities from IMT-Advanced to IMT-2020.

In some low-latency communications and ultrareliable scenarios, low latency is of the highest importance, e.g., in order to enable the safety critical applications. Such capability would be required for some high mobility uses as well, e.g., in transportation safety, while, e.g., high data rates, might be less important.

The high connection density is needed in the massive machine-type communication scenario to support large number of devices in the network that, e.g., may transmit only occasionally at low bit rate and with zero or very low mobility. A low-cost radio device with long operational lifetime is of great significance for this usage scenario.

4. Future possible IMT frequency bands

In order to encourage increased data traffic capacity and to enable the transmission bandwidths needed to encourage very high data rates, 5G will extend the range of frequency bands used for mobile communications. This includes new radio spectrum below 6 GHz, as well as spectrum in higher frequencies.

For wireless communications, lower frequencies provide better coverage. Currently, almost all countries use spectrum below 6 GHz for IMT systems. Spectrum relevant for 5G wireless access therefore ranges from below 1 GHz up to approximately 100 GHz.

Frequency bands identified for IMT in the RR (MHz)	World Radiocommunication Conference (WRC)	Licensed mobile frequency bands in CEPT countries (MHz)
450–470	WRC-07	450–457.5 / 460–467.5
470–608	WRC-15	—
614–698	WRC-15	—
694–960	WRC-2000, WRC-07, WRC-12	694–790
		790–862
		880–915 / 925–960
1427–1518	WRC-15	1427–1518
1710–2025	WARC-92, WRC-2000	1710–1785 / 1805–1880
		1900–1920
		1920–1980 / 2110–2170
		2010–2025
2110–2200	WARC-92	1920–1980 / 2110–2170
2300–2400	WRC-07	2300–2400
2500–2690	WRC-2000	2500–2690
3300–3400	WRC-15	—
3400–3600	WRC-07	3400–3600
3600–3700	WRC-15	3600–3800
4800–4990	WRC-15	—

Table 2. Spectrum already identified for IMT.

High frequencies, e.g., those above 10 GHz, can only serve as a complement to lower frequencies and will mainly provide additional system capacity with very wide transmission bandwidths for extreme data rates for dense deployments. Spectrum use at lower-frequency bands will remain the backbone for mobile radio communication networks in the 5G era, providing excellent ubiquitous wide area connectivity.

4.1. Spectrum below 6 GHz

Besides achieving high data rates, it is also necessary to guarantee wide area coverage and outdoor to indoor coverage in 5G. Therefore, spectrum below 6 GHz forms a very important part of the 5G spectrum solution. Until now in Europe, more than 1200 MHz of spectrum for mobile broadband in the frequency range from 694 to 3800 MHz was harmonized.

For providing ubiquitous coverage in next-generation (5G) or pre-5G networks, the important role will be of LTE (4G) bands already harmonized below 1 GHz, including particularly the 700 and 800 MHz band, in order to enable nationwide and indoor 5G coverage. The 450 MHz band, which harmonized conditions for LTE use in the band currently is under development in the European Conference of Postal and Telecommunications Administrations

Frequencies to be studied until WRC-19 for possible identification to IMT (GHz)	Primary allocations to radiocommunication services in RR in ITU regions (including WRC-15 results)	Comments
24.25–27.5	Earth exploration-satellite (space-to Earth), fixed, fixed-satellite (Earth-to-space), inter-satellite, mobile, radionavigation, space research (space-to-Earth)	Frequencies have already allocation to the mobile service on a primary basis in RR
37–40.5	Earth exploration-satellite (Earth-to-space), fixed, fixed-satellite (space-to-Earth), mobile, mobile except aeronautical mobile, mobile-satellite (space-to-Earth), space research (Earth-to-space), space research (space-to-Earth)	
42.5–43.5	Fixed, fixed-satellite (Earth-to-space), mobile except aeronautical mobile, radio astronomy	
45.5–47	Mobile, mobile-satellite, radionavigation, radionavigation-satellite	
47.2–50.2	Fixed, fixed-satellite (Earth-to-space), fixed-satellite (space-to-Earth), mobile	
50.4–52.6	Fixed, fixed-satellite (Earth-to-space), mobile	
66–76	Broadcasting, broadcasting-satellite, fixed, fixed-satellite (space-to-Earth), inter-satellite, mobile, mobile-satellite (space-to-Earth), radionavigation, radionavigation-satellite	
81–86	Fixed, fixed-satellite (Earth-to-space), mobile, mobile-satellite (Earth-to-space), radio astronomy	
31.8–33.4	Fixed, inter-satellite, radionavigation, space research (deep space) (space-to-Earth)	Frequencies may require additional allocation to the mobile service on a primary basis in RR
40.5–42.5	Broadcasting, broadcasting-satellite, fixed, fixed-satellite (space-to-Earth)	
47–47.2	Amateur, amateur-satellite	

Table 3. Possible new spectrum for IMT in frequencies above 24 GHz.

(CEPT), in author's opinion will also play a significant role for enabling wide coverage for services of next-generation mobile networks in Europe. Frequency bands currently identified for IMT in the ITU Radio Regulations (RR) are presented in **Table 2**.

The *priority* frequency band suitable for the introduction of 5G use in Europe even before 2020 with wide channel bandwidths (50–100 MHz and more) could be 3400–3800 MHz band, noting that this band is already harmonized for mobile networks in Europe. This frequency band has the ability to put Europe at the forefront of the 5G or *pre*-5G deployment.

4.2. Spectrum above 6 GHz

5G envisages very high data rates, which will need much larger bandwidths than ever before. Those very high data rates may only be found in higher frequency bands (above 6 GHz). To deliver higher data rates and lower latency, there is an expectation that new wireless solutions at higher frequencies—millimeter wave (*mmWave*) bands—will be deployed. Therefore, implementation of frequency bands even above 24 GHz remains needed to ensure all the performance targets of 5G, e.g., multi-gigabit per second data rates. Implications of very low-latency drive to millimeter wave deployments with their highly directive antennas and small cell sizes.

The 2015 World Radiocommunication Conference (WRC-15) decided the following frequency bands 24.25–27.5 GHz, 31.8–33.4 GHz, 37–43.5 GHz, 45.5–50.2 GHz, 50.4–52.6 GHz, 66–76 GHz, and 81–86 GHz as presented in **Table 3** to study for possible identification to IMT (aimed for 5G) at WRC-19.

Europe identified 26 GHz as a *pioneer* frequency band for early European harmonization, as it provides over 3 GHz of contiguous spectrum and has the greatest potential to be a globally harmonized band.

The 31.8–33.4 GHz (referred to as 32 GHz) and 40.5–43.5 GHz (referred to as 40 GHz) bands were also identified as priority bands for study in the CEPT.

5. Examples of 5G usage scenarios

According to the recent ITU theoretical assessment, simulations, measurements, technology development, and prototyping described in the Report ITU-R M.2376-0, utilization of bands between 6 and 100 GHz is feasible for studied IMT deployment scenarios and could be considered for the development of IMT for 2020 and beyond [6].

It is expected that LTE will develop in a way of providing wide area coverage for mobile users. 5G networks will include LTE access based on OFDM along with new radio interfaces using possibly new techniques, e.g., filter bank multicarrier (FBMC) transmission technique.

5.1. IMT-2020 architecture

Deployment architecture of IMT-2020 can be classified into two architecture types: *stand-alone* or *overlay*. The stand-alone architecture refers to the network deployment consisting of millimeter wave (*mmWave*) small cells. The overlay architecture refers to the network deployment of mmWave small cells developed on top of the existing macro-cell networks (LTE, etc.). In overlay architecture case, the macro-cell layer of the existing 4G mobile communication serves mainly for providing coverage, whereas the mmWave small cell layer should be used to provide capacity [6] as shown in **Figure 4**.

For mmWave small cells explicated using cellular technologies, the typical service range is expected to be around 10 to 200 m under non-line-of-sight (NLOS) circumstances, which is a lot shorter than the range of a cellular macro-cell that can provide several kilometers. The small cells can be deployed both indoors (e.g., femto cells) and outdoors. When deployed outdoors, mmWave small cells are typically deployed at a lower antenna height than a macro-cell (on street lamp posts, on building walls, in parks, etc.) and with lower transmit power to cover a targeted area. For mmWave small cell deployment, scenarios can be identified three categories: indoor, hotspot, and outdoor [6].

http://dx.doi.org/10.5772/intechopen.72431

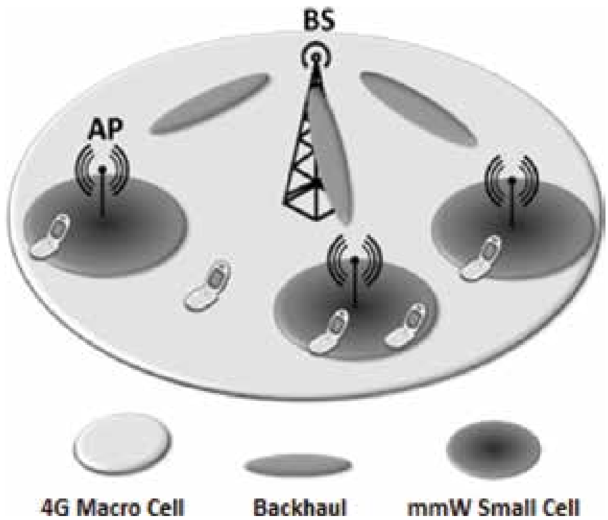

Figure 4. System deployment architecture proposed for 5G.

5.2. Channel bandwidth

For single-input/single-output (SISO) scenario, the maximum capacity of a radio channel can be described by Shannon-Hartley formula. This formula relates the maximum capacity (transmission bit rate) that can be achieved over a given channel to certain noise characteristics and bandwidth. For an *additive white Gaussian noise* of power the N, the maximum capacity can be calculated by

$$C = B \cdot \log_2\left(1 + \frac{S}{N}\right), \tag{1}$$

where C is the maximum capacity of the channel (bits(s)) otherwise known as *Shannon's capacity limit for the given channel*, B is the bandwidth of the channel (Hz), S is the signal power (W), and N is the noise power (W). The ratio S/N is named *signal-to-noise ratio* (SNR) [8].

It can be perceived that the maximum transmission data rate, at which the information can be transmitted without any error, is limited by the bandwidth, the signal level, and the noise level. With the increase in bandwidth, the noise power also increases; that is why the channel capacity does not become infinite. By using wider channels, increasing the number of antennas, and reducing interference, it is possible to increase the capacity.

One of the benefits of higher-frequency adaptation for mobile communications is system capability to implement wide channels.

In the authors' opinion, to achieve objectives set for future IMT-2020 systems, it is necessary to provide contiguous, broad, and harmonized frequency bands, which will minimize 5G device complexity and possible interference issues.

In the 5G era, FDD will remain the main duplex scheme for lower-frequency bands. However, for higher-frequency bands—especially above 10 GHz—targeting very dense deployments, TDD will play a more important role.

In author's opinion, 5G wireless access may be realized by the improved LTE systems for existing spectrum in combination with new radio access technologies that primarily target new spectrum [2].

5.3. Antenna technology

Antennas that can operate well enough in distant frequencies at the same time, e.g., at between 450 MHz, 700 MHz, and 26 GHz, are a difficult task. Therefore, most likely two separate antennas, each operating at the specific frequency band, will be required. The wavelengths above 24 GHz provide a possibility to put more antenna elements in the restricted area. The antenna technology with the increased number of particular antenna elements can be used to provide high beamforming gain. The incremented path loss of above 24 GHz frequency bands can be mitigated by beamforming techniques with exact pointing direction. The phased array beamforming is used to raise the received signal power by using beamforming gain. Greater antenna gains may be achieved applying narrower beams [4].

For 5G communication systems, massive MIMO (*multiple-input and multiple-output*) solutions would be used to compensate additional propagation loss in higher frequencies [2] and to minimize interference. Array antennas should be integrated in the terminals or user equipment. In this case, it should be possible since the transmission wavelengths would become smaller.

5.4. 5G development scenarios

The 5G radio access is based on the evolution of LTE and the other one on New Radio (NR) access. In the LTE-5G, enhancements will continue to enable it to support as many 5G requirements and used cases as possible. Unlike the LTE-5G, the NR-5G is free from backward compatibility requirements and thereby able to introduce more fundamental changes, such as targeting spectrum at high (mmWave) frequencies. However, NR is being designed in a scalable manner so it could eventually be migrated to frequencies that are currently served by LTE.

The process of making LTE-5G involves a variety of enhancements and new features in 3GPP Release 14 and Release 15. The most significant ones are enhancements to user data rates and system capacity with FD MIMO (*full-dimension multiple-input multiple-output*), improved support for unlicensed operations, and latency reduction in both control and user planes. FD MIMO is a technology that arranges the signals transmitted to antennas in the form of virtual beams that are able to power multiple receivers in three dimensions. It is expected that this technology significantly will increase spectrum efficiency.

In authors' opinion, LTE along with NR-5G will continue to play a major role in mobile communications for many years to come.

6. Electromagnetic compatibility studies with existent technologies

The 2019 World Radiocommunication Conference (WRC-19) will take place over 4 weeks, from 28 October to 22 November 2019. The Conference will address a number of questions

http://dx.doi.org/10.5772/intechopen.72431

within the radiocommunication sector, and consideration of spectrum for 5G above 24 GHz is expected to feature heavily through the Resolution of WRC-19 Agenda Item 1.13.

In the framework of WRC-19 Agenda Item 1.13, identification of frequency bands 24.25–27.5 GHz, 31.8–33.4 GHz, 37–43.5 GHz, 45.5–50.2 GHz, 50.4–52.6 GHz, 66–76 GHz, and 81–86 GHz for the future development of IMT will be considered, including possible additional allocations to the mobile service on a primary basis, in accordance with Resolution 238 (WRC-15) [9].

Compatibility studies must be carried out in each frequency band, and each band must balance with the requirement of other existing services and allocations. Any identification of frequency bands for IMT should take into account the use of the bands by other services (to protect existing services) and the evolving needs of these services. WRC-19 Agenda Item 1.13 states to conduct and complete in time for WRC-19 the appropriate sharing and compatibility studies, taking into account the protection of services to which the band is allocated on a primary basis.

Spectrum sharing will be an important element to facilitate the requirements of 5G, and the development of new technological solutions in higher-frequency ranges, such as the more extensive use of MIMO technique, may provide more opportunities for sharing but also challenges for National Regulatory Authorities (NRAs).

In the future, there may be a need to adapt the harmonized regulatory framework for 5G in the existing mobile frequency bands in Europe (700 MHz, 800 MHz, 900 MHz, 1500 MHz, 1800 MHz, 2.1 GHz, 2.3 GHz, 2.6 GHz).

7. Electromagnetic compatibility evaluation methods

Congestion of the radio spectrum is growing with an ongoing rise in the demand for more wireless services. National communications regulators are faced with the challenge of identifying new frequencies for new uses while preventing interference to existing users of the spectrum.

7.1. Radio frequency interference

Interference occurs when a transmission from one system disrupts the reception of signals at the receiver of another nearby system. It can occur between systems operating on the same frequency known as co-channel interference or between systems in frequencies that are close known as adjacent channel or adjacent band interference. It is worth noting that there are also other types of interference such as intermodulation [10].

Co-channel interference is a result of the stronger interfering signal, which affects the victim signal. In adjacent channel interference, there are two main causes of interference: *unwanted emissions* and *blocking* or *receiver selectivity* as presented in **Figure 5**.

Unwanted emissions are any off-channel noise of the interfering equipment falling within the receive band of the victim receiver and thus acting as co-channel interference to the wanted signal. This sort of interference can only be removed at the source.

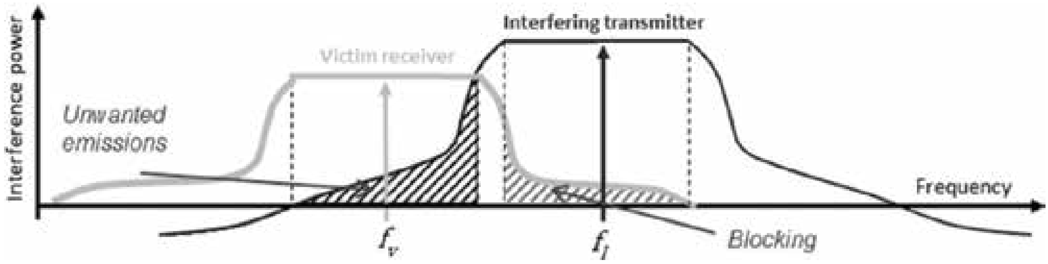

Figure 5. Causes of interference: unwanted emissions and blocking.

Blocking, i.e., a strong signal off the receive band of a victim receiver, desensitizes its reception. This sort of interference can only be removed at the victim. In most cases, adoption of power control for the interferer and efficient site engineering can improve the situation.

In practice, both of these can occur simultaneously. Sometimes, it is necessary to improve the design of both the transmitter and receiver to prevent interference, which is becoming increasingly important as the spectrum becomes more congested [10].

7.2. Interference prediction methods

Interference prediction typically is done through theoretical calculations known as sharing studies, which usually refer to in-band studies, and compatibility studies, which refer to adjacent band studies. Theoretical studies are necessary because it is not always possible to perform measurements on real systems, particularly in cases where the systems are still under development. Two types of studies are commonly used:

Deterministic studies based on fixed parameters, using the *minimum coupling loss* (MCL) method. This is a worst-case assessment of interference. The results usually determine the minimum required separation distance (in space or in frequency) between two systems to avoid interference. The MCL approach is relatively straightforward, modeling only a single interferer-victim pair. It provides a result that is spectrally inefficient.

Statistical studies based on variable parameters, using the *Monte Carlo* method. This is a more realistic assessment, which takes into account the real-world variation and randomization of certain parameters such as the relative positioning of systems. The result of a Monte Carlo simulation is a measure of system performance—it is commonly a probability of interference for the scenario under investigation, which can be compared against a relevant threshold to determine if the level of interference is considered to be a problem or not. Care must be taken when interpreting a probability of interference. A mobile system operator specifies that a system can provide a system availability of 95%. It is capable of modeling highly complex systems including LTE networks. The result is spectrally efficient but requires careful interpretation.

A mobile system operator specifies that a system can provide a system availability of 95%.

These studies can be performed using a range of software tools, for example, SEAMCAT for Monte Carlo analysis, which is an open-source software tool. In order to compare both methods, the same radiowave propagation prediction model should be adopted for all three methods.

In some cases, the assumptions used in the studies can be validated through laboratory measurements of real systems or through field measurements [10].

If the results of studies show that interference may occur, it may be necessary to investigate different mitigation techniques to minimize the risk of interference. This could include specification of additional filtering to be applied to the transmitter or receiver, additional frequency separation between both systems, and restrictions on the usage of the new system such as limits on maximum transmit power or geographical restrictions on where the new system can be used. In addition, with the emergence of new technologies, new and innovative sharing solutions are being explored, for example, techniques such as geolocation and licensed shared access (LSA).

This overall process, including the prevention of interference by quantifying the risk through studies and the cure through identifying suitable mitigation, is known as spectrum engineering [10].

Sharing studies need to have accurate input assumptions in order to produce meaningful and reliable results. Spectrum engineers are working with future technologies where not all the parameters can be defined in advance of the deployment of the new technology. Spectrum engineering results can be used to optimize frequency planning.

7.3. Protection criteria for mobile service

Different interference criteria can be used for interference assessment to the mobile service stations or other services, e.g., C/I, C/(N + I), (N + I)/N, or I/N.

The protection criterion for use in sharing and compatibility studies between IMT-Advanced and IMT-2020 and other systems and services irrespective of the number of cells and independent of the number of interferers is I/N value of −6 dB. This criterion applies to interference from a single source or to the aggregate interference from multiple sources of interference. The same protection criterion should be used for both co-channel and adjacent band studies [11, 12].

For the assessment of the interference of LTE and other services in 700 MHz band [13, 14], authors used both of these methods, namely, MCL and Monte Carlo method, and the criterion of $I/N = -6$ dB was used for interference assessment to the mobile service. In this assessment the predetermined trigger field strength values also was used. Additionally, some field measurements were also performed.

8. Conclusion

Global harmonization of IMT spectrum will be essential for developing 5G. The benefits of spectrum harmonization include facilitating economies of scale, enabling global roaming, reducing equipment design complexity, improving spectrum efficiency, and potentially reducing cross border interference. 5G mobile communication systems will require frequencies for their development and usage. Frequencies below 6 GHz are very valuable because of its optimum radio wave propagation, especially frequencies below 1 GHz. The results of the present study have shown that implementation of frequency bands even above 24 GHz remains needed to ensure all the performance targets of 5G, e.g., multigigabit per second data rates. In the authors' opinion, for the deliberative development of IMT systems, it is necessary to timely provide wide and contiguous spectrum resources for implementation of new technologies and services.

It is important to note that the properties of higher-frequency bands, such as shorter wavelength, would better enable the use of advanced antenna systems including MIMO and beamforming techniques in supporting enhanced broadband.

The electromagnetic compatibility between LTE and different existent technologies in 700 MHz band authors was evaluated with different methods: MCL method, Monte Carlo method, and predetermined trigger field strength values; some field measurements were also done. According to results of electromagnetic evaluation, additional mitigation techniques were proposed in order to assure the compatibility between considered radio systems. Similar electromagnetic compatibility evaluation methods and approach can be also applied for IMT-2020 studies in frequencies above 24 GHz.

The results obtained within the framework of the research can be used by National Regulatory Authorities, equipment manufacturers, mobile operators, researchers, and other interested parties when planning 4G and 5G mobile services.

Author details

Guntis Ancans* and Vjaceslavs Bobrovs

*Address all correspondence to: guntis.ancans@rtu.lv

Institute of Telecommunications, Riga Technical University, Riga, Latvia

References

[1] IMT Vision – Framework and overall objectives of the future development of IMT for 2020 and beyond. Recommendation ITU-R M.2083-0, September 2015

[2] Ancans G, Bobrovs V, Ancans A, Kalibatiene D. Spectrum Considerations for 5G Mobile Communication Systems. ICTE 2016. Procedia Computer Science. 2017;**104**:509-516

[3] 5G Automotive Vision. P. 36, 20-10-2015. [Online]. Available: https://5g–ppp.eu/wp-content/uploads/2014/02/3G.-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf

[4] Future spectrum requirements estimate for terrestrial IMT. Report ITU-R M.2290-0, pp. 10-14, December 2013

[5] Ancans A, Bogdanovs N, Petersons E, Ancans G. Evaluation of Wi-Fi and LTE integrated channel performance with different hardware implementation for moving objects. ICTE 2016, Procedia Computer Science. 2017;**104**:493-500

[6] Technical feasibility of IMT in bands above 6 GHz. Report ITU-R M.2376-0, July 2015

[7] 7G. Radio Access. Ericsson White paper, April 2016. [Online]. Available: https://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf

[8] Saunders SR, Zavala AA. Antennas and Propagation for Wireless Communication Systems. 2nd ed. John Wiley & Sons Ltd.; 2007

[9] Resolution 238 (WRC-15). Studies on Frequency-Related Matters for International Mobile Telecommunications Identification Including Possible Additional Allocations to the Mobile Services on a Primary Basis in Portion(S) of the Frequency Range between 24.25 and 86 GHz for the Future Development of International Mobile Telecommunications for 2020 and beyond. Switzerland, Geneva: International Telecommunications Union; 2015

[10] A comparison of the minimum coupling loss method, enhanced minimum coupling loss method, and the Monte-Carlo simulation. ERC Report 101, Menton, May 1999

[11] Characteristics of terrestrial IMT-Advanced systems for frequency sharing/interference analyses. Report ITU-R M.2292-0, December 2013

[12] Characteristics of terrestrial IMT systems for frequency sharing/interference analyses in the frequency range between 24.25 GHz and 86 GHz. SWG Work for TG 5/1, Revision 3 to Document 5D/TEMP/265-E, 22-02-2017

[13] Radio Regulations. International Telecommunications Union (ITU), vol. 1 (articles edited in 2016), 2016

[14] Ancans G, Bobrovs V, Ivanovs G. Frequency arrangement for 700 MHz band. Latvian Journal of Physics and Technical Sciences. February 2015;**52**(1):63

Chapter 4

5G Backhaul: Requirements, Challenges, and Emerging Technologies

Md Maruf Ahamed and Saleh Faruque

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.78615

Abstract

5G is the next generation cellular networks which is expected to quench the ever-ending thirst of data rates and interconnect billions of smart devices to support not only human centric traffic, but also machine centric traffic. Recent research and standardization work have been addressing requirements and challenges from radio perspective (e.g., new spectrum allocation, network densification, massive multiple-input-multiple-output antenna, carrier aggregation, inter-cell interference mitigation techniques, and coordinated multipoint processing). In addition, a new network bottleneck has emerged: the backhaul network which will allow to interconnect and support billions of devices from the core network. Up to 4G cellular networks, the major challenges to meet the backhaul requirements were capacity, availability, deployment cost, and long-distance reach. However, as 5G network capabilities and services added to 4G cellular networks, the backhaul network would face two additional challenges that include ultralow latency (i.e., 1 ms) requirements and ultradense nature of the network. Due to the dense small cell deployment and heavy traffic cells in 5G, 5G backhaul network will need to support hundreds of gigabits of traffic from the core network and today's cellular backhaul networks are infeasible to meet these requirements in terms of capacity, availability, latency, energy, and cost efficiency. This book chapter first introduce the mobile backhaul network perspective for 2G, 3G, and 4G networks. Then, outlines the backhaul requirements of 5G networks, and describes the impact on current mobile backhaul networks.

Keywords: 5G, mobile backhaul network, wired backhaul, wireless backhaul, microwave, millimeter wave, free space optics

1. Introduction

During the last few decades, mobile communication has evolved significantly from early wireless voice systems to today's intelligent communication systems [1, 2]. With the advancement of each generation, the mobile communication systems become more sophisticated and unleashed new consumer services that support countless mobile applications used by billions of people around the world, shown in **Figure 1** [1–3]. In 2000, when 3G brought us the wireless data, the consumers got access to the internet anytime and anywhere they go [2]. This mobile broadband network with a combination of the innovation of smartphone technologies brought a significant change of mobile internet experience where users can access their email, social media, music, high definition video streaming, online gaming, and many more, which we see today as the app-centric interface [4].

Due to the advancement of technologies, mobile devices getting smarter every day in terms of advanced computing and multimedia capabilities that supports a wide range of applications and services (e.g., high quality image transfer, ultrahigh definition video streaming, live video games, and cloud resources) [5, 6]. Therefore, more and more users are expecting to have the same quality of internet experience anytime, and everywhere they go. These trends will even more pronounced when 5G network becomes available with intelligent network capabilities and numerous services. 5G network will extend the wireless connectivity beyond the people, to support the connectivity for everything that may benefit from being connected that might include everything from personal belongings, household appliances, to medical equipment, and everything in between [1]. Numerous 5G network use cases, services and network

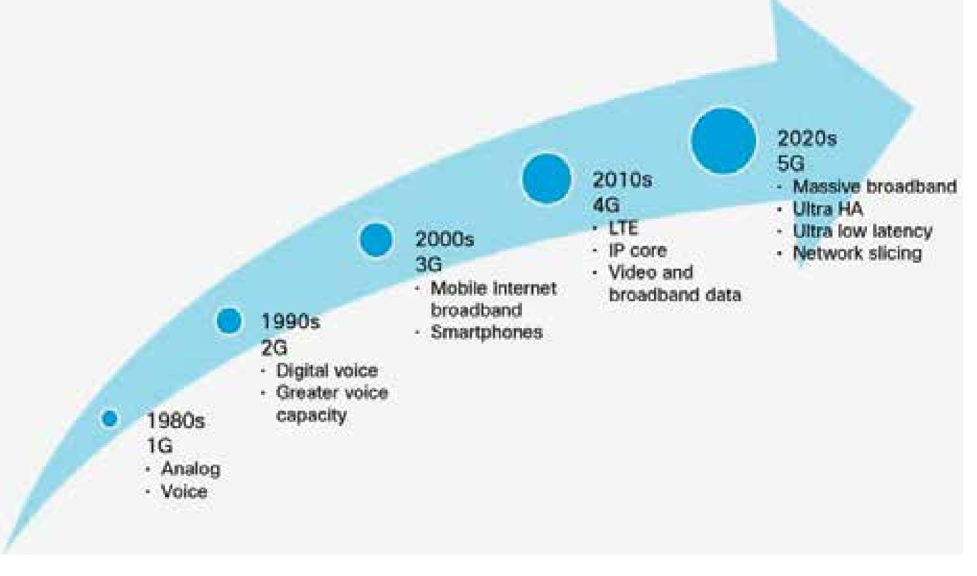

Figure 1. The evolution of mobile standards [3].

requirements are discussed in [7]. Here are the two most significant trends of 5G services are discussed below [6]:

- Everything will be connected to the mobile network wirelessly that will enable the billions of smart devices interconnected autonomously while ensuring the security and privacy [8]. 5G network will enable emerging services that include remote monitoring and real-time control of a diverse range of smart devices, which will support machine-to-machine (M2M) services and Internet of Things (IoT), such as connected cars, connected homes, moving robots and sensors [6, 8–10]. According to Cisco VNI Mobile 2017, the most noticeable growth will occur in M2M connections. The number of M2M connection will reach 3.3 billion by 2021, which is 29% of the total devices and connections and it was only 10% in 2016 [5]. Another mobile traffic forecast by UMTS presented that the total number of connected IoT devices will reach to 50 billion by 2020 which was only 12.5 billion in 2010 [5].
- 5G networks will deliver richer content in real time ensuring the safety and security that will make the wireless services more extensive in our everyday life. Some example of emerging services may include high resolution video streaming (4K), media rich social network services, augmented reality, and road safety [6]. According to Cisco mobile data traffic forecast, the maximum mobile data traffic will be generated by video-based mobile application, which is going to be 72% of mobile data traffic by 2019 compared to 55% in 2014 [5, 11].

It is clear that the future mobile network (i.e., 5G) will no longer human centric, it will be more on machine centric which will interconnect billions of smart devices to the mobile network. According to Cisco, smart devices are those that have advanced computing and multimedia capabilities with a minimum of 3G network connectivity [5]. Globally the growth of smart devices will reach 82% by 2021 and some regions it will reach 99% by 2021 (e.g., North America). The main impact of this growth will be on mobile data traffic because a smart device generates much higher traffic compared to non-smart device. According to Cisco forecast, a smart device generated 13 times more traffic compared to non-smart device in 2016 and by 2021 a smart device will be able to generate 21 times more traffic [5]. According to another mobile traffic forecast by Cisco, the expected growth will reach 24.3 Exabytes per month by 2019 which was only 2.5 Exabytes in 2014 [12]. This ever-increasing traffic growth becomes the key driver for the evolution of next generation mobile networks, called 5G, envisioned for the year 2020 [13, 14]. The key requirements of 5G network include, extreme broadband delivery, ultrarobust network, ultralow latency (i.e., less than 1 ms latency) connectivity, and support massive smart devices for the human and for the IoT services [15].

To bring the 5G network in reality, a simple upgrade of mobile network will not be enough where we just add new spectrum and enhance the capacity or use advanced radio technology. It will need to upgrade from the system and architecture levels down to the physical layer [15]. Although some research and standardization work addressing the corresponding challenges from radio perspective (e.g., new spectrum exploration, carrier aggregation, network densification, massive multiple-input-multiple-output, and inter-cell interference mitigation techniques) but there is a new challenge has emerged: the backhaul [16–18]. Because in 5G networks, the ultradense and heavy traffic cells will need to support hundreds of gigabits of traffic from the core network through backhaul and today's cellular system architecture is infeasible to meet this extreme requirement in terms of capacity, availability, latency, energy, and cost efficiency [16, 19].

A details survey about the evolution of mobile backhaul solution from 2G to 3G networks and from 3G to 4G networks are presented in [20, 21], respectively. The hybrid of millimeter wave and optical backhaul solution is proposed in [22] where the software-defined backhaul resource manager is proposed as a novel software defined networking (SDN) approach for realizing high utilization of the backhaul network capacity in a fair and dynamic way, while providing better end-to-end user quality of experience (QoE). Also, [23] provides another hybrid solution that combines an optical laser (through free space) and millimeter-wave radio to provide a combination of guaranteed high capacity, extended reach, and high availability with affordable cost.

If given a choice, fiber always remains the first backhaul choice for service provider due to its inhibitive bandwidths more than 10 Gbps and allowed maximum latency of hundreds of microseconds [16, 23]. But, laying fiber to connect all the cells to the core is not possible in some cases due to the availability problem and the deployment cost is high as well. In addition, fiber deployment, even when it is feasible can take several months [4, 23]. Since the massive deployment of small cells will be the key techniques for 5G networks and the backhaul requirements of the small cells can significantly vary with the small cell location, the fiber cannot be the optimal approach for 5G backhaul solution [23–25]. On the other hand wireless backhauling (e.g., microwave and millimeter wave) becomes popular due to its availability, deployment time and cost-effective approach [24]. But the weather condition and multipath propagation have significant impact on microwave and millimeter-wave radio systems which can affect the transmission performance. So it is obvious that there will no unique backhaul solution for 5G networks. The backhaul evolution for 5G networks will include both wired and wireless backhaul solution [23].

The contributions of this book chapter are listed below:

- First, this chapter provides a brief introduction of mobile backhaul network and the evolution of mobile backhaul network.
- Second, provides a comprehensive overview of backhaul requirements of 5G networks and highlight the potential challenges.
- Finally, it outlines the existing mobile backhaul solutions (i.e., wired and wireless) and list their features, benefits, drawbacks, application areas and deployment challenges.

2. Introduction to mobile backhaul network and evolution

The mobile backhaul network connects radio access network air interfaces at the cell sites to the inner core network which ensures the network connectivity of the end user (e.g., mobile phone user) with the mobile networks (shown in **Figure 2**). In **Figure 2**, UE refers to end user, eNodeB refers to cell or cell site or base stations. Each user data is added with other components of the backhaul traffic (shown in **Figure 3**), to calculate the single eNodeB transport provisioning and then aggregate with all other eNodeB's traffic before it connects with the core network.

It is found that the capacity requirements on the transmission network to support the backhaul traffic from the core network is raises with the evolution of mobile/cellular networks [23]. Cost and reliability always been a concern and major challenges for cellular network operators and there is no magic solution to the demand [19]. This section describes, how the mobile backhaul network evolve with the evolution of each mobile network (e.g., 2G, 3G, and LTE).

2.1. Typical GSM, 3G, and LTE Network Overview

A typical GSM (Global System for Mobile Communications) network architecture is shown in **Figure 4**, where the BTS (base station transceivers) are located at the cell site and provide the control and radio air interface for each cell. The BSC (Base station controllers) provides control

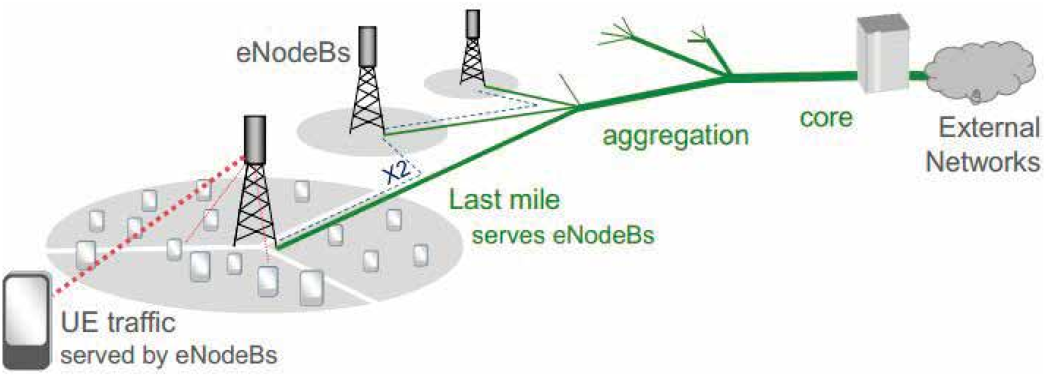

Figure 2. Mobile backhaul network of LTE (long-term evolution) [26].

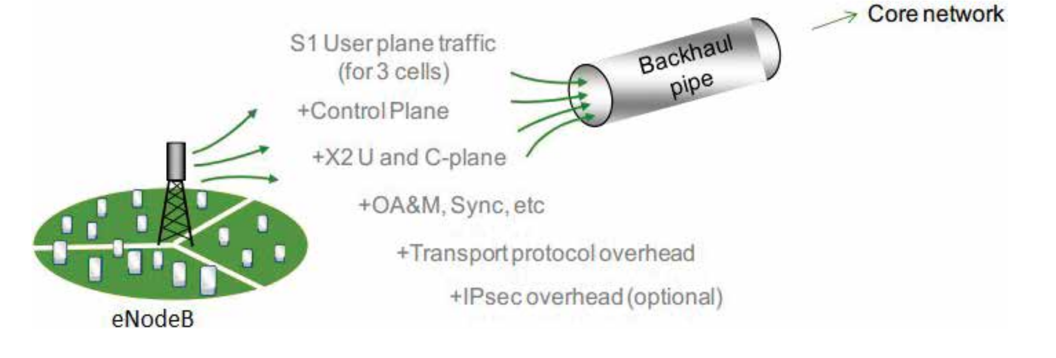

Figure 3. Components of backhaul traffic [26].

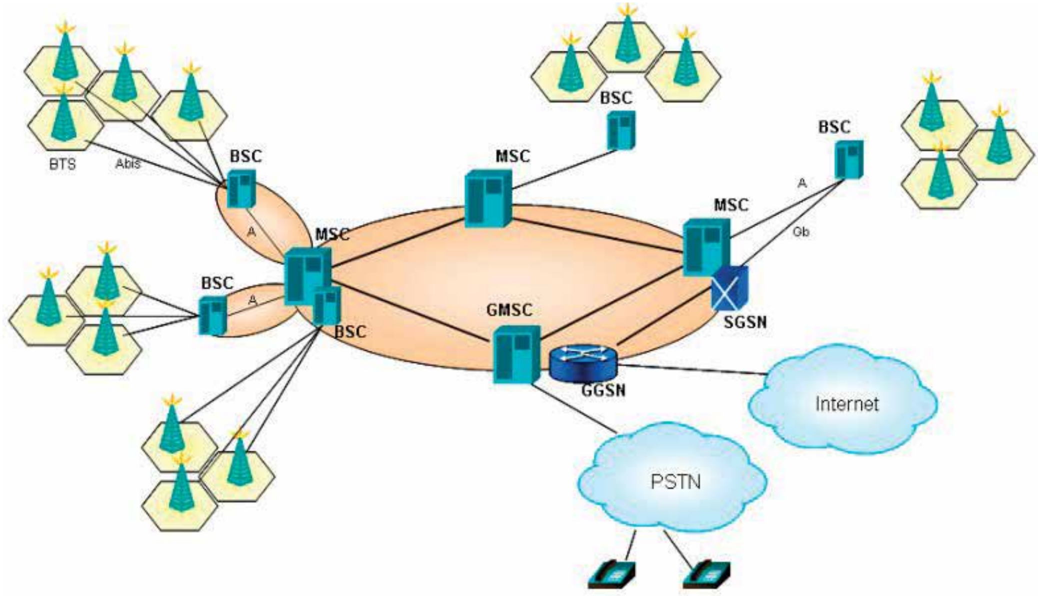

Figure 4. Typical GSM network with wireless interface requirements [27].

over multiple cell sites and multiple base station transceivers. The base station controllers can be located in a separate office or co-located at the mobile switching center (MSC).

There are standard interfaces developed by the wireless industry for interconnecting these devices, so they could deploy interoperable systems from multiple vendors. These physical interfaces define the wireless backhaul transport services and requirements. Thus, a basic understanding of these interfaces is very important. Some standard interfaces for GSM network are listed below [27]:

- Abis: the Abis interface connects the base station transceivers to base station controllers.
- A: the A interface in **Figure 2** connects the base station controller to the mobile switching center.
- Gb: voice services continue over the A interface, while data services are handled over the Gb interface.

Although the functions of these devices are similar, the 3GPP wireless standards body adopted slightly different names for the functional nodes and logical interfaces for UMTS (3G) networks, shown in **Figure 5**. But, historically the 2G/3G wireless standards were based on T1 (TDM) physical interfaces for interconnection between these devices because of the wide availability of T1 copper, fiber, and microwave services [28].

T1 physical interfaces has driven mobile backhaul transport requirements for 2G/3G wireless standards, but 4G wireless standards (i.e., LTE: Long-Term Evolution) are based on entirely new packet-based architecture, including the use of Ethernet physical interfaces for interconnection between the various functional elements, shown in **Figure 6**.

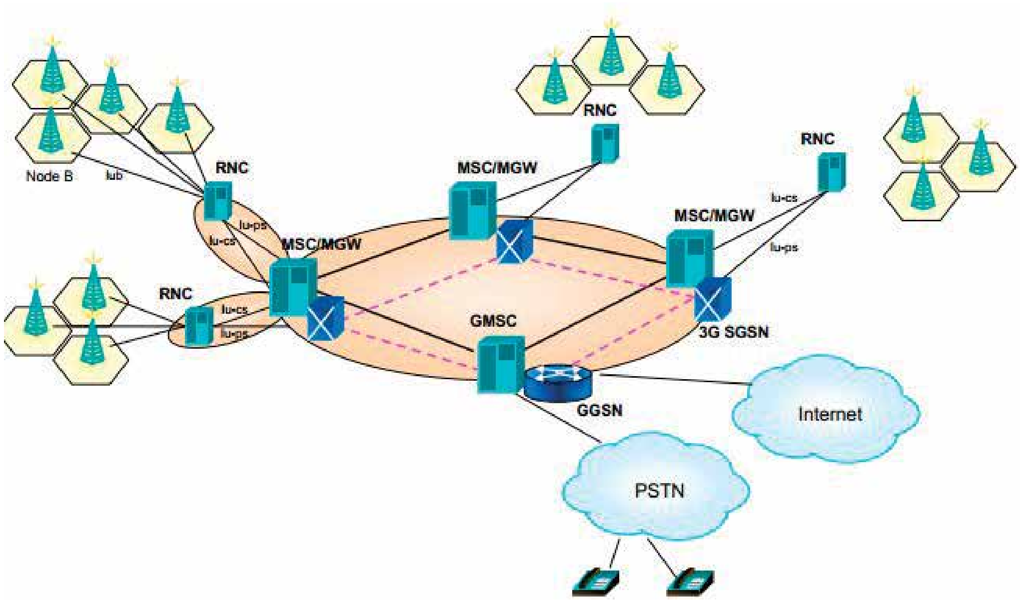

Figure 5. Typical 3G network with wireless interface requirements [27].

Another objective of the LTE standards was to flatten and simplify the network architecture. This resulted in pushing more intelligence into the radios (eNodeB) and elimination of the radio controllers as a separate device. In effect, the radio controller function has been distributed into

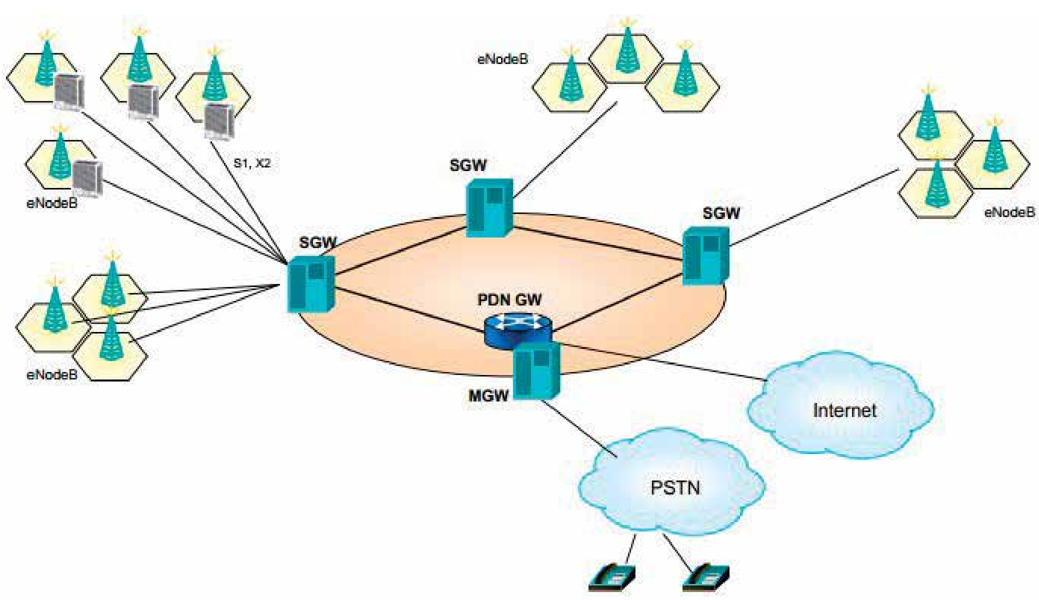

Figure 6. Typical LTE network with wireless interface requirements [28].

Standards	Voice spectrum (MHZ)	Data spectrum (MHZ)	Voice spectral efficiency (bits/Hz)	Data efficiency (bits/Hz)	# Sectors	Total bandwidth (Mbps)	# T1 s
GSM 2G	1.2		0.52		3	1.3	1
GSM/ EDGE 2.75G	1.2	2.3	0.52	1	3	6.1	4
HSDPA 3G		5	0	2	3	21.0	14
LTE 4G		5	0	3.8	3	39.9	n/a
LTE 4G		10	0	3.8	3	79.8	n/a

Table 1. Wireless capacity requirements for 2G, 3G and LTE [27, 28].

each eNodeB radio. So, the resultant network is indeed much simpler and flatter, with far fewer functional devices.

From a mobile backhaul perspective: most cell sites will continue to support GSM 2G and UMTS 3G networks for many years, the addition of LTE means backhaul transport carriers need to implement systems that can support both native T1 TDM services and Ethernet services. But, the major changes are the higher capacities required by LTE cell sites. A detail comparison of wireless capacity requirements for 2G, 3G, and LTE networks are shown in **Table 1**.

3. 5G backhaul requirements and challenges

Enhance the network reliability and reduce the cost efficiency always been a major challenge for the cellular network operator and there is no magic solution to that demand [19]. With the evolution of mobile network, the capacity requirement of the transport network from the core raises significantly [23]. The major backhaul challenges that mobile network operator had to deal with up to 4G network includes capacity, availability, deployment cost, and long distance reach [16]. But, 5G network will interconnect billions of new start devices with the numerous use cases and services, which will support machine-to-machine (M2M) services and Internet of Things (IoT) to the mobile network [6, 8]. These new smart devices will not only enhance the backhaul capacity requirement, but it will also add two additional challenges in the backhaul network: (a) ultralow latency of ~1 ms (round trip) connectivity requirements, and (b) denser small cell deployment. This section describes the 5G backhaul requirements and potential challenges.

3.1. Capacity

The evolution of 5G cellular network is positioned to address new services and demands for business contexts of 2020 and beyond [29]. It is expected that 5G network will enable a fully mobile and connected society that empower socio-economic transformation in many ways and even some of which are unimagined today. To fulfill the demand of fully mobile and connected society can be characterized by the tremendous increase in the number of connectivity and traffic volume density [29]. According to Nokia, the number of connecting devices per mobile users will be ten to one hundred that includes, mobile phone, laptop, tablet, smart watch to smart shirts [11]. In addition, the number of connected machines and sensor devices in the industry and public infrastructure will increase. According to Ceragon, the forecasted capacity increase could be $\times 1000$ compared to the capacity density in current 4G/4.5G networks [30, 31]. So, it is obvious that the evolution of 5G networks from LTE/LTE-A will need higher capacity backhaul links per cell site: while LTE/LTE-A networks need hundreds of Mbps, 5G network will need to support tens of Gbps, shown in **Figure 1**.

3.2. Availability

Availability is the major consideration for any backhaul networks, if the backhaul services are not in operation the system performance are negatively affected. In case of fiber systems, if there is any interruption of current path, the systems will automatically switch to the protection path within <50 ms [23]. Even in the wireless backhaul (e.g., microwave and millimeter wave), the backhaul link can be affected by multipath propagation and bad weaher condition. To overcome this, adaptive modulation technique is used that lowers the line rates to maintain the availability. Although the 5G network requirements is not standard yet, but to provide the expected new services such as autonomous vehicle/autonomous driving, tactile internet, and many machine to machine applications need high availability and very low latency [30].

3.3. Deployment cost requirements

Cellular network provider has to spend billions of dollars each year to acquire wireless spectrum for building excellent network coverage [23]. Since dense small cell deployment will be the key for 5G networks to support 1000 times more capacity, the cost efficient backhaul solution for the small will be a major challenge. An application specific traffic-engineering model needs to be formulated so that both customers and service providers can be happy.

3.4. Long distance reach requirements

Reach defines how far a cell site can get backhaul support from the core network with the required quality of service. Long distance reach is always a big issue for the backhaul network in terms of cost and additional equipment (e.g., total deployment cost of fiber backhaul will increase with the distance) [23]. Typically, cell sites are interconnected in a hierarchical mesh and all the traffics are transported back to an aggregation point (sometimes-called super cell) where all the traffics are aggregated and transport to the core network. Due to the dense small cell deployment in the 5G networks, massive backhaul traffic will be aggregated at the super cell that can create congestion and can even collapse the backhaul networks [19]. Therefore, long distance reach will be a big challenge for the 5G backhaul network.

3.5. Ultralow latency requirements

One of the major requirements of 5G network is ultralow latency ~1 ms (round trip) [31]. Some 5G use cases and services, such as real-time monitoring and remote control, autonomous driving, tactile internet, and M2M applications need to support by mission-critical network because this type of services will need high availability, ultralow latency and tight security. In addition, the risks of network failure are too high [30]. Therefore, it will be a big challenge for 5G backhaul to support massive traffic and maintain the required quality of service with lower latency requirement. Since propagation delay is inherent, a solution has to be formulated based on physical layer.

3.6. Ultradense network

Since, 5G will use higher RAN frequencies, the cell site coverage will become very small compared to today's cell site (i.e., macro or micro cell). It is also not feasible to increase the cell site capacity by 1000 times. Therefore, dense small cell deployment is the only efficient way to support 1000 time more capacity in 5G network [30]. This dense nature of the small cell grid will present the following challenges for 5G backhaul:

- Denser backhaul link due to the denser small cell grid will highly limit the frequency reuse, which will require better utilization of wireless backhaul spectrum [30].
- There will be some set to unprecedented requirements for cell site synchronization. According to the forecast, 5G network will need three times stricter accuracy requirements than LTE-A (i.e., 1.5 μs to approx. 0.5 μs) [17].

4. Available mobile backhaul solutions and key challenges

Small cell backhaul requirements (e.g., traffic load intensity, latency, target quality of service, and cost of implementing backhaul connections) can vary significantly depending on the locations of the small cells [24]. Besides, the available backhaul scenarios can greatly vary, for example, some places fiber connection may be available but it may not be available for other places. Some places may be good line of sight microwave connectivity but it may not be available for every places. So, there is no single technology that can dominate backhaul technology [25]. There are many options, and the operators have to decide which backhaul solution will be most economical for any particular deployment scenarios. This section describes available backhaul solution that can be used for 5G networks.

4.1. Wired backhaul solution

A compressive study of wired backhaul solution is presented in [16]. So this subsection describes the fiber backhaul connection only. Fiber is the most popular backhaul solution that can provides highest capacity with low bit error rate and it allowed highest reach before any signal needs to be retransmitted. That's why fiber is always remains as the first choice for backhaul if it is available. Unfortunately, fiber is not available for most of the places. For example, one study conducted in 2014 present a fact that the fiber backhaul is not available nationwide in Europe and the alternative microwave backhaul solution will not sustain the traffic growth of LTE/LTE-A beyond 2017–2018 [16]. There are some other instances as well, such as nearby backbone fiber does not exist and there is no right of way available, or there are some obstructions (e.g., highways or rivers or buildings) that will make the fiber connection impossible. In addition, the deployment of new fiber connection will take several months compared with wireless deployments that can be completed within a week. Initial deployment cost of fiber connection is also high that includes cable costs, splicing, trenching, right of way, etc., plus the cost of equipment's for the optical transport and aggregation [23]. So, it is obvious that the fiber connection will be first choice for backhaul if it is available otherwise we have to look for other backhaul solution that meets the requirements of 5G networks.

4.2. Wireless backhaul solutions

Support the backhaul traffic from radio switch to cell site wirelessly (i.e., wireless backhaul) becomes popular due to the viability, and cost-efficiency. Wireless backhaul (e.g., microwave, millimeter wave) allows operator to have end-to-end control of their network instead of leasing third party wired backhaul (e.g., fiber) connections. However, the optimal selection among wireless backhaul solution depends on several factors that includes cell site location, propagation environment, desired traffic volume, interference conditions, cost efficiency, energy efficiency, hardware requirements, and the availability of spectrum [24]. This subsection describes the available wireless backhaul solution that can be used for small cell backhaul in 5G networks.

4.2.1. Microwave

Wireless backhaul technology supports approximately 50% of mobile traffic globally where, microwave RF technology has the vast majority [23]. It typically operates in the 6, 11, 18, 23, and 28 GHz frequency bands. The deployment of microwave radio requires one-time capital cost and there is some other costs that includes space/power rental and maintenance. So, microwave RF can be deployed at a much lower cost and the deployment time is much faster compared to fiber connection. However, the performance of microwave RF backhaul solution significantly varies with the propagation environment and bad weather condition. Often the system need to lower the transmission rate to maintain the availability requirements. Typically microwave RF can support up to 500 Mbps capacity and the maximum reach could be 30 miles which may also vary with the applications. For examples, if we use lower frequency band we can achieve longer reach but lower frequency bands are congested and may not be available for backhaul solution. On the other hand, if we use higher frequency we can have higher date rate but the system has to sacrifice the long reach. Microwave RF system capacity can be increased up to 10 Gbps for long and medium distance connectivity by utilizing wider channel spacing (e.g., 112 MHz for traditional microwave bands 4–42 GHz), higher order modulation schemes (e.g., 4096 QAM and up), and the use of ultrahigh spectral-efficiency technique (e.g., line of sight MIMO) [30].

4.2.2. Millimeter wave

According to International Telecommunication Union (ITU), millimeter wave operates in Extremely High Frequency (EHF) which is range in 30–300GHz. So the millimeter wave RF can become prominent for small cell backhaul solution due its enormous spectrum [32]. Besides, smaller wavelengths will enable the integration of large number of antennas in a simple configuration and small cell can take this advantage of massive MIMO for LOS or non-LOS backhaul solutions. Typically, millimeter wave RF can support high data rate up to 1–2 Gbps range but the reach is shorter compared to Microwave RF due to the high propagation loss at millimeter wave [16]. The major factor that causes propagation loss includes, absorption, rain fading, and multipath propagation. Typically, the millimeter wave beams are much narrower compared to microwave beams that creates a major constraint which is line of sight (LOS) backhaul solution. Narrow beam can also create alignment problem and to avoid this problem, millimeter wave RF equipment's need to be installed in a solid structure [23].

4.2.3. Free space optics (FSO)

Free space optics (FSO) is a line of sight backhaul technology which is similar as fiber optics except FSO uses invisible beam of light (e.g., LED, LASR) for data transmission instead fiber cable [33, 34]. FSO has enormous spectrum in the range of 300 GHz to 1 THz and the maximum data rate can support up to 10 Gbps (both upstream and downstream) [16]. The

Technology	Deployment cost	Latency	Reach	Upstream throughput	Downstream throughput	Options
Satellite	High	300 ms one-way latency	∼Ubiquitous	15 Mbps	50 Mbps	LOS
TVWS	Medium	10 ms	1–5 km	18 Mbps/ch	18 Mbps/ch	NLOS
Microwave PtP	Medium	<1 ms/hop	2–4 km	1 Gbps	1 Gbps	PtP
Microwave PtmP	Medium	<1 ms/hop	2–4 km	1 Gbps	1 Gbps	PtmP
Sub-6 GHz 800 MHz–6 GHz	Medium	5 ms single-hop one way	1.5–2.5 km urban, 10 km rural	170 Mbps	170 Mbps	NLOS
Sub-6 GHz 2.4, 3.5, 5 GHz	Medium	2–20 ms	250 m	150–450 Mbps	150–450 Mbps	NLOS
MmWave 60 GHz	Medium	200 μs	1 km	1 Gbps	1 Gbps	LOS
MmWave 70–80 GHz	Medium	65–350 μs	3 km	10 Gbps	10 Gbps	LOS
FSO	Low	Low	1–3 km	10 Gbps	10 Gbps	LOS

Table 2. Wireless backhaul solutions [16, 23, 33].

http://dx.doi.org/10.5772/intechopen.78615

major advantage of FSO is low power consumption compared to microwave or millimeter wave RF. But FSO system has number of constraints that includes LOS communication, fading due to fog, interference due to ambient light, scattering and physical obstructions. However, FSO technology can be one of the possible solution for 5G backhaul due to its scalability and flexibility [33].

Although there are some other wireless backhaul solutions (e.g., Satellite and TV White Spaces (TVWS)), but maximum data rate can support is less than 1Gbps and the latency requirement is also high. This is the main reason not to add much details about this two backhaul technologies. The available features (e.g., latency, reach, and throughput) of all wireless backhaul technologies are summarized in **Table 2**.

As it is seen from **Table 2**, FSO provides highest throughput with lowest latency which is the basic requirements of 5G backhaul. This is the main motivation of this paper where FSO is used for 5G backhaul networks with addition of ambient light cancelation technique at the receiver, described in Section 4.

5. Conclusions

According to the use cases, services, and network requirements of 5G, the next generation mobile network will not only human centric. This network will allow to connect new type of devices that support machine-to-machine (M2M) services and Internet of Things (IoT). Therefore, the backhaul network must meet diverse network requirements based on the type of user traffic, such as, some user traffic may care more about the maximum network speed not latency and on the contrary the users may care about the low latency not the speed. So, it is obvious that there will be no unique backhaul solution for 5G networks. Based on the deployment area and user traffic, the 5G backhaul network will be a combination of wired (e.g., fiber) and wireless backhaul (millimeter wave, and free space optics). Thus, understanding the basic backhaul network requirements is the key to choose the right technology and type of network. In an effort, this book chapter first introduce the backhaul network perspective for 2G, 3G, and 4G networks and then outlines the backhaul requirements of 5G networks. This chapter also describes the available backhaul solutions and describes the key challenges.

Author details

Md Maruf Ahamed[1,2]* and Saleh Faruque[2]

*Address all correspondence to: ahamedm@uwplatt.edu

1 University of Wisconsin-Platteville, Platteville, USA

2 University of North Dakota, Grand Forks, USA

References

[1] Dahlman E, Mildh G, Parkvall S, Peisa J, Sachs J, Selén Y. 5G radio access. Ericsson Review. 2014;**91**:42-48

[2] Sawanobori TK. The Next Generation of Wireless: 5G Leadership in the U.S. Washington, DC, USA: CTIA; 2016

[3] Cisco 5G Vision Series: Laying the Foundation for New Technologies, Use Cases, and Business Models. Cisco; 2016. http://www.atis.org/5G2016/media/C11-736764-00_Cisco%205G%20Vision%20Series_WP_v3a.pdf

[4] Understanding 5G: Perspectives on future technological advancements in mobile. GSMA Intelligence. December; 2014. https://www.gsmaintelligence.com/research/?file=141208-5g.pdf&download

[5] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021. Cisco; 2017. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[6] Docomo 5G White Paper. NTT Docomo, Inc; 2014. https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_5g/DOCOMO_5G_White_Paper.pdf

[7] Ericsson White Paper: 5G Systems. Ericsson; 2017. https://www.ericsson.com/assets/local/publications/white-papers/wp-5g-systems.pdf

[8] 5G Vision. DMC R&D Center, Samsung Electronics Co., Ltd; 2015. http://images.samsung.com/is/content/samsung/p5/global/business/networks/insights/white-paper/5g-vision/global-networks-insight-samsung-5g-vision-2.pdf

[9] Palattella MR, Dohler M, Grieco A, Rizzo G, Torsner J, Engel T, et al. Internet of things in the 5G era: Enablers, architecture, and business models. IEEE Journal on Selected Areas in Communications. March 2016;**34**:510-527

[10] Skouby KE, Lynggaard P. Smart home and smart city solutions enabled by 5G, IoT, AAI and CoT services. In: Presented at the 2014 International Conference on Contemporary Computing and Informatics (IC3I); Mysore, India

[11] 10 Key Rules of 5G Deployment-Enabling 1 Tbit/S/km2 in 2030. Nokia Networks; 2015. http://info.networks.nokia.com/10_key_rules_of_5g_deployment.html

[12] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update: 2016-2021. Cisco; March 28, 2017. https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[13] Pierucci L. The quality of experience perspective toward 5G technology. IEEE Wireless Communications. 2015;**22**:10-16

[14] Soldani D, Manzalini A. Horizon 2020 and beyond: On the 5G operating system for a true digital society. IEEE Vehicular Technology Magazine. 2015;**10**:32-42

[15] 5G masterplan—five keys to create the new communications era. 2015. Available: https://pages.nokia.com/5g-master-plan.html

[16] Jaber M, Imran MA, Tafazolli R, Tukmanov A. 5G backhaul challenges and emerging research directions: A survey. IEEE Access. Vol. 4. 2016. pp. 1743-1766

[17] Hannula M. Getting in Sync with Mobile Backhaul and the Challenges of 5G. IL, USA: Coriant; 2016

[18] Thomas S. Poll: Backhaul Presents 5G's Biggest Challenge. Light Reading; 2016. http://www.lightreading.com/mobile/5g/poll-backhaul-presents-5gs-biggest-challenge/a/d-id/715304

[19] Ge X, Cheng H, Guizani M. 5G wireless backhaul networks: Challenges and research advances. IEEE Network. 2014;**28**:6-11

[20] Raza H. A brief survey of radio access network backhaul evolution: Part I. IEEE Communications Magazine. 2011;**49**:164-171

[21] Raza H. A brief survey of radio access network backhaul evolution: Part II. IEEE Communications Magazine. 2013;**51**:170-177

[22] Bojic D, Sasaki E, Cvijetic N, Wang T, Kuno J, Lessmann J, et al. Advanced wireless and optical technologies for small-cell mobile backhaul with dynamic software-defined management. IEEE Communications Magazine. 2013;**51**:86-93

[23] Mavrakis RKD. Laser-Radio Provides New Alternative for Backhaul. Ovum; 2015. http://www.ovum.com/laser-radio-provides-new-alternative-for-backhaul/

[24] Siddique U, Tabassum H, Hossain E, Kim DI. Wireless backhauling of 5G small cells: Challenges and solution approaches. IEEE Wireless Communications. 2015;**22**:22-31

[25] Bovberg J. Overcoming Backhaul Challenges. Fierece Wiereless; 2016. https://pages.questexweb.com/rs/294-MQF-056/images/Overcoming_Backhaul_Challenges_FINAL.pdf

[26] Guidelines for LTE Backhaul Traffic Estimation. NGMN Alliance; 2011. https://www.ngmn.org/fileadmin/user_upload/NGMN_Whitepaper_Guideline_for_LTE_Backhaul_Traffic_Estimation.pdf

[27] Understanding Mobile Wireless Backhaul. Fujitsu Network Communications Inc; 2009. http://www.fujitsu.com/downloads/TEL/fnc/whitepapers/Fujitsu_Wireless_Backhaul.pdf

[28] 4G Impacts to Mobile Backhaul. Fujitsu Network Communications Inc; 2009. https://www.fujitsu.com/us/Images/4Gimpacts.pdf

[29] NGMN 5G White Paper. NGMN; 2015. https://www.ngmn.org/fileadmin/ngmn/content/images/news/ngmn_news/NGMN_5G_White_Paper_V1_0.pdf

[30] Cohen D. What you Need to Know about 5G Wireless Backhaul. CERAGON; 2016. https://www.ceragon.com/solutions/mobile-backhaul/5g-wireless-backhaul

[31] Gupta A, Jha RK. A survey of 5G network: Architecture and emerging technologies. IEEE Access. 2015;**3**:1206-1232

[32] Orainy AAA. Wireless backhauling for 5G small cell networks. International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering. 2016;**10**:267-270

[33] Kehr W. Free Space Optics. University of Missouri-Rolla; 2002-2004. http://web.mst.edu/~mobildat/Free%20Space%20Optics/

[34] Dahrouj H, Douik A, Rayal F, Al-Naffouri TY, Alouini MS. Cost-effective hybrid RF/FSO backhaul solution for next generation wireless systems. IEEE Wireless Communications. 2015;**22**:98-104

Chapter 5

Radio Access Network Backhauling Using Power Line Communications

Francesco Marcuzzi and Andrea M. Tonello

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.73390

Abstract

Nowadays, radio access networks (RANs) are moving toward a small cell-based paradigm, where the macro-cell antennas are aided by additional ones with lower coverage capabilities. This paradigm shift brings a new problem into the equation: a backhaul link is needed to carry the traffic from the small cell base stations to the network gateway. Currently, both wired and wireless solutions exist, but none is universally considered optimal. Power line communications (PLC) can be considered as a broadband access solution for this backhaul branch. Recent developments helped to push PLC performances to a point where state-of-the-art solutions can achieve very high-speed data transfers. Aided by traffic generation-based simulations, we will show how the PLC technology can be assessed for the described application. The reader will be guided through the process by discussing small cell networks, the power line infrastructure and basics of traffic generation modeling. The chapter will then discuss a quality-of-service (QoS)-driven analysis and use numerical results to show how requirements can be defined for the backhauling technology. Overall, this chapter will address how PLC and small cell network technologies can be brought together in a unified model to foster future small cell technology.

Keywords: power line communications, small cell networks, backhauling, optimization, broadband access, random channel modeling, stochastic geometry, transmission line theory, traffic generation

1. Introduction

Nowadays, Internet connectivity is a consumer good, and its usage is growing faster and faster. A recent report forecasts that the worldwide mobile data traffic will clock in at 71 exabytes per month in 2022, while in 2016 this figure equalled about 9 EB/month [1]. It also predicts that by 2022 five out of nine billion of mobile subscriptions will be LTE (in 2016 this value was 1.5 out of 7.5), and that video services will generate 75% of the amount of mobile data traffic (in 2016 it was 50%). This growth in the amount of traffic is causing a shift in the way we design networks and the technologies used for their deployment. In fact, other sources predict that new-generation network technologies will see a manifold increase in the strictness of the requirements on capacity, latency, energy consumption, cost and mobility [2, 3].

In the world of mobile communications, connectivity is mainly deployed through what are technically called radio access networks (RANs). A RAN is a part of a system of telecommunications: it is the conceptual intermediary between the mobile end-user, which is the person requesting service through a user device, and the core network, which enables the exchange of data from all the endpoints of the network. A RAN is often implemented through a system of high-powered antennas which partition the geographical territory in areas of a few kilometers of diameter each. These are referenced to as macro-cells, because of their capacity of coverage. RANs are widely spread to enable provision of communication standards as GSM, EDGE, UMTS and LTE. On the other hand, there are also other solutions to provide mobile connectivity.

Small cells are low-power, high-capacity radio access nodes that improve directly the spatial fragmentation and frequency reuse of the network and indirectly local capacity and global coverage. They also bring the network closer to the user, thus helping to save energy and stretching end-device battery life. Small cells can be characterized based on their capacity of coverage, both in terms of distance and number of simultaneous users, into the following broad families: microcells, picocells and femtocells. The latter, which are the smallest and less powerful of the bunch, are used mostly for in-home, private purposes, because of their low cost and high capacity [4], and are the kind of cells that will be considered throughout this work. Small cells are strategically deployed in environments where the network coverage generated by the macro-cell is weak, in order to bring the said network closer to the final user.

This technology introduces the necessity for a backhaul segment to bring data from the small cells to the RAN access points (network gateways) and back further to the core network. Solutions for this link can be wired or not; popular choices are optic fibers, ADSL on twisted pair or radio connections. Furthermore, this additional element in the network's hierarchy needs to be designed properly, taking into account the end-user requirements, and granting easy deployment and no use of end-user spectrum are all necessary action points in network planning [5].

Power line communications (PLC) is a technology that uses the same medium both for power supply and for data connectivity. An immediate advantage of this technology is the fact that there is no need for a new infrastructure, leading to savings in deployment time and costs. Despite the power cables not being designed for data transmission, they are currently employed both for low data rate (e.g., meter reading, alarm signals, emergency communications) and for data-hungry (e.g. in-home HD video streaming) applications [6, 7]. New advancements in the PLC research boosted the communication capacity, even beyond gigabit per second in the in-home environment. While PLC promises easy deployment, its performance for backhauling applications still needs to be thoroughly assessed.

To the best of our knowledge, never-before PLC was regarded as a broadband solution for RAN backhauling. Considering how femtocells are nowadays used inside buildings and how power is fed to this premises through the power line infrastructure, this could be an opportunity to better integrate PLC and provide general connectivity in the small cell environment. To analyze the problem, the contribution in this chapter is organized as follows.

In Section 2 the rationale of the wireless small cell networks, the power line infrastructure paradigm and the traffic generation methods used for the first part of the assessment will be explained.

In Section 3 the overall network structure will be described. A bottom-up transmission line model will be used to simulate the physical PLC channel between network nodes.

Section 4 will explain how, on top of the previous results, an optimization problem can be formulated to perform resource allocation. This also enables smart routing for improved connectivity of far cells.

Section 5 will be dedicated to showing and commenting results from the simulations carried out for several system setups.

Eventually, conclusions will be drawn in Section 6, along with comments on possible future applications and research paths.

2. Setting up the network: the physical layer

The network to investigate comprises small cells and power lines in a joint infrastructure. This section specifically aims at finding out the requirements that a PLC network needs to satisfy to be able to support a small network of private small cells. To do so, the following main aspects have to be addressed: (a) the small cell network deployment on the geographical territory, (b) the power line infrastructure and how it topologically connects the radio base stations (c) the traffic generated by the cells that needs to be backhauled.

Most of the material presented in this section is a synthesis of [8, 9].

2.1. Small cells

Small cells are low-power, high-capacity radio nodes, controlled either by an operator or a private customer. As mentioned before, small cells can be categorized into the following classes:

- Microcells: their coverage can reach up to a few kilometers and is used in large premises such airports or transportation hubs. They enable macro-cell base station controllers to manage the power in the network while optimizing spectrum usage.
- Picocells: these can cover up to a few hundred meters. While being employed for medium-sized venues (malls, train stations and the like), they also enable smoother handoffs of end-users between different macro-cells.

- Femtocells: lowest coverage of all the classes, it clocks in at a few tens of meters. They provide advantages both to operators and end-users, by bringing the network closer to the devices and improving battery life and quality-of-service (QoS). Femtocells can usually support 4, 8 or 16 simultaneous users.

While the first two categories are generally deployed and managed by operators, femtocells are directly controlled by the private user. It is legit to assume that, in a residential neighborhood, connection to the core network is provided in each building through one of these femtocells, meaning that there is a one-to-one relationship between cells and buildings in our model.

The next generation of cellular technology envisions a very different paradigm for networks, where small cells will play a fundamental role in assisting the macro-cells in providing the service in blind spots or where the QoS needs to be improved; in fact, a great increase in the density of small cell is forecasted for the coming years. Our model inquires the possibility of providing the connection to the small cells through simulations where the density of such elements is adjustable, thus allowing the model to relax or tighten the capacity requirement on the PLC technology.

In order to simulate the envisioned network, the first thing needed is a deployment of small cells. To do this, we employ the stochastic geometry tool [10], as to create a random dispersion of small cells, based on their individual coverage and global density, which is indirectly controlled through the portion of geographically territory to be actually covered. Stochastic geometry is a tool that simulates spatial averages and has been recognized as valid and powerful in the wireless communication world; herein, it is applied to generate and study the spatial distribution of PLC nodes (residing in the radio base stations) yielding, when wiring is considered, a random topology model whose idea was already presented in [11].

In our model, base stations are deployed through the following random processes:

- The number of base stations to be deployed is determined through random Poisson's process, whose average depends on the area that each cell covers (it is supposedly constant for every base station) and the quantity of geographical territory to be covered.
- A two-dimensional coordinate plane simulates our geographical territory; each cell's position is determined as a pair of uniform random variables.

A certain degree of overlapping is also allowed between different cells in order to account for neighboring houses or even cells in the same building but located on different floors. **Figure 1** shows a random deployment of cells. The small red circles represent the connected area reached by each base station located in their centers. The served area is the same for all the cells. To implement the specific deployment in **Figure 1**, a 20% value of covered territory was set.

2.2. Power line infrastructure for the small cell network

Once the small cells are deployed, the supporting power line infrastructure is created. Some assumptions are made to develop the model, which is based on the European paradigm

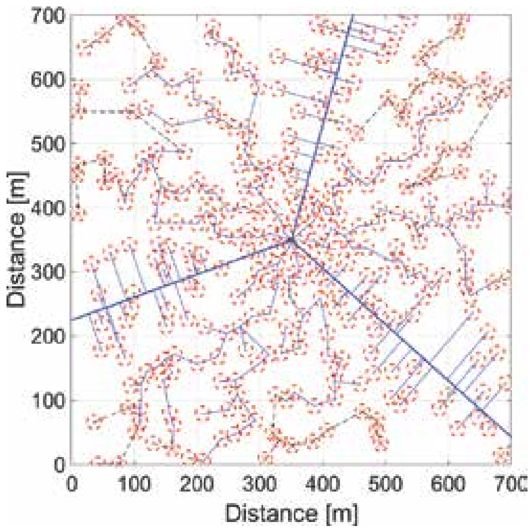

Figure 1. A random network deployment is shown, as produced by the implemented simulation software. The dashed circles represent small cells, while the continuous lines represent power cables from the supply infrastructure.

presented in [6]. This basically means the following things: the network is assumingly structured in a radial architecture, whose branches are all departing from a central concentrator that acts as a front-haul gateway (FHG), which may be located in a medium-voltage (MV) to low-voltage (LV) transformer station. In **Figure 1** the FHG is shown as a triangle in the center of the area. This is where high-speed data links are assumingly realized towards the core network, so the LV network of power lines might enable the backhaul portion of the network, between the private femtocells and the FHG.

On average, from each MV/LV transformer station, a maximum of ten branches are deployed to feed the end-users. Each one of these branches supplies a maximum of 30 to 35 buildings and extends to a maximum distance of 1 km from the FHG, measured in power line cable length. The LV network is the most geometrically complex part of the supply network: the model described in the following stems from observation and real-life cases described in literature [6].

In our model, different topologies for the power line infrastructure are realized to depict different geometrical properties of real networks. The bus topology drives a central backbone through the territory and connects the nearby cells following a minimum policy regarding the global amount of cable used; then, it draws branches to connect the cells to the backbone. The tree and chain topologies connect directly the cells to each other, preferring, respectively, a branching or serial connection policy. The only condition in the creation of these two structures is to avoid intersections of the power line medium between different segments.

While this completes the description of the physical layout of the network, the mobile traffic that is generated needs to be addressed and evaluated to define a requirement on the PLC backhaul technology.

2.3. Traffic generation models

We assume that small cells enable two different types of traffic: voice and data. Based on [4], we assume the probability of each type of call to be 3% for voice and 97% for data. We consider Poisson's distribution for the number of connection requests performed by the population of end-users; this, however, requires to know beforehand the average number of calls per time unit. This parameter is tuned in the model to account for different traffic loads in the considered networks.

Voice traffic is analytically treated as a phone call: 64 kbps are allocated to the user for the whole duration of each call, which follows an exponential distribution. This makes the whole statistic of voice traffic requests Poissonian. On the other hand, data traffic is trickier to model, as amount of data and duration of each communication are variable and co-dependent. This can be modeled through self-similar processes, such as the Pareto and Weibull distributions, which implement the well-known 80/20 rule [12]. Data to model the related probability distribution functions were retrieved from the same paper.

The developed network manages the traffic and resources on the PLC branch under analysis. It is implicitly assumed that resource management of the radio users inside the femtocells is carried out by the cell computational unit itself, and the same goes for the front-haul gateway (FHG) for the traffic to be routed towards the core network.

Results regarding the requirements on the PLC technology are reported in Section 5.1.

3. Transmission line-based bottom-up statistical PLC model

Transmission line theory allows to describe the behavior of a cable used as a telecommunication channel by modeling its medium as a circuit and including effects due to the propagation of short-length waves [6]. This method takes into consideration the nature of the transmission medium by means of:

- Propagation parameters, which allow to describe how signals are attenuated and distorted along the power lines based on the cable type
- Geometry of the cables, meaning how they are actually deployed on the territory based on length and direction
- Topology of the network, how cells are connected to each other and how their presence introduces interference through signal reflection on the channel

Translating this theory into a mathematical model enables bottom-up analysis of the network, as it starts from its lowest layer (physical) and builds models upon it to retrieve typical performances in a deterministic fashion. On the other hand, top-down approaches use real data measurements to infer the possible behavior of the network. As this application for broadband outdoor PLC, to our best knowledge, is not yet described in literature through statistical data, the former type of analysis was preferred.

While geometry and topology are defined in the first part of the random network generation, propagation parameters are generated afterwards based on the type of cable that is chosen to represent the network.

The method herein employed is the same described in [11]. Basically, the medium whose channel behavior is to be calculated is split into different elementary sections, and for each one of these sections, a transfer function is calculated. Sections of the medium are delimited either by loads (private premises) or discontinuities (change of cable type). Since the power line channel is not symmetrical, it is important to identify the direction of the signal that is being transmitted and whose attenuation profile we want to obtain. This is due to the fact that the transfer function of a certain section depends on all the previous sections passed by the signal. In this phase of the experiment, we consider the cells as transmitting towards the FHG. We consider the MV/LV transformer station to completely separate the channels of different sectors; thus, each one of the main branches departing from the station accounts for its own set of channel responses.

The channel under analysis sweeps the frequency domain between 1 MHz and 30 MHz, as this is the range of current PLC applications, although performances at higher frequencies are being explored [13]. Each private premise is modeled as an impedance whose value is statistically chosen in a uniform interval between 5 and 200 ohms, as the frequencies of the broadband channel impedances tend to drop to very low magnitudes (order of a few ohms). Further information on impedance modeling for broadband PLC channels can be found in [14]. The channel is also assumed to be time invariant for simplicity; in reality the channel in PLC is LPTV (linearly periodic time variant), although mostly at low frequency and only in the presence of strongly time variant loads connected closed to the transmitter or receiver (**Figure 2**).

The channel is evaluated for each cell of the network and each one of the 100 KHz subchannels. The evaluation of the channel is a computation of the chain of transfer functions between the receiving node and the transmitting one, which returns a channel gain as a ratio between the magnitudes of voltages at the ports of the two nodes. When combining this data with the transmitting power of nodes and the power spectral density of the noise, it is possible to retrieve the upper bound of capacity of the physical link through the Shannon formula. By using this method, it is implicitly assumed that inputs and noise behavior are Gaussian, which

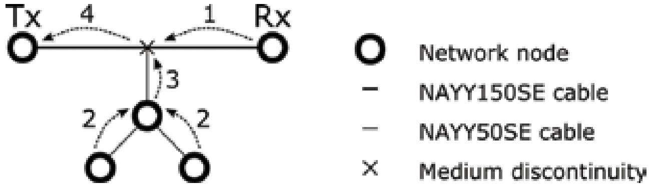

Figure 2. The channel transfer function is computed with the voltage ratio approach method. Starting from the receiving node, each section voltage ratio is used to retrieve the transfer function of that portion of the channel. Whenever a branch is found on the backbone connecting the receiving and the transmitting nodes, it is necessary to compute the transfer function related to this part of the network external to the communication path in order to model the reflections and attenuation it introduces.

is not actually the case of PLC channels, but the results can still be used as a frame of reference to a very well-documented case in literature.

$$C_{TOT} = \sum_{i=1}^{N} B_i \log_2(1 + SNR_i) \tag{1}$$

3.1. Remarks

As mentioned, this transmission line analysis allows to characterize the transfer function of the channel for each cell based on the power line distance of the cell from the FHG, the local density of loads connected to the power line infrastructure and the power spectral density of the noise on the channel.

This transfer function represents the channel gain of the signals transmitted between two points as a chain of voltage ratios, which can be used to compute the Shannon link capacity. The average channel gain (ACG) is the value of channel gain average over the whole spectrum considered for data transmission [15] and is thus strongly correlated to the value of this capacity upper bound calculated with Shannon's formula. By creating a large database of simulations through this model, it is possible to let a node infer the capacity of the medium simply by measuring the intensity of the noise on the cables, thus reducing the computational overhead load needed to map the channel and create an adequate scheduling. This would allow a node (the FHG in a downlink situation, a private cell in the uplink) to infer the capacity between itself and a receiver without sounding the medium.

4. Resource allocation and optimization

Each link between the FHG and a generic cell is characterized by a capacity value. This allows the server of the network, located in the FHG, to allocate the time resource to all the users that in each instant require a certain amount of throughput to connect to the core network.

The MAC protocol is assumed to be a TDMA as the channel is assumed time invariant; nevertheless, the time resource is divided into time frames as the PLC channel displays an LPTV (linear periodically time variant) behavior to simplify the integration of frequency variability in a more advanced stage of the study. The time frame naturally is meant to represent the duration of the mains cycle in an AC power supply network. As a frame of reference, in the European paradigm, this duration equals 20 ms.

Resource allocation schemes are applied to traffic requests that are generated over the network according to the models and metrics discussed in Section 2.3. Whenever one of the small cells generates a connection request, a throughput requirement is created and added for the whole duration of the call to the throughput already required during the said time range. As mentioned before, connection requests can be of two kinds, specifically voice or data. The former kind requires a continuous connection and a fixed throughput of 64 kbps for the whole duration of the call, while the latter generates a bigger quantity of traffic not necessarily requiring continuity.

Initially, the network's performance is evaluated by operating a very simple TDMA protocol: during each time frame, the time frame is equally shared between the pools of users requesting access to the medium. If the allocated resource completely covers the required throughput for one or more users, then a smaller chunk is assigned to these users, and the remainder of the resource is shared accordingly.

The TDMA method described above is ideal as it allows a continuous allocation of the time resource, while real communication standards divide the time frame into slots that can be distributed to users. This is due to the fact that the time resource must be split into a finite number of units to respect the packet structure of data transmission. HPAV [16] envisions a duration of the elementary OFDM symbol in the tens of microseconds, which is quantitatively adapted here to divide each frame into 500 elementary symbols. For the optimization problem presented in the following, the time frame is divided into time slots in order to satisfy a system where the number of users and the magnitude of their throughput requests in each time frame are not fixed. Each time slot consists of an integer number of elementary symbols such that the total number of symbols is eventually covered by an integer number of time slots.

The aim of the optimization problem is structured in the following way: first, all the users requesting resource to satisfy calls of the voice kind are assigned enough resource to cover this throughput. When voice calls are satisfied, the rest of the throughput requests are considered, and the rest of the resource is allocated according to the following:

$$\begin{aligned}
&\max_{\alpha, N_s} \sum_{k}^{N_U} \sum_{j}^{N_S} \alpha^{(k,j)} C_k(N_S), s.t. \\
&(i) \sum_{k}^{N_U} \alpha^{(k,j)} = 1, j = 1, 2, \ldots, N_S \\
&(ii) \max_p \sum_{j}^{N_S} \alpha^{(k,j)} C_k(N_S) \geq \frac{p}{100} C_k, k = 1, 2, \ldots, N_U
\end{aligned} \tag{2}$$

where N_S and N_U equal, respectively, the number of slots in which the time frame is divided and the number of base stations operating an access request during the current time frame, $\alpha^{(k,j)}$ is the logic operator that equals 1 when the j^{th} time slot is assigned k^{th} base station and C_k is the theoretical capacity of the link between the FHG and the k^{th} base station in case where the whole resource is assigned to the said base station, while $C_k(N_S)$ is the capacity that the k^{th} base station gets from a time slot when the time frame is split into N_S slots. The p parameter allows the equalization of the resource, and it is maximized in order to allow each base station to exploit at least p (%) of the capacity of the link towards the FHG. Condition (i) makes sure that each time slot is assigned to only one user. Condition (ii) on the other hand assures that each user is assigned enough resource to exploit a minimum percentage (p/100) of their actual capacity. The aim of the optimization process is to maximize the aggregated throughput in

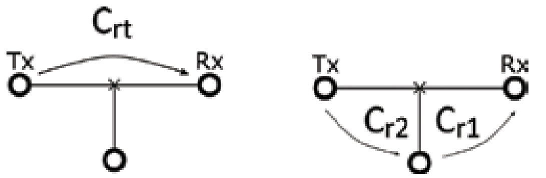

Figure 3. Routing enables a better coverage of the cells far from the FHG by decoding or amplifying the message received from the transmitting node before forwarding it to the receiver. In the network depicted on the left, the central cell is only creating interference and reflections on the PLC medium, while on the right is tentatively considered as a repeater for routing purposes.

each sector under a condition of resource scarcity by changing the overall number of time slots and the number of slots assigned to each user.

4.1. Enabling routing

The optimization problem can also be updated to include routing in it. A node can be chosen to act as a repeater based on the capacity of the channel between the FHG, the repeater and the cells that would be reached by the extended coverage. This also requires to distinguish between uplink and downlink cases because of the asymmetry of the PLC channel.

Simple routing can be enabled by assigning the role of repeater to one of the cells of the network, depending on the typical throughput required by a generic cell and the capacity of the channel between the end-user cell and the FHG with or without the repeater.

Figure 3 shows a case where only one cell is considered for routing-enabled extended coverage. This technique can actually be used to reach multiple cells in the region far from the FHG. Firstly, for each cell to be tentatively covered with routing, it is necessary to consider if the capacities created by a hopped communication give a better performance than the original one:

$$\begin{aligned} C_{DT} &= C_{rt} \\ C_{DF} &= \min\{\tau C_{r2}, \tau C_{rt} + (1-\tau)C_{r1}\} \end{aligned} \tag{3}$$

where C_{DT} and C_{DF} represent, respectively, the power line link capacity in the direct transmission case and the one where Detect-and-Forward strategy is used in the relaying node, as per **Figure 3**. Also, τ represents the fraction of the time frame that is dedicated to a specific connection. It is possible to adapt the problem to a case where the repeater needs to cover multiple cells. Basically, optimization is operated by tuning the aforementioned parameters so that the resource is used at its best. More details about this can be found in [17].

5. Results

In this section, numerical results from the simulations are presented. All the simulated networks consider small cells with a round coverage area whose surface is constant across the whole territory, while the simulation area (visible also in **Figure 1**) is square shaped. The length of the side equals the supposed average distance between neighboring FHGs. Cells that are further than a kilometer away (considering power line distance) from the FHG are considered as not connected neither to the communication nor to the supply network. This set of geographical parameters enabled the representation of both sparsely populated areas (as rural environments) and strongly occupied ones (as urban settings) through the tuning of the density parameter. By varying gradually this parameter, it is possible to see how a change in population affects the requirements imposed on the PLC technology and its overall performance. In Section 3 it was mentioned that the type of cable affects the behavior of the network: in our case, two types of cable were used to describe the network, specifically NAYY150SE for the LV lines that depart from the transformer station and NAYY50SE to connect this main bus to the point of connection of private premises. Furthermore, for the resource allocation problem, only cells with a nominal capacity over a certain threshold are considered as covered by service. It is also supposed that the FHG is transmitting with a power spectral density of −50 dBm/Hz and that the channel is characterized by a Gaussian noise with a power spectral density having a floor of −140 dBm/Hz. To evaluate the performance of the network, the grade of service (GoS) will be used. This is defined as the ratio between the throughput supported (ST) by the technology and the one required (RT) by the aggregated users across all the base stations:

$$GoS = \begin{cases} 1, RT \geq ST \\ \dfrac{ST}{RT}, otherwise \end{cases} \tag{4}$$

With the model described in Section 2, it was possible to assess preliminarily how the cells generate traffic and how they define a requirement for the PLC technology to enable broadband access. It was found that, when generated traffic resembled models of real cellular networks, the magnitude of this traffic was compatible with the highest capacity found on the PLC medium in outdoor applications. For a more detailed report, please refer to [18].

As mentioned before, transmission line theory enables bottom-up analysis of the network performance. In **Figures 4** and **5**, it is possible to see how the channel varies its behavior based on frequency and the power line distance between the communicating nodes; this behavior changes for different densities of cells deployed on the geographical territory.

Figure 4 shows how the channel gain behaves at different frequencies and power line lengths of analyzed links for a fixed density. In this specific case, the density was tuned in order to obtain an average of 35 cells for each one of the main branches departing from the FHG: this is reportedly [6] the maximum number of private premises fed by one main LV bus in the European paradigm. As in the Japanese/American paradigm, this number is much lower; it made sense to consider the European one as it sets the strictest conditions on the capacity of the channel. The underlying scatter plot shows the actual real channel gain of cells at different frequencies and distances; from the clear pattern these data create, a second-order interpolation was extracted in order to easily describe the behavior.

Figure 5, on the other hand, shows how the capacity of a generic link between a cell and the FHG can be correlated to the ACG of the same link. As seen in **Figure 4**, it is possible to infer the ACG of a link by knowing the distance of a cell from the FHG and the density of cells in the territory. Furthermore, it is easily possible to see how the ACG strongly depends on the length

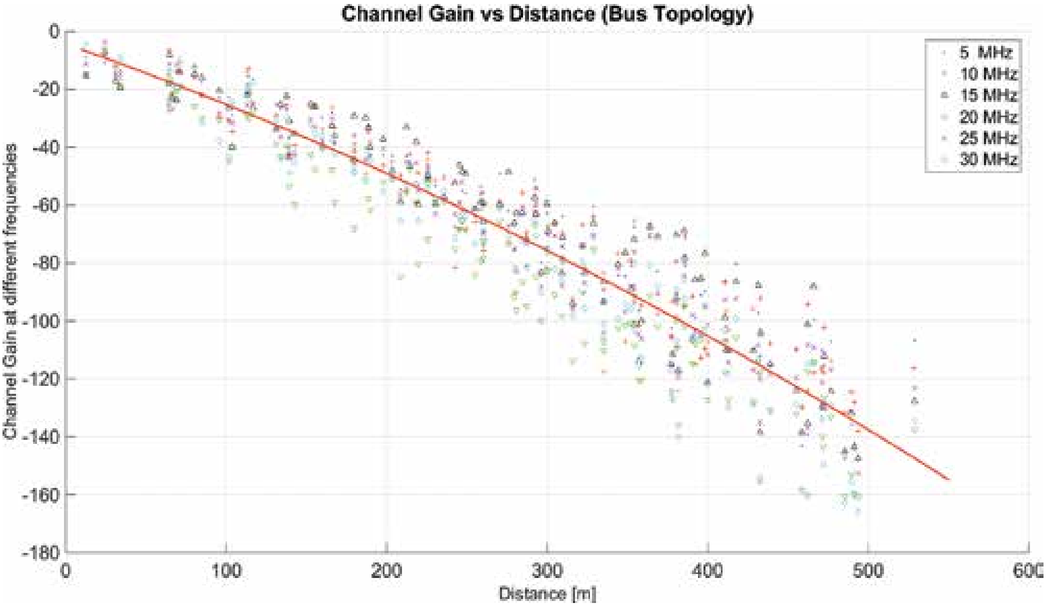

Figure 4. The scatter plot shows the channel gain computed with Shannon's formula and transmission line-based techniques in relation to the power line distance of the considered cells. A pattern is very clear; thus, a second-order interpolation was driven to retrieve the central red curve, which represents the ACG. This depends on the density of cells on the territory.

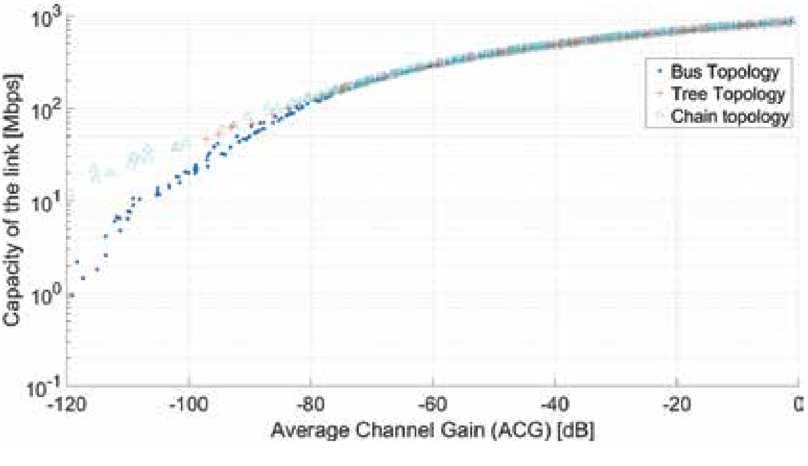

Figure 5. This semilogarithmic scatter plot shows the relation between the capacity of a link and the ACG attributed to the same. This relation does not depend on the distance; thus, it is possible to use it to infer a cell's capacity towards the FHG by measuring the ACG of the link.

of medium a signal must go through to enable communication. In **Figure 6**, the FHG (represented with a triangle in the center of the network) is always considered the transmitting node, thus enabling a downlink kind of communication. As mentioned in Section 3, it is possible to correlate this value with the capacity of the link between the FHG and the generic cell, when a few parameters of the network are known, such as distance of the communicating cell, density of the cells and power spectral density of the noise on the medium. Analysis from further simulations revealed that capacity correlated to the ACG neither depend on distance nor density of cells in the network. With reference also to **Figures 4** and **5**, it is possible to see that for a high-density network, coverage can reach almost 400 meters of power line length. In situations of lower cell density, cells that are further from the FHG can be covered by the PLC connection service. It follows that PLC coverage depends on distance also in relation to the density of base stations in the territory and their average coverage, i.e., the number of loads connected to the same shared medium.

Figures 7 and **8** report the grade of service (GoS) and the average aggregated throughput on the main buses departing from the FHG for the evaluation of the time resource assignment. The first one represents the network's performance when a simple TDMA protocol is employed, namely, each transmitting user gets a time slot whose duration is inversely proportional to the total number of transmitting users, while the second one shows the same values when the optimization process is operated. The average aggregated throughput is calculated as the sum of the effective throughput carried on the medium. Each user contributes to this figure either with the minimum value chosen between the throughput assigned by the FHG and the throughput required to satisfy the users in the cell. On the other hand, the grade of service is how much the throughput required by the users in the cell is actually satisfied: 1

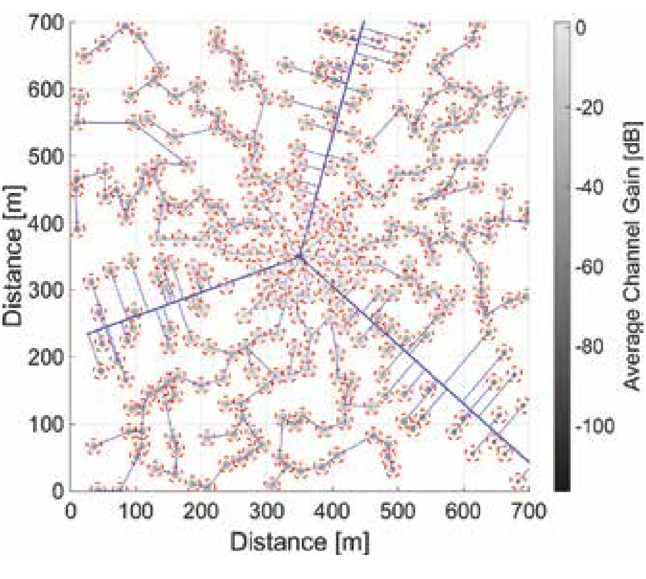

Figure 6. In this randomly generated network, a map of colors is used to show how the ACG fares in comparison with topology and distance of the cells from the FHG. As mentioned in section 5, cells whose ACG is circa −120 dB or higher can achieve a capacity of about 1 mbps or more.

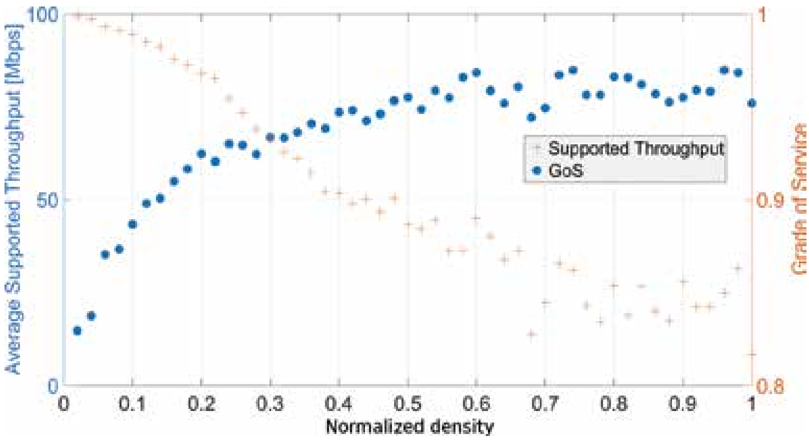

Figure 7. This graph shows the average supported throughput on each main bus departing from the FHG and the grade of service in relation to the normalized density of cells in a network where a simple TDMA protocol is applied.

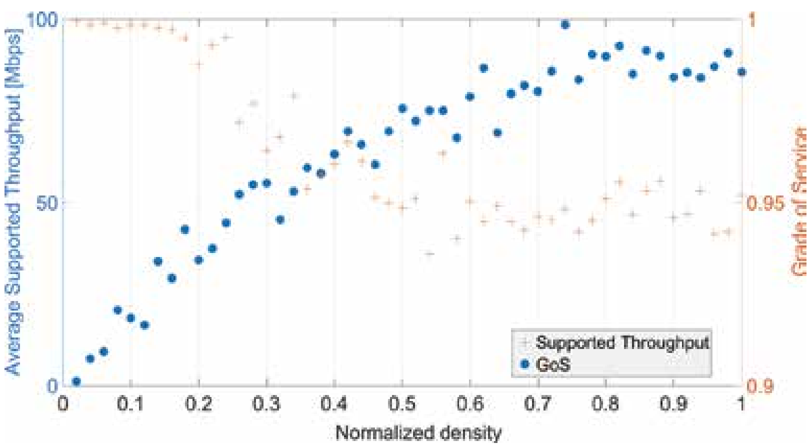

Figure 8. This graph shows the average supported throughput on each main bus departing from the FHG and the grade of service in relation to the normalized density of cells in a network where the optimization of the resource is carried out.

accounts for a network where all the users are satisfied, whereas 0 represents the opposite condition.

Also, for each density step considered for the simulations, the p factor mentioned in the definition of the optimization problem is reported to show how much resource is on average assigned to the requesting users to satisfy the generated calls. Its trend can be observed in **Figure 9**. This value is higher for low-density values because of the low number of users that request connections to the FHG.

From **Figures 7** and **8**, it is possible to see that the GoS is behaving better for the cases where the optimization is carried out. On the other hand, for lower densities, optimization seems to decrease the average supported throughput, probably due to the granular nature of the time

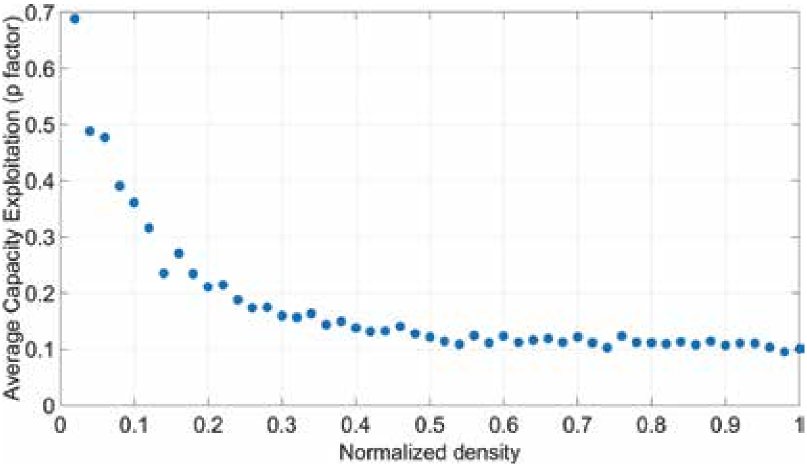

Figure 9. When the optimization is carried out on the time resource in a network, the p factor actualizes how much of the nominal capacity of the link towards the FHG each cell is able to exploit, based on how much time resource is assigned to each user.

resource in the optimization case versus the continuous one assumed in the simple TDMA case used for the former graph.

6. Conclusions

PLC is regarded nowadays as an enabling technology for IoT, smart grid and smart home applications. This study explores the possibility of using PLC as a broadband access solution for radio access networks by exploiting traffic generation models, transmission line theory and an optimization process to show how a TDMA protocol can be improved upon to let a central coordinator element schedule the time resource in an appropriate manner.

The experiments carried out on the simulation level showed how it is possible to employ PLC as a technology for broadband access in radio access networks. Studies carried out in the past showed which requirements were put on the shared medium, whereas here a general model for PLC performance was developed to infer which performance the technology is fit to provide. Specifically, herein it was shown that the technology is able to support connectivity for a small network of private femtocells covered by a LV power supply network, ranking as an adequate last mile solution. It was also shown how this performance can be enhanced for high-density networks by employing an optimization method, aiming at maximizing the supported throughput.

Moreover, it is mentioned in additional remarks (Sections 3.1 and 4.1) how the techniques used in the study will be exploited more in the future to gain a deeper insight into how hybrid broadband PLC-radio access networks can be designed to offer a better connection service to the end-users that populate them, namely:

- The transmission line theory model will be employed in conjunction with the topological and geometrical information to develop a system to infer the capacity of the link based on these information, which do not require a sounding of the channel, thus decreasing overhead and enhancing data speed.
- By using the method above, it will also be possible to consider different methods of routing inside the network to extend the coverage provided by the FHG. This will involve considering the combined capacities of the links that are created by partitioning the network. This will also depend on the type of link enabled, be it uplink or downlink.

A more realistic model could be developed by including different cell coverage sizes while also considering a more realistic behavior of users in the base stations via a traffic generator based on inferred statistics. Future endeavors will be devoted to experimentation in the field.

Author details

Francesco Marcuzzi* and Andrea M. Tonello

*Address all correspondence to: francesco.marcuzzi@aau.at

Networked and Embedded Systems Institute, Alpen-Adria-Universität Klagenfurt, Austria

References

[1] Cerwall P, Jonsson P, Carson S, editors. Ericsson Mobility Report (June 2017). 1st ed. Heuveldop N; 2017. p. 36

[2] KMS H, Rodriguez J, editors. Backhauling/Fronthauling for Future Wireless Systems. 1st ed., Chichester, United Kingdom: Wiley; 2016. p. 232

[3] Osseiran A, Boccardi F, Braun V, Kusume K, Marsch P, Maternia M, et al. Scenarios for 5G mobile and wireless communications: The vision of the METIS project. IEEE Communication Magazine. 2014;**52**(5):26-35. DOI: 10.1109/MCOM.2014.6815890

[4] Quek TS, de la Roche G, Güvenç İ, Kountouris M, editors. Small Cell Networks: Deployment, PHY Techniques, and Resource Management. 1st ed., Cambridge, United Kingdom: Cambridge University Press; 2013. DOI: 10.1017/CBO9781139061421

[5] Small Cell Forum, editor. Backhaul Technologies for Small Cells: Use Cases, Requirements and Solutions. Small Cell Forum; 2013. p. 78

[6] Lampe L, Tonello AM, Swart T, editors. Power Line Communications: Principles, Standards and Applications from Multimedia to Smart Grid. 2nd ed., Chichester, United Kingdom: Wiley; 2016. p. 618. DOI: 10.1002/9781118676684

[7] Cano C, Pittolo A, Malone D, Lampe L, Tonello AM, Dabak AG. State of the art in power line communications: From the applications to the medium. IEEE Journal on Selected Areas in Communications. 2016;**34**(7):1935-1952. DOI: 10.1109/JSAC.2016.2566018

[8] Tonello AM, Marcuzzi F. Statistical assessment of PLC networking for front-hauling in small radio cells. In: WSPLC 2016; Oct 10–11 2016; Paris, France. IEEE; 2016

[9] Marcuzzi F, Tonello AM. Stochastic geometry for the analysis of small radio cells and PLC back-hauling. In: SCC 2017; 11th International ITG Conference on Systems, Communications and Coding; Proceedings of; 6–9 Feb. 2017; Hamburg, Germany. VDE; 2017

[10] ElSawy H, Hossain E, Haenggi M. Stochastic geometry for modeling, analysis, and design of multi-tier and cognitive cellular wireless networks: A survey. IEEE Communications Surveys & Tutorials. 2013;**15**(3):996-1019. DOI: 10.1109/SURV.2013.052213.00000

[11] Tonello AM, Versolatto F. Bottom-up statistical plc channel modelling - part I: Random topology model and efficient transfer function computation. IEEE Transactions on Power Delivery. 2011;**26**(2):891-898. DOI: 10.1109/TPWRD.2010.2096518

[12] Han Y, Seo S, Hwang C, Yoo J, Won-Ki Hong J. Flow-level traffic matrix generation for various data center networks. In: Network Operations and Management Symposium (NOMS), 2014 IEEE; 5–9 May 2014; Krakow, Poland. IEEE; 2014. DOI: 10.1109/NOMS.2014.6838394

[13] Versolatto F, Tonello AM. PLC channel characterization up to 300 MHz: Frequency response and line impedance. In: GLOBECOM 2012; Global Communication Conference; Proceedings of; 3–7 Dec 2012; Anaheim, CA, USA

[14] Pittolo A, Tonello AM. A synthetic MIMO PLC channel model. In: ISPLC 2016; 20th International Symposium on Power Line Communications and its Applications; Proceedings of; 20–23 Mar 2016; Bottrop, Germany

[15] Tonello AM, Versolatto F. Bottom-up statistical PLC channel Modelling - part I: Inferring the statistics. IEEE Transactions on Power Delivery. 2012;**25**(4):2356-2363. DOI: 10.1109/TPWRD.2010.2053561

[16] Tonello AM, Cortes JA, D'Alessandro S. Optimal time slot design in an OFDM-TDMA system over power-line time-variant channels. In: ISPLC 2009; 13th International Symposium on Power Line Communications and its Applications; Proceedings of; 29 Mar-1 Apr 2009; Dresden, Germany

[17] D'Alessandro S, Tonello AM. On rate improvements and power saving with opportunistic relaying in home power line networks. EURASIP Journal on Advances in Signal Processing. 2012;**2012**:194. DOI: 10.1186/1687-6180-2012-194

[18] Marcuzzi F, Tonello AM. Stochastic geometry for the analysis of small radio cells and PLC back-hauling. In: SCC 2017; 11th International ITG Conference on Systems, Communications and Coding; Proceedings of; 6–9 Feb 2017; Hamburg, Germany

Chapter 6

Co-Channel Interference Cancellation for 5G Cellular Networks Deploying Radio-over-Fiber and Massive MIMO Beamforming

Sheng Xu

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72727

Abstract

In future fifth-generation (5G) wireless cellular networks, distributed massive multiple-input multiple-output (MIMO) techniques will be applied worldwide. Recently, much more challenges on efficient resource allocation to large numbers of user equipment (UE) are raised in order to support their high mobility among different micro-/pico-cells. In this chapter, we propose a framework to enable an optical back-haul cooperation among different optical network units (ONUs) with distributed MIMO techniques in wireless front-haul for next-generation optical-wireless cellular networks. Specifically, our proposal is featured by a downlink resource multi-cell sharing scheme for OFDMA-based passive optical network (PON) supporting radio-over-fiber (RoF). We first consider system architecture with the investigation of related works, and then we propose a co-channel interference mitigation and delay-aware sharing scheme for real-time services allowing each subcarrier to be multi-cell shared by different active ONUs corresponding to different micro-/pico-cells. Furthermore, a heuristic algorithm to mitigate co-channel interference, maximize sharing capacity, and minimize network latency is given by employing the graph theory to solve such sharing problems for future 5G. Finally, simulations are performed to evaluate our proposal.

Keywords: co-channel interference, 5G, distributed MIMO, passive optical networks, OFDMA, radio-over-fiber

1. Introduction

The increasing demand of real-time services (e.g., VoIP, the video telephony, and streaming) poses high requirements on communication quality (e.g., interference mitigation and delay constraint) and bandwidth increase in the network for the future era of big data [1]. Nowadays, the OFDMA-based passive optical network (PON) has been applied to provide such a large-capacity and also high-flexibility solution for wireless cellular networks with radio-over-fiber (RoF) technology [2, 3].

Figure 1 describes a RoF-based optical-wireless system adopting OFDMA-PON, while one of prominent challenges in such networks for future 5G communications is the algorithm for effective resource allocation. In the related works, a dynamic bandwidth allocation (DBA) in OFDMA-PON has been implemented in [4] with fixed subcarriers for data scheduling, which adopts a traditional grant/report polling scheme. Moreover, dedicated resource allocation (DRA) and shared resource allocation (SRA) as two DBA methods were proposed in [5]. The DBA protocol in OFDM-PON therefore has been proposed in [6], where protocols are summarized in two schemes: the fixed burst transmission (FBT) and the dynamic circuit transmission (DCT). FBT employs a round-robin, IPACT algorithm [6] while DCT adopts bandwidth estimation. Furthermore, a power-efficient DBA scheme of OFDM-PON has also been given in [7] for the purpose of minimizing the optical network units (ONUs) transmitting power. In addition, a lot of works such as for attaining the low power consumption with OFDM-PON have been finished on a system hardware level. Specifically, a 36.86-Gb/s optical wavelength conveying six 100-MHz-bandwidth LTE-A signals has been proposed in [8]. The system supports 5-carrier aggregation, 2×2 MIMO, and three sectors, over a 40-km SSMF front-haul adopting a single 1550-nm directly modulated laser. In addition, the system [2] adopts a fixed RF channel on subcarriers; however, it becomes inflexible to satisfy DBA when high mobility of large number of user equipment (UE) occurs in the wireless front-haul. The structure [2, 3] deploys an optical distribution network (ODN), which is different with [4, 5], but the DBA problem is still the same.

However, considering a very high density level of UEs and their high mobility in future 5G cellular networks, a prominent problem waiting to be solved is the co-channel interference jointly employing radio-over-fiber, massive multiple-input and multiple-output (MIMO), and beam-forming [9] technologies. For example, in optical-wireless networks, when the same wireless frequency resources carried by different optical wavelengths overlap in the same beam direction

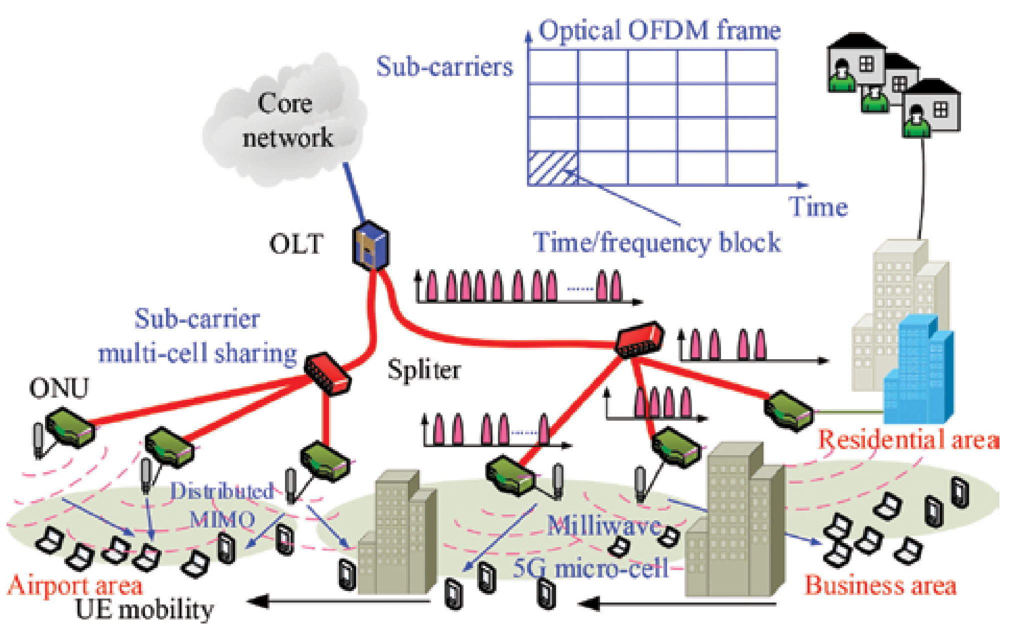

Figure 1. Network architecture of RoF-OFDM-PON based on 5G [9].

and are received by different UEs, these UEs which emerge in this beam direction under their mobility will suffer high interference. Typically, in the case of distributed massive MIMO, numbers of UEs move frequently among a few of micro-/pico-cells in any time; wireless data carried over the wavelength are shared by all served UEs in the pico-cell and adjacent cells. The minimizing of co-channel interference as mentioned will become much more imperative. Hence, it is expected to seek a scheduling optimization of wireless resources to each micro-cell mitigating the interference due to high UE mobility. Furthermore, we consider a given limited number of optical subcarriers, when an ONU needs additional resources, and in order to support the bandwidth demand for the rest of ONUs, optical subcarriers will not be reallocated in congestion cases, the problem herein is also to find ways to share optical subcarriers among local different cells by ONUs. However, it brings the additional delay problem and also configuration and control problem for selecting ONUs because of resource sharing and transmission. To achieve these targets, we propose and observe an interference mitigation and delay-aware sharing scheme for real-time services allowing that each subcarrier of RoF-OFDM-PON [2] can be multi-cell shared by UEs accessed from different micro-cells. Namely, each UE is arranged to receive multiple data streams demodulated from different ONUs simultaneously.

In this chapter, we address the aforementioned problems in the system, which have not been studied in other works before. The proposed method in this chapter can be employed by a future 5G operator to run radio-over-fiber based optical OFDM (OOFDM) [3] networks with multiple micro-/pico-cells as shown in **Figure 1**; it could be used as a method on network design to reduce resource waste and improve the performance of network.

The rest of this chapter is organized as follows. Section 2 presents our system architecture and resource allocation model. Section 3 introduces our resource sharing proposal. A heuristic algorithm guaranteeing minimum co-channel interference, maximum sharing capacity, and minimum delay time is presented in Section 4. Section 5 provides evaluation results with simulations. Finally, the chapter is concluded in Section 6.

2. RoF-OFDM-PON system

2.1. Link architecture in RoF-OFDM-PON networks

Figure 2 illustrates an experimental link architecture for signal processing in RoF-OFDM-PON [10] to support our proposal and to be employed as a physical infrastructure for one data stream in system. From the transmitter side, this implementation firstly modulates experimental data through an OFDM processing with performing of the PRBS, NRZ pulse, and QAM sequence generation (e.g., 4-QAM) in advance [10]. After that, a RF-IQ mixer is used to deal with the OFDM signal to analog RF with a proper quadrature modulation. The output signal then experiences an optical OFDM (OOFDM) modulation with a 193.1 THz CW laser by LiNbO3 mach-zehnder modulator (MZM) [10, 11] and then is sent into fiber through EDFA to amplify the signal.

On each receiver of ONU side, signals from fiber are received by photo-detector (PD) [10, 11] and are executed with a RF de-multiplexing and OFDM demodulation followed by QAM sequence generator and NRZ pulse generator in order to recovery the experimental data [10]. It is important to note that one set of optical OFDM subcarrier on fiber could be modulated to

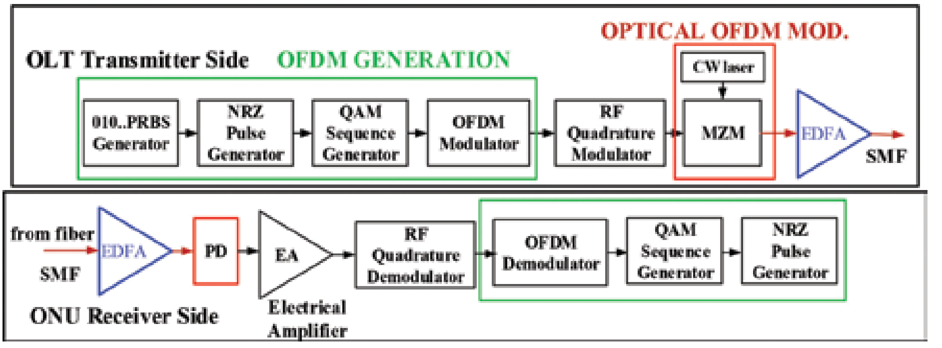

Figure 2. The downlink signal processing of RoF-OFDM-PON.

accommodate different UEs belonging to two or more receivers/ONUs in cellular networks, and the wireless data allocated to UEs belonging to any micro-cell could be transmitted by the broadcasting of multiple sharing streams from other ONUs with antennas in other micro-cells nearby (e.g., by the distributed massive MIMO [9]). In this chapter, our work thus mainly consider these resource allocation problems, while detailed physical discussions on the control and configuration issues (e.g., protocol specification) for dynamic transmission from multiple ONUs for resource sharing are out of the scope of this chapter.

2.2. Optical subcarrier allocation model

Employing the downlink signal processing mentioned in Section 2, the current OFDM-PON access networks flexibly allocate the time/frequency blocks in OFDM frame logically as shown in **Figure 3** under a mixed access rate. **Figure 3** describes an example about resource allocation of optical time/frequency block distributed to three different ONUs under different time slots and optical subcarriers. In this case, multiple wireless UE data are modulated onto each time/frequency block, and each block could be allocated to a single ONU, while each ONU could receive several such time/frequency blocks in the same time slot [12, 13]. Each optical subcarrier of time/frequency block in **Figure 3** could be addressed by a RoF modulation with radio frequency in the same or different wireless radio frequency spectra (e.g., a LTE radio frame frequency spectrum from 2110 to 2170 MHz) [14]. Moreover, according to the bandwidth capacity of single optical carrier, multiple radio frame could be conveyed on a single carrier (e.g., as reported in [8], six 100 MHz LTE-A signals are conveyed on a 36.86 Gb/s optical carrier).

Adopting this subcarrier allocation method, different UEs are fed by its ONU within its cell, and the subcarrier number allocating to each ONU could be appended according to the increase of traffic in this cell. However, the very high UE mobility in future 5G pico-cells [9] results that a few idle resource appears on a signal optical time/frequency block so that much more wasted resource is produced during resource allocation. In order to rationally allocate these idle resources (e.g., the remnant resource in **Figure 4** during slot t_2), it should be noted that the physical optical modulation process in **Figure 2** can be easily controlled to make UE data belonging to different ONUs be modulated onto the same optical time/frequency block with radio-over-fiber, as shown in **Figure 4**. For instance, in **Figure 4**, ONU 3 has idle resources in time slot t_1; however, its data requirement exceeds the allocated amount from time slot t_3 on.

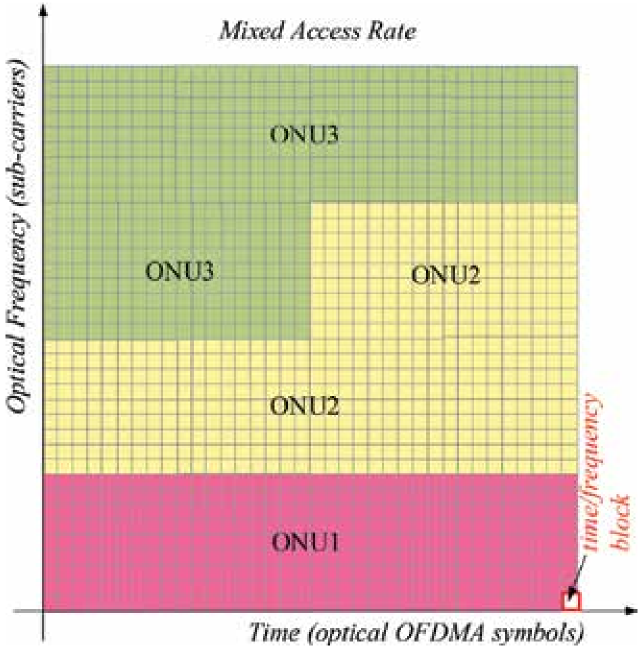

Figure 3. Optical OFDMA frame with time/frequency block allocation to different ONUs in RoF-OFDM-PON systems [12].

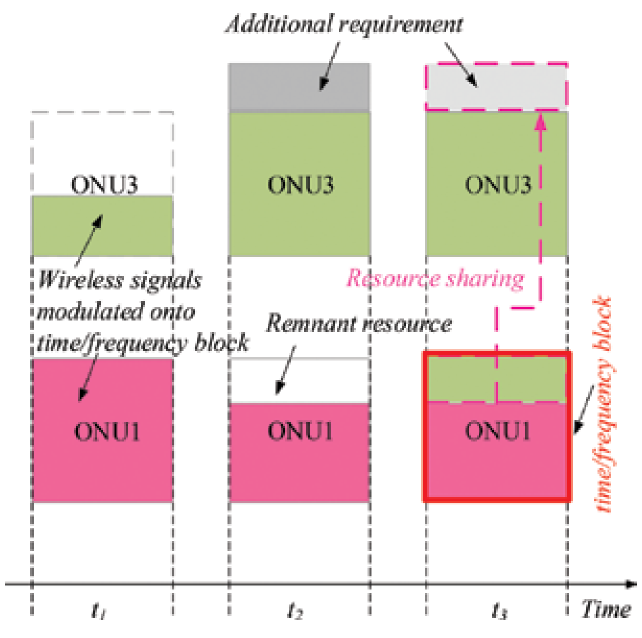

Figure 4. Multi-cell sharing of wireless resources on optical time/frequency blocks allocated to different ONUs corresponding to different wireless cells.

With our proposal, it will receive remnant resources of ONU 1 in time slot t_3 by real-time resource sharing for its additional requirement.

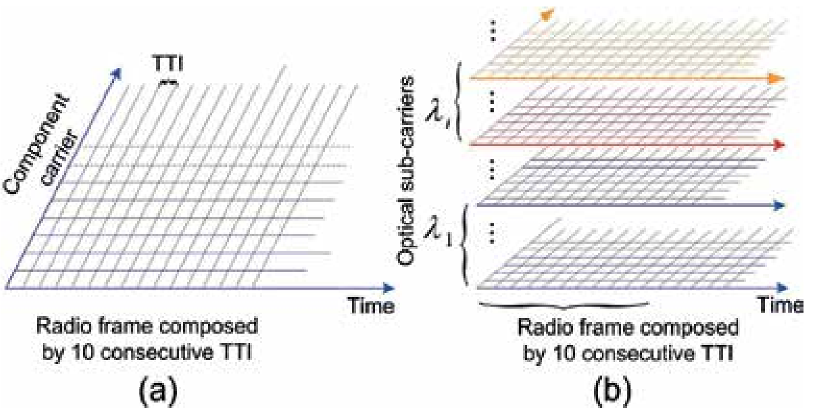

Figure 5. Wireless resource sharing logically on optical time/frequency blocks by different ONUs.

2.3. Radio frame model on optical subcarriers

The wireless spectrum resource can be illustrated by **Figure 5(a)**. In contrast to the single-layer radio frame which is illustrated in **Figure 5(a)**, by allocating more subcarriers, multiple radio frames can be delivered on fiber to each ONU, forming the multi-layer radio frames which are shown in **Figure 5(b)** for each ONU. Note that **Figure 5(b)** describes multiple wireless frames carried by a single optical subcarrier λ_i. Therefore, the time slot in **Figure 5** is different from that in **Figures 3** and **4**, for example, optical scheduling time slot t_1 in **Figure 4** contains several consecutive time slots in **Figure 5** which is named as transmission time interval (TTI) [15].

Note that the smallest optical resource unit in **Figures 3** and **4** is named as time/frequency block, while the smallest radio resource unit in **Figure 5** is named as resource block (RB). They are different concepts in this chapter. Each component carrier (CC) contains several RBs [14, 15]. One UE can receive several CCs in a certain time slot simultaneously.

3. Mathematical optimization

3.1. Assumptions of the model

Assumption 1: There are total $N^i_{\text{sub-c}}$ optical subcarriers allocated to ONU i, and l is the indicator of optical subcarrier. For each l-th optical subcarrier, it contains C_l layers of frames, as shown in **Figure 6(b)**. p is the indicator of frame on each optical subcarrier. The concept of multi-layer frames will be adopted in the following problem description and resource sharing algorithm.

Assumption 2: Denote $R_{k,t}$ as the minimum capacity requirement for user k in slot t. The UE set $\{1, 2,\ldots, \tilde{k},\ldots, \tilde{K}\}$ is served by ONU i, while UE set $\{1, 2,\ldots, k,\ldots, K\}$ is served by ONU j. Denote $N_{\text{sub-c}}$ as the consecutive subcarrier number on frequency of each RB and N_{sym} as OFDM symbol [14] number on time domain of each RB. In addition, denote $N_{\text{sc}}{}^{(\text{d})}{}_s$ as the subcarrier number for date transmission in the s-th OFDM symbol, and $N_{\text{sc}}{}^{(\text{d})}{}_s < N_{\text{sub-c}}$, because of the existence of subcarriers used for control signals in each RB.

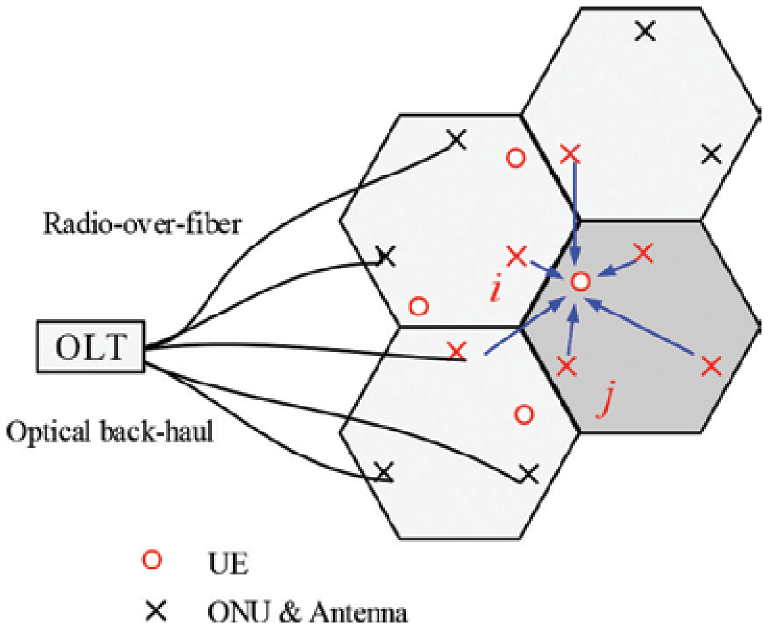

Figure 6. The illustration of multi-cell RoF-OFDM-PON scenarios with distributed massive MIMO deployment. (Each cross represents an ONU/antenna location, while the red circle represents the UE. Six red crosses in different cells highlight that one UE receives multiple data streams in the subcarrier sharing scenario).

Assumption 3: There are J numbers of modulation and coding scheme (MCS) for each RB to choose, and $R_o^{(c)}$ is code rate in a dedicated MCS, $o \in \{1, 2, \ldots, J\}$. M_o describes the constellation size of MCS [14]. In each TTI of scheduling, the capacity $r^{(o)}{}_{\mathrm{RB}}$ for one RB under MCS o is then given by Eq. (1):

$$r_{\mathrm{RB}}^{(o)} = R_o^{(c)} \log_2(M_o) \sum_{s=1}^{N_{\mathrm{sym}}} N_{\mathrm{sc}\ \ s}^{(d)} \tag{1}$$

Assumption 4: Suppose that $g_{k,l,p,n}$ indicates the wireless channel quality indicator (CQI) [14] of the n-th RB in each CC carried on the p-th optical subcarrier dedicated to UE k. The CQI of RB in each CC carried on the p-th frame on l-th optical subcarrier can be given by $g_{k,l,p} = [g_{k,l,p,1}, g_{k,l,p,2}, \ldots, g_{k,l,p,N_{\mathrm{RB}}}]^{\mathrm{T}}$. Each UE can employ at most z CCs to receive data in each slot. Each UE can only adopt one MCS for its assigned RB of CCs.

Assumption 5: We consider beam-forming [7] for wireless signal propagation. In terms of beam direction of the i-th ONU/antenna pair, two categories of UEs' conflict relationship according to any two UEs' locations (from the view point of the i-th ONU/antenna pair) are (1) same angle UEs and (2) different angle UEs. Define a matrix $\boldsymbol{\chi}_i = [\chi_i(1, 2), \ldots, \chi_i(k, k'), \ldots, \chi_i(K\text{-}1, K)]$. The value of $\chi_i(k, k')$ is then defined in the Eq. (2).

$$\chi_i(k, k') = \begin{cases} 0; \text{different angle UEs} \\ 1; \text{the same angle UEs} \end{cases} \tag{2}$$

Naturally, different beam directions can mitigate interference. For any two UEs, from a view point of the ONU/antenna pair, the first category (i.e., $c_i(k, k') = 1$) is that two UEs locate at the same angle of beam direction. The second category (i.e., $c_i(k, k') = 0$) is that two UEs locate at different beam directions. Hence, the co-channel interference of second category UEs will be mitigated, even if these UEs employ the same RB of CCs on different radio frames which are carried by different optical subcarriers. However, for the first category UEs, interference still occurs if the UEs employ the same RB of CCs on different radio frames.

Therefore, $\gamma^{(n,y,t)}{}_{k,\ k'}$ is also defined as a binary variable. As shown in Eq. (3), $\gamma^{(n,y,t)}{}_{k,\ k'} = 1$ indicates that the same RB n of the y-th CC is allocated to UE k and k' in slot t at different frames. The same RB of CCs here means the RB on component carriers in the same frequency and also the same time slot carried by different frames:

$$\gamma_{k,k'}^{(n,y,t)} = \begin{cases} 1; & \text{UE } k \text{ and } k' \text{ are allocated with the same RB of CC} \\ 0; & \text{otherwise} \end{cases} \tag{3}$$

Definition 1: The UE set {1, 2,…, $\tilde{k}$,…, $\tilde{K}$} is located outside the cell ξ and served by ONU i, while UE set {1, 2,…, k,…, K} is in the cell ξ and served by ONU j.

Definition 2: For any two UEs, from the viewpoint of the MIMO antenna, we define that the same angle UEs in Eq. (2) are two UEs located at the same angle of beam direction. The angle space depends on the coverage of a beam released by antennas (e.g., 30° or the case of narrow beam in 5G). Otherwise, they are different angle UEs which locate at different beam directions (base station MIMO antenna arrays in the cell are in the same place and treated as one point).

3.2. Modeling of resource sharing proposal

The optimization model we proposed is more applicable for the deployment of small cell coverage with a high UE mobility scenario, so that the sharing capacity can be maximized and the delay time from the OLT to each UE could be minimized by the model. In the system, we suppose a remnant resource of bandwidth of each ONU after its inter-cell allocation can be delivered to the UEs in different cells for resource sharing by the broadcasting of distributed antennas. It is assumed that the antenna transmission for wireless signals in each cell could well reach the UEs in several adjacent cells. We also suppose that each ONU is attached by one antenna element in its location by default. In this chapter, for simplicity, we directly denote i or j as an ONU/antenna pair, that is, the ONU i means the ONU in i-th ONU/antenna pair, and the ONU j means the ONU in j-th ONU/antenna pair. Especially, in terms of UE k which located in cell ξ, we define ONU i for the ONU placed outside cell ξ and ONU j for the ONU placed in cell ξ. The UE set {1, 2,…, $\tilde{k}$,…, $\tilde{K}$} is located outside the cell ξ and served by ONU i, while UE set {1, 2,…, k,…, K} is in the cell ξ and served by ONU j. The classification of different UE sets and different ONU/antenna pairs is to clearly describe the optimization problem of subcarrier multi-cell sharing.

Consider the single UE k which is accommodated by ONU j, and UE k receives data from ONU j and a shared ONU i. For the data stream from ONU i to UE k, we define $d^{\,k}{}_{i,t}$ as its delay of sharing data for UE k in time slot t by ONU i from optical back-haul in OLT to the UE.

We consider the case that multiple ONUs share their data for a single UE k. As the system model depicted in **Figure 6**, we define a set $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ which represents the set of ONUs outside the cell where UE k is located for bandwidth sharing to UE k.

On the other hand, we define $\mathcal{P} = \{j \mid j = 1, 2, \ldots, b, \ldots, n\}$ representing the set of ONUs inside the cell where UE k belongs and $\mathcal{N} = \{i \mid i = 1, 2, \ldots, b, \ldots, n\}$ representing the set of total ONUs in a local network, respectively. Here, m is less than n and $\mathcal{M} \cup \mathcal{P} \subseteq \mathcal{N}$. For the parameter b, note that the delay time $d^k_{b,t}$ is the maximum delay in sharing links among all the links through the selected ONUs to the UE k satisfying:

$$b = \arg\max_{i \in \mathcal{M}} d^k_{i,t} \tag{4}$$

We could enrich our model to the real-time scenario for multiple UEs. The joint objective to (*i*) maximize sharing capacity with minimum delay and (*ii*) to minimize co-channel interference in a time duration T can be formulated by Eq. (5) in detail.

Objective:

$$\max\left\{\sum_{t=1}^{T}\sum_{k=1}^{K}\left(\sum_{i=1,\, i\neq j}^{m} w^k_{i,t}\cdot c^k_{i,t} - \sum_{i=1,\, i\neq j}^{m} q^k_{i,t}\cdot d^k_{i,t}\right) - \sum_{i=1,\, i\neq j}^{m}\sum_{t=1}^{T}\sum_{y=1}^{N_{\mathrm{CC}}}\sum_{n=1}^{N_{\mathrm{RB}}}\left(\sum_{C_i(k,k')=1}\gamma^{(n,y,t)}_{k,k'} + \sum_{C_i\left(k,\tilde{k}\right)=1}\gamma^{(n,y,t)}_{k,\tilde{k}}\right)\right\} \tag{5}$$

where $w^k_{i,t}$ and $q^k_{i,t}$ can be further described in Eqs. (6) and (7), respectively:

$$w^k_{i,t} = \beta^k_{i,t}\cdot G^k_{i,t}\cdot \min\left\{h^k_{i,t-1}, h^k_{i,t-2}, \ldots, h^k_{i,1}\right\} \tag{6}$$

$$q^k_{i,t} = \beta^k_{i,t}\cdot D^k_{i,t}\cdot \max\left\{U^k_{i,t-1}, U^k_{i,t-2}, \ldots, U^k_{i,1}\right\} \tag{7}$$

Here, $\beta^k_{j,t}$ represents a binary indicator that UE k is served or not by ONU i in slot t. Different from $c^k_{i,t}$ which is a current capacity that could be provided to UE k in slot t, while $G^k_{i,t}$ is a current capacity which is obtained by UE k finally in slot t. Moreover, $h^k_{i,t-1}$ is a historical capacity obtained by UE k in slot t-1. It should be noted that $c^k_{i,t}$ is the shared capacity available for UE k from ONU i. Meanwhile, in Eq. (7), $D^k_{i,t}$ and $U^k_{i,t-1}$ are current delay constraint of UE k in slot t and historical delay record of UE k in slot t-1, respectively.

The optimization objective in Eq. (5) may be seemed indeed as an interference mitigation problem of finding $c^k_{i,t}$ subjected to the delay requirement from a set of $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ for UE k severed by ONU j. This will be solved in more details in our heuristic algorithms later. Firstly, we discuss all the constraints of objective as follows;

1) Capacity constraints for UE k:

$$\sum_{i \in \mathcal{M},\, i\neq j} c^k_{i,t} \geq A^k_t - \sum_{j \in \mathcal{P}} F^k_{j,t} \tag{8}$$

Equation (8) describes the total sharing capacity should not be less than the capacity requirement for each UE k in each time slot t. A^k_t is total data capacity demand of UE k and $F^k_{j,t}$ is data capacity provided by ONU j to UE k.

2) Delay constraints for UE k:

$$d_{b,t}^{k} \leq D_{i,t}^{k} \tag{9}$$

The delay constraint in Eq. (9) in each slot t means that the delay time spent on the path from the source of OLT to the destination of UE should not exceed the maximum tolerable transmission delay time (TDT) of UE k in a real-time service.

3) Capacity constraint for ONU i

Considering the 5G communication with carrier aggregation from [14, 16], we may further discuss the constraint of $c^{k}{}_{i,t}$:

$$\sum_{k=1}^{K} c_{i,t}^{k} \leq E_{i,t} \tag{10}$$

Denote $E_{i,t}$ as the total remaining capacity of ONU i after the allocation for its UEs accommodated. Equation (10) describes that the total amount of sharing capacity of UEs should be less than $E_{i,t}$.

With respect to our resource allocation model for optical time/frequency blocks with RoF and downlink signal processing in Section 2, we formulate $E_{i,t}$ approximately with the aforementioned assumptions which are detailed in the aspect on resource allocation.

The remaining capacity $E_{i,t}$ of ONU i in time slot t is then given as Eq. (11) approximately:

$$E_{i,t} = \sum_{l=1}^{N_{\text{sub-c}}^{i}} C_l \left(\frac{1}{Q} N_{\text{cc}} N_{\text{RB}} \sum_{o=1}^{Q} r_{\text{RB}}^{(o)} \right) - \sum_{l=1}^{N_{\text{sub-c}}^{i}} \sum_{p=1}^{C_l} \sum_{\tilde{k}=1}^{\tilde{K}} \sum_{y=1}^{N_{\text{cc}}} \sum_{n=1}^{N_{\text{RB}}} \sigma_{\tilde{k},l,p}^{(n,y,t)} \sum_{o=1}^{Q} \mu_{\tilde{k},o} \cdot r_{\text{RB}}^{(o)} \tag{11}$$

Especially, $\sigma^{(n,y,t)}{}_{k,l,p}$ is a binary variable to define whether or not the n-th RB of the y-th CC is assigned to the k-th UE on the p-th frame on the l-th optical sub-carrier in slot t, and $\sigma^{(n,y,t)}{}_{k,l,p} = 1$ expresses that allocating the n-th RB of the y-th CC to the k-th UE in slot t. Here, we define a binary variable $\mu_{k,o} = 1$ to express that the k-th UE employs the o-th MCS. Q in Eq. (13) describes the highest MCS employed by the k-th UE corresponding to a CQI of RB "$max(g_{k,l,p,\delta^*})$" in each CC of the p-th frame on the l-th optical subcarrier. Here:

$$\delta^* = \underset{n \in \{1,2,\ldots,N_{\text{RB}}\}}{\arg\max} \left(g_{k,l,p,n} \right) \tag{12}$$

$$Q_{k,l,p,\max\left(g_{k,p,\delta^*}\right)} = \underset{j \in \{1,2,\ldots,J\}}{\operatorname{argmax}} \left(R_j^{(c)} \log_2 \left(M_j \right) \middle| g_{k,l,p,\delta^*} \right) \tag{13}$$

In the Eq. (11), $N^{i}{}_{\text{sub-c}}$, N_{cc}, and N_{RB} are the number of optical subcarriers allocated to ONU i by OFDM-PON, the total number of wireless component carriers (CC) [15, 16] modulated onto a single optical subcarrier, and total number of wireless resource blocks (RB) carried by a single wireless component carrier (CC), respectively. The second term of polynomial in Eq. (11)

should be larger than the summation of total UE minimum capacity requirement. Hence, a constraint of $c^k_{i,t}$ can be further formulated as in Eq. (14),

$$\sum_{k=1}^{K} c^k_{i,t} \leq E_{i,t} \leq \sum_{l=1}^{N^i_{\text{sub-c}}} C_l \left(\frac{1}{Q} N_{\text{cc}} N_{\text{RB}} \sum_{o=1}^{Q} r^{(o)}_{\text{RB}} \right) - \sum_{\tilde{k}=1}^{\tilde{K}} R_{\tilde{k}} \tag{14}$$

In the following allocation algorithm, we will satisfy all the aforementioned constraints to find the sharing capacity $c^k_{i,t}$ for UE k from ONU i.

4) Interference constraint for UEs

In terms of the i-th ONU/antenna pair, the constraint of $\gamma^{(n,y,t)}{}_{k,\tilde{k}}$ is described in Eq. (15). It means that the number of same RB of CC allocated to different UEs on all the frames carried by optical subcarriers has an upper limitation. Since $\gamma^{(n,y,t)}{}_{k,\tilde{k}}$ is a binary indicator, Eq. (15) could be treated as two cases. First, when $\gamma^{(n,y,t)}{}_{k,\tilde{k}} = 1$, the same RB n of the y-th CC is allocated to UE k and $\tilde{k}$ in slot t at different frames, the product of all the numbers of these RBs allocated to UE k and all the number of same RBs allocated to UE $\tilde{k}$ must not be larger than their arithmetic mean square, while the upper limitation of their arithmetic mean equals half of the number of total frames. Second, when $\gamma^{(n,y,t)}{}_{k,\tilde{k}} = 0$, any RB n of the y-th CC is not allocated to both UE k and $\tilde{k}$ in slot t at different frames; therefore, for all the number of RBs allocated to UE k and $\tilde{k}$, their product must equal to 0 (i.e., without RB overlapping on the same time/frequency domain) as described in Eq. (15):

$$\sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} \cdot \sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{\tilde{k},l,p} \leq \gamma^{(n,y,t)}_{k,\tilde{k}} \left[\frac{1}{2} \left(\sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} + \sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{\tilde{k},l,p} \right) \right]^2 \leq \gamma^{(n,y,t)}_{k,\tilde{k}} \left[\frac{1}{2} \left(\sum_{l=1}^{N^i_{\text{Sub-c}}} C_l \right) \right]^2, \quad \forall k, \tilde{k}, \forall y, \forall n, \forall t \tag{15}$$

Similarly, we hereby obtain the following constraint of $\gamma^{(n,y,t)}{}_{k,k'}$ as described in Eq. (16):

$$\sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} \cdot \sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k',l,p} \leq \gamma^{(n,y,t)}_{k,k'} \left[\frac{1}{2} \left(\sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k,l,p} + \sum_{l=1}^{N^i_{\text{Sub-c}}} \sum_{p=1}^{C_l} \sigma^{(n,y,t)}_{k',l,p} \right) \right]^2 \leq \gamma^{(n,y,t)}_{k,k'} \left[\frac{1}{2} \left(\sum_{l=1}^{N^i_{\text{Sub-c}}} C_l \right) \right]^2, \quad \forall k, k', \forall y, \forall n, \forall t \tag{16}$$

4. Proposed resource sharing algorithm

In this section, we propose a heuristic algorithm for obtaining sub-optimal solutions because solving the objective in Section 3 is highly complex. A natural and simple approach to address the joint objectives of Eq. (5) is to treat it as a maximum flow and minimum cost problem about resource allocation (e.g., RB of CC) by assigning $\sigma^{(n,y,t)}{}_{k,l,p}$ we defined for UE in each slot, approximately. We record and observe some historical information (e.g., $h^k_{j,t\text{-}1}$ and $U^k_{j,t\text{-}1}$) as timely references and evaluations for finding a maximum flow ($c^k_{i,t}$) subjected to the delay requirement ($d^k_{i,t}$) with minimum delay time from a set of $\mathcal{M} = \{i \mid i = 1, 2, \ldots, b, \ldots, m\}$ for UE

k severed by ONU j. In the following sections, the sharing path assignment and the corresponding resource allocation will be detailed in algorithm descriptions by transferring the problem into the optimization with the aid of graph theory.

4.1. Problem statement

- *Given parameters*:
  - $G(\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is UE k and set of all ONUs and $\mathcal{E}$ is the set of resource sharing path through multiple ONUs to UE k
  - Matrix $\mathcal{C}_{k,t} = [c^k_{1,t}, \ldots, c^k_{i,t}, \ldots, c^k_{N,t}]$, $\forall\, k$ in $\mathcal{K}$, $c^k_{i,t} > 0$
  - Matrix $\mathcal{D}_{k,t} = [d^{\,k}_{1,t}, \ldots, d^{\,k}_{i,t}, \ldots, d^{\,k}_{N,t}]$, $\forall\, i$ in $\mathcal{N}$, $\forall k$ in $\mathcal{K}$
  - Matrix $\mathcal{X}_i = [\chi_i(1,2), \ldots, \chi_i(k,k'), \ldots, \chi_i(K\text{-}1,K)]$
  - Matrix $\mathcal{Y}_i = [Y_i(k,1), \ldots, Y_i(k,\tilde{k}), \ldots, Y_i(k,\tilde{K})]$
  - Set of UE in the cell: $\mathcal{K} = \{k \mid k = 1, 2, \ldots, K\}$
  - Set of UE outside the cell: $\tilde{\mathcal{K}} = \{\tilde{k} \mid \tilde{k} = 1, 2, \ldots, \tilde{K}\}$
  - Set of ONU: $\mathcal{N} = \{i \mid i = 1, 2, \ldots, b, \ldots, N\}$
  - $R_{k,t}$: Minimum data capacity requirement for UE k
  - $B_{k,t}$: Allocated data capacity to UE k

Note that $\mathcal{X}_i$ and $\mathcal{Y}_i$ are two matrices which store UE conflict relationships.

- *Objective*:
  - Minimize the co-channel interference which is generated by sharing data received for the same angle UEs in their located cell and also in their adjacent cells.
  - Maximize the sharing capacity in terms of UEs.
  - Minimize the average delay of sharing data transmission by ONUs to satisfy UE requirements.

4.2. Algorithm description

The sharing algorithm tries to search the idle RBs over each optical subcarrier delivering to each cell and shares them to the UEs in the adjacent cells. From the sharing paths with minimum delay time, the algorithm selects the paths with maximum number of idle RBs for each UE so that it could maximize the sharing capacity for each UE. Consequently, the algorithm as a solution of optimization problem for our resource allocation model is suggested to be executed on the OLT side of optical back-haul. Considering the single UE k ($k = 1, 2, \ldots, K$) which is accommodated by multiple sharing paths from different ONUs, UE k can receive the data from each sharing ONU i ($i = 1, 2,\ldots, N$). The algorithm is divided into several steps as shown in **Tables 1**, **2**, and **3**.

In the first step, a dynamic sharing graph is generated, for instance, **Figure 7(a)** illustrates an example of a sharing tree that five ONUs share bandwidth resources to UE k. The rooted vertex represents UE k, and any leaf vertex i represents ONU i, respectively. Each edge represents a sharing path. In the second step, the weights of vertex and edge are assigned. For the data of UE k from ONU i, $d^k_{i,t}$ denotes the overall delay on sharing path through ONU i to UE k in time slot t. $c^k_{i,t}$ denotes the available sharing data capacity for UE k from ONU i in time slot t.

ALGORITHM 1 Real-time Sharing Algorithm (RTSA)
<Note>:**Algorithm 1** contains **FUNCTION 1, 2** and **3**

Input: Matrix $C_{k,t} = \{c^k_{i,t} \mid c^k_{1,t},\ldots, c^k_{i,t},\ldots, c^k_{N,t}, \forall\, c^k_{i,t} > 0\}$
Matrix $\mathcal{D}_{k,t} = \{d^k_{1,t},\ldots, d^k_{i,t},\ldots, d^k_{N,t}\}$
Set of UE: $\mathcal{K} = \{k \mid k = 1, 2,\ldots, K\}$
Set of ONU: $\mathcal{N} = \{i \mid i = 1, 2,\ldots, b,\ldots, N\}$
Initialization: $\boldsymbol{G_k} = \emptyset$ for all $k = 1, 2,\ldots, K$; $\boldsymbol{G_k}$: total RB set to UE k
While ($B_{k,t} < R_{k,t}$) do
STEP 1: Make a sharing graph $\boldsymbol{G}$ ($\mathcal{V}$, $\mathcal{E}$); $\mathcal{V} = \{1,\ldots, N\}\mathrm{U}\{k\}$, $\mathcal{E} = \{(i, k) \mid d^k_{i,t} \le D_k\}$; D_k is a maximum tolerable delay(MTD).
STEP 2: Assign weights to $\mathcal{E}$ by matrix $\mathcal{D}_{k,t}$ and sort the edges in $\mathcal{E}$ according to its weight in graph $\boldsymbol{G}$ ($\mathcal{V}$, $\mathcal{E}$) in an ascending order.
STEP 3: Find an edge (i, k) with minimum $d^k_{i,t}$ in the ascending order of $\mathcal{E}$ as a current link (i, k) for resource sharing.
STEP 4: $B_{k,t} \leftarrow c^k_{i,t}$ by allocating $\sigma^{(n,y,t)}_{k,l,p}$ until $B_{k,t} \ge R_{k,t}$, otherwise, go to **STEP 3**
STEP 5: Traverse all UEs in set $\mathcal{K}$ to allocate resource satisfying, $i = \mathrm{argmin}\, d^k_{i,t}$, $B_{k,t} \leftarrow c^k_{i,t}$, find $c^k_{i,t}$ employing
FUNCTION 1: form $\boldsymbol{G}_k$
End while
STEP 6: Traverse all ONUs in set $\mathcal{N}$ to allocate resources, repeat **STEPS 1–5**.
STEP 7: Record historical information (e.g., $h^k_{j,t}$ and $U^k_{j,t}$) by **FUNCTION 2**
STEP 8: Update matrix $C_{k,t}$, and matrix $\mathcal{D}_{k,t}$.
End
Out put: $\boldsymbol{G_k} = \{\boldsymbol{C}_k(1), \boldsymbol{C}_k(2), \boldsymbol{C}_k(i),\ldots\ldots, \boldsymbol{C}_k(N)\}$
End
FUNCTION 1: form $\boldsymbol{G}_k$, for any k; $\boldsymbol{C}_k(i)$: a RB set from ONU i to UE k
Initialize $\boldsymbol{G}_k = \{\boldsymbol{C}_k(1), \boldsymbol{C}_k(2), \boldsymbol{C}_k(i),\ldots, \boldsymbol{C}_k(N)\} = \emptyset$, set $B_{k,t} \leftarrow 0$.
1: Find ONU i for current link(i, k), where $i = \mathrm{argmin}\, d^{\,k}_{i,t}$, set $p \leftarrow 1$.
2: Set $l \leftarrow 1$; l is layer (radio frame) indicator.
3: Find the idle RB for any UE by **FUNCTION 3**, where $\sigma^{(n,y,t)}_{k,l,p} = 0, \forall k$.
4: Allocate a corresponding idle RB to UE k, put RB of $\sigma^{(n,y,t)}_{k,l,p} = 0$ into set $\boldsymbol{C}_k(i)$, then for the UE k,$\sigma^{(n,y,t)}_{k,l,p} \leftarrow 1$. Increase capacity $B_{k,t}$, where, $B_{k,t} = B_{k,t} + r^{(n,y,t)}_{\mathrm{RB}k,l,p}$.
5: If $B_{k,t} \ge R_{k,t}$, **break**; $r^{(n,y,t)}_{\mathrm{RB}\ k,l,p}$ is the capacity of RB.
Else if $\forall k, \forall \sigma^{(n,y,t)}_{k,l,p} = 1$, i.e., the set of RBs on current layer have been occupied and cannot be scheduled to UE k for sharing, and **if** $l \le L$ add l, go to **STEP 3**.
Else if $p \le P_{\text{max-sub}}$, add p, go to **STEP 2**.
Else go to **STEP 1**.
End IF
6: Output $\boldsymbol{G}_k = \{\boldsymbol{C}_k(1), \boldsymbol{C}_k(2), \boldsymbol{C}_k(i),\ldots\ldots, \boldsymbol{C}_k(N)\}$

Table 1. Real-time sharing algorithm.

FUNCTION 2 Historical State Recording

1: Set $h^k_{i,t} = \sum_{p=1}^{N^i_{\text{sub-c}}} \sum_{y=1}^{N_{\text{cc}}} \sum_{n=1}^{N_{\text{RB}}} \sigma^{(n,y,t)}_{k,p} \sum_{o=1}^{Q} \mu_{k,o} \cdot r^{(o)}_{\text{RB}}$
2: Set $U^k_{i,t} = d^{\,k}_{i,t}$
3: For t from 1 to T, $\forall k,i$
4: If $h^k_{i,t} < h^{(\min)}_{k,t}$, then $h^{(\min)}_{k,t} = h^k_{i,t}$
5: If $U^k_{i,t} > U^{(\max)}_{k,t}$, then $U^{(\max)}_{k,t} = U^k_{i,t}$
End

Table 2. Historical state recording function.

FUNCTION 3 Interference Mitigation
<Note>:This function addresses RB allocation according to UE conflict relationships

1: If $\chi_i(k, k') = 0, \forall k, k'$ and $Y_i(k, \tilde{k}) = 0, \forall k, \tilde{k}$
Find any RB where $\sigma^{(n,y,t)}_{k,l,p} = 0$, $\sigma^{(n,y,t)}_{k',l,p} = 0$, and $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 0$ to allocate, for UE k
2: Else if $Y_i(k, \tilde{k}) = 1$, find RB where $\sigma^{(n,y,t)}_{k',l,p} = 0$ and $\sigma^{(n,y,t)}_{k,l,p} = 0$ to allocate,
but refrain RBs where $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 1, \forall l,p$
3: **Else if** $\chi_i(k, k') = 1$, find RB where $\sigma^{(n,y,t)}_{\tilde{k},l,p} = 0$ and $\sigma^{(n,y,t)}_{k,l,p} = 0$ to allocate,
but refrain RBs where $\sigma^{(n,y,t)}_{k',l,p} = 1, \forall l,p$
End

Table 3. Interference mitigation function.

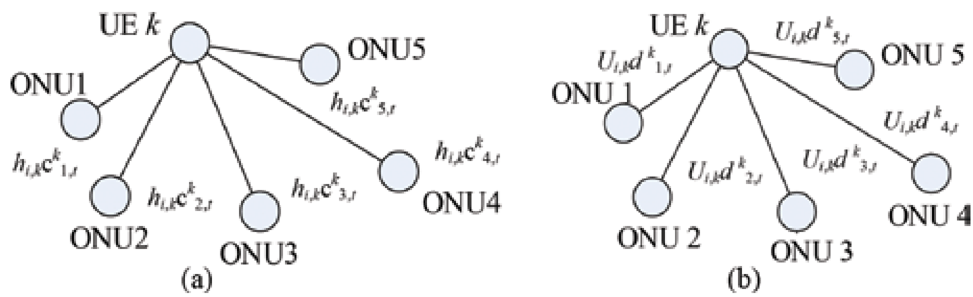

Figure 7. Wireless resource sharing logically on optical time/frequency blocks by different ONUs.

For each time slot t, the weight of each leaf is $h_{i,k}c^k_{i,t}$. The weight of edge between the rooted vertex and any leaf vertex i is $U_{i,k}d^{\,k}_{i,t}$, as shown in **Figure 7(b)**. Specifically, for $t = 1, 2,\ldots, T$, we define $h_{i,k} = min\{h^i_{k,t}, h^k_{j,t\text{-}1},\ldots, h^k_{j,1}\}$ and $U_{i,k} = min\{U^i_{k,t}, U^k_{j,t\text{-}1},\ldots, U^k_{j,1}\}$, respectively.

In the third step, the sharing path selection is performed by finding minimum delay time on each ONU. A sharing ONU combination could be found for UE k aiming to a minimum delay under its data rate demand.

In the fourth step, we allocate RB of CC to UEs in each ONU. Resource sharing for each UE (e.g., RB of CC) is executed by assigning $\sigma^{(n,y,t)}_{k,l,p}$. Here, $\sigma^{(n,y,t)}_{k,l,p}$ is a binary variable to define whether or not the n-th RB of the y-th CC on the l-th radio frame on p-th optical subcarrier is assigned to the k-th UE in slot t for finding a proper $c^k_{i,t}$ subjected to the minimum delay $d^{\,k}_{i,t}$ from a set of ONUs.

In the fifth step, the algorithm loops to another UE allocating RBs to it until all the UEs of current set have been finished. In the sixth step, the algorithm traverses all served ONUs to finish the RB allocation. In the seventh step, we compute the capacity obtained by UE k from ONU i in slot t and its delay time. Then we compare them to the previous historical values and update the historical peak value in the case that the current one exceeds it. After executing these steps, the algorithm outputs the RB set allocated to UE k classifying them into each subset of RBs obtained by each sharing ONU individually.

To meet all constrains of mathematical descriptions in Section 3, we formulate three different functions for the heuristic algorithm herein as practical approaches to achieve the target of optimization. Function 1 is one of solutions to search idle RBs for resource sharing satisfying minimum optical wavelength cost. Function 2 addresses historical information recording and their updating. Function 3 finds idle RBs and allocates them according to different UE conflict relationships in order to mitigate co-channel interference, which will be intensively evaluated and discussed in the next section of this chapter.

5. Simulations and numerical results

5.1. Simulation parameters

In this section, we provide a deep observation for the proposed resource sharing approach on the performance of wireless UEs in the OFDM-PON system. The simulation and analytic evaluation by large-scale C++ programming mainly focus on the interference mitigation of mobile UEs in the cell under different mobility and times.

In intensive large-scale C++ simulations, a RoF-OFDM-PON covering up to 256 cells assuming random UE mobility is deployed to evaluate our proposal as shown in **Figures 8** and **9**. Optical subcarriers with per λ_i 10-Gb/s digital-equivalent data rate are adopted. LTE-like wireless resources carried on optical subcarriers are assigned to UEs corresponding to the scheduling solution in the well-known network simulator 3 (ns-3) [17], supporting maximum five-carrier aggregation, simultaneously. MCS is assigned to UEs corresponding to Eqs. (12) and (13) by the scheduling in the ns-3. The main simulation parameters are described in **Table 4**.

In the C++ simulations, according to LTE-EPC model [18] in ns-3 simulator and with respect to its resource allocation models, we modify the scheduler significantly based on our proposed real-time sharing algorithm (RTSA). We evaluate the throughput performance of UEs by comparing RTSA with maximum throughput (MT) and proportional fair (PF) schemes [6, 19]. Note that for a fair comparison, we also modify the scheduler to serve multiple wavelength scheduling (i.e., multiple radio frames carried on one optical wavelength) since MT and PF themselves have no such functionality.

From the entire network perspective, the total UE number is set to 36,000 in simulations under different mobility ratios (a = number of mobile UEs/number of total UEs). We set a position for each UE with position allocator by model library of NS3 [17] (e.g., random waypoint). Meanwhile, we aim to simulate a difference specifically on UE mobility, for example, changing the residential position of UEs (e.g., migrate and recall UEs) within the scope of all cells regularly

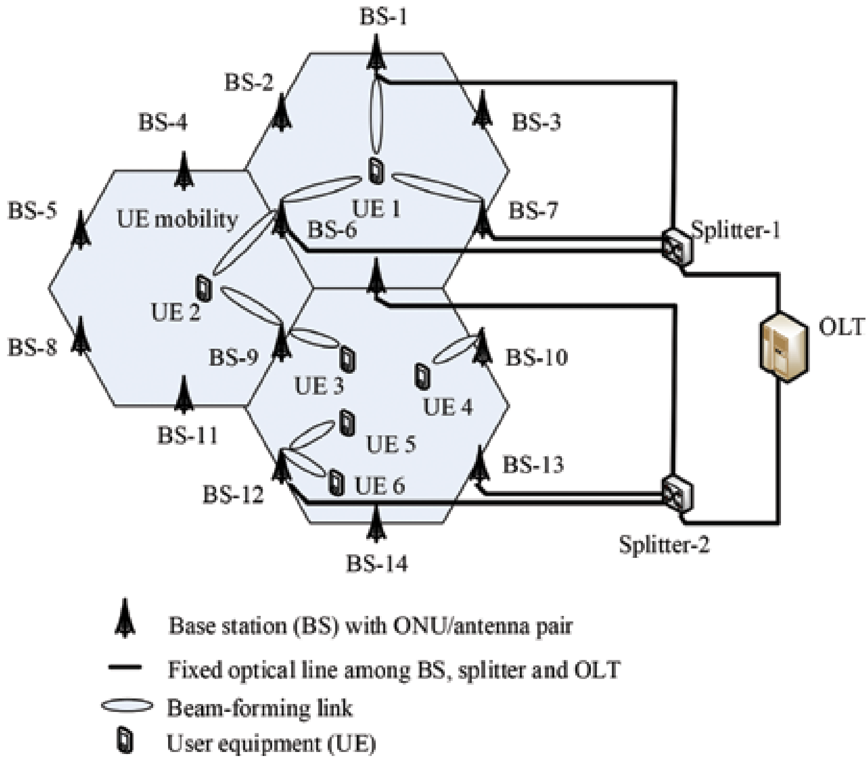

Figure 8. Scenarios of UEs and distributed antenna allocation in simulations with beamforming in the cluster of wireless cells.

in different times as shown in **Figure 9(a)** and **(b)** respectively. For instance, we define different mobility ratios of UEs equaling to 0.2 and 0.8, respectively.

5.2. Results and analysis on interference mitigation

Next, we observe the effect on interference mitigation under different UE mobility rates within different time slots by generating massive number of same angle UEs (the conflict UEs) in each beam direction. As an example, in **Figure 10** we typically illustrate four windows of UE distribution states under random mobility in four different time slots, respectively. With the irregular movement of UEs, new UE conflict relationships will be generated randomly in terms of different antennas. For instance, in the time slot 1, UE k_4 and k_5 are located in different directions in terms of antenna 1. Simultaneously, UE k_1 and k_7 which have a conflict relationship with each other are located in the same direction for antenna 1. However, in the time slot 2, a new UE conflict relationship is generated between UE k_4 and k_5, while UE k_1 and k_7 are located in different directions, and their conflict relationship disappears. Similarly, in a continuous time scope (e.g., 10 minutes) containing many more time slots, we then observe our proposed scheme in the aspect of interference cancellation. We therefore evaluate the block error rate (BLER) of RTSA, PF, and MT under Gauss interference model in ns-3 model library so as to compare our proposed scheme with conventional schemes in terms of their effectiveness on interference mitigation. In simulations, BLER is observed under fair channel condition, for instance, the same level of signal power which is set by signal-to-noise ratio (SNR), and the

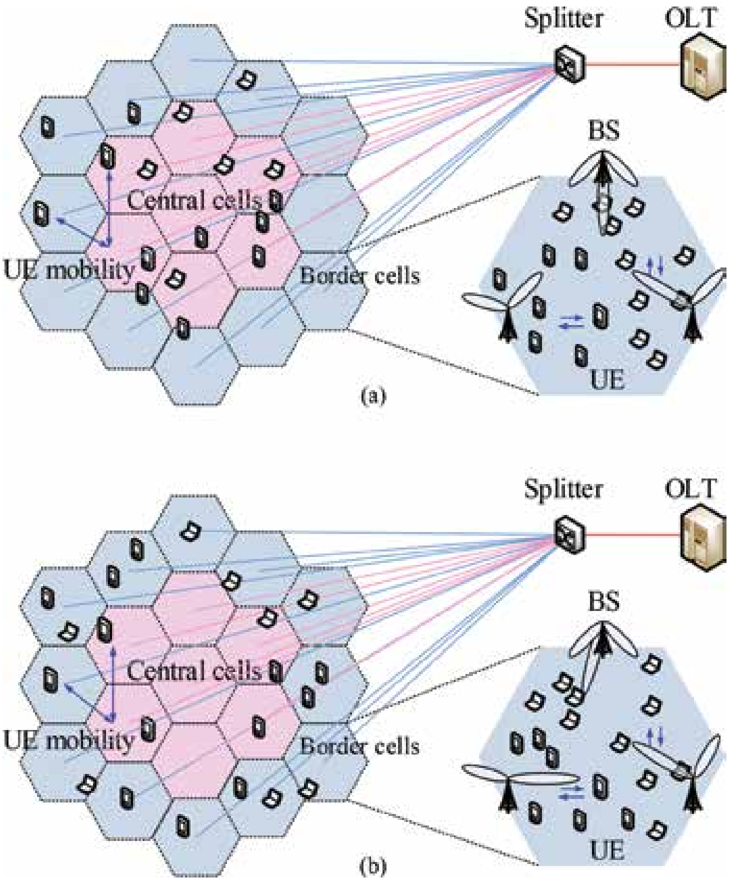

Figure 9. Scenarios of UE long-distance migration with position distribution for each UE by random model library (e.g., random waypoint) in simulations. (a) The aggregation of UEs at central cells. (b) The spreading of UEs to border cells.

Parameter	Value	Parameter	Value
LTE subcarrier	15 kHz	Frame duration	10 ms
Resource block	180 kHz	TTI	1 ms
RB carriers ($N_{sub\text{-}c}$)	12	UE date rate$_{min}$ (R_k)	200 Mbps
RB OFDM symbols	7	MCS (J)	29
UE received CC$_{max}$ (z)	5	Bandwidth of CC	20 MHz
Single CC length (m)	100 RBs	Testing MIMO per cell	4 × 4
BS TX power	30 dBm	Number of cell	256
Noise spectral density	−174 dBm/Hz	Cell radius	500 m
Path loss (distance R), in dB	128.1 + 37.6lg$R$	SMF fiber distance	20 km

Table 4. Simulation parameters.

SNR is obtained according to the parameters in **Table 4**. In addition, we set random UE location and irregular mobility with the increase of time.

The change of BLER is observed under low mobility (a = 0.2) and high mobility (a = 0.8) cases, respectively. As shown in **Figure 11**, RTSA has the lowest level of BLER than MT and PF, which

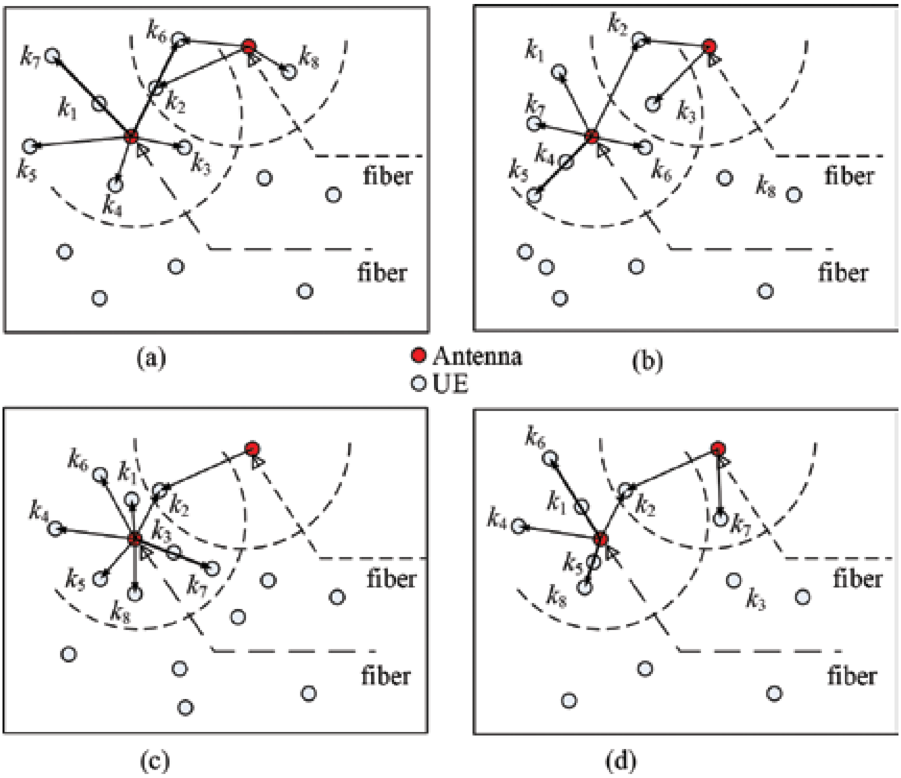

Figure 10. A group of observations about interference with UE/antenna distribution under different time slots and UE migration when employing the proposed scheme. In (a)–(d), the dashed line represents the propagation scope of each antenna. The solid arrow line represents a link with one beam direction (the thick arrow line represents a link which has conflict UEs with the potential interference).

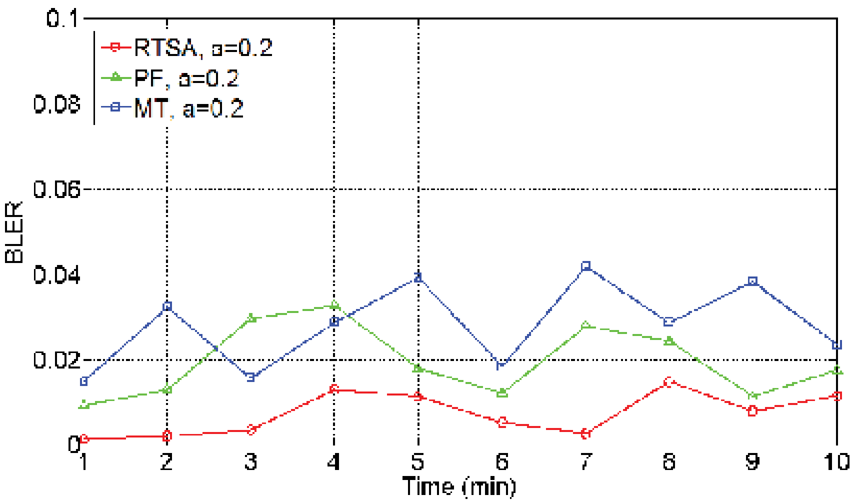

Figure 11. Comparisons of BLER under different time durations (mobility ratio 0.2).

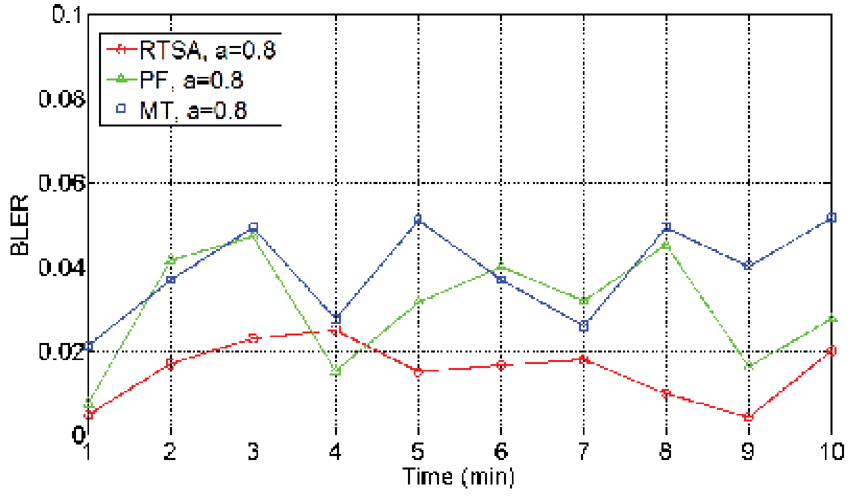

Figure 12. Comparisons of BLER under different time durations (mobility ratio 0.8).

also can be found in **Figure 12**. This feature reflects that the proposed RTSA has a benefit on BLER improvement. Compared with the results in **Figures 11** and **12**, we could further observe that the level of BLER for a = 0.8 is larger than that in a = 0.2 and the fluctuation of BLER in different time is also obvious in terms of the case where a = 0.2. The reason for a higher level of BLER is that a higher mobility brings much more interference with the conflict UEs generated in each beam direction. Nevertheless, our proposed RTSA will still has the lowest BLER level even in the case of higher UE mobility.

6. Conclusions

This chapter investigates the resource sharing problems for future 5G cellular networks [20], which jointly employ distributed massive MIMO, beamforming, and OFDMA-based passive optical network supporting radio-over-fiber (RoF). We have surveyed the system and its physical transmission features to explore reasonable solutions. With the assumptions based on physical features of system given in this chapter, we describe the latest hot problem with mathematical optimization for minimizing co-channel interference, etc. Since it is highly complex to get optimal results, then we heuristically formulate a real-time sharing algorithm as a practical solution. Simulation results also reveal that the proposed scheme is the most efficient one at the interference mitigation compared to conventional schemes.

Author details

Sheng Xu

Address all correspondence to: xusheng@akane.waseda.jp

Global Information and Telecommunication Institute, Waseda University, Tokyo, Japan

References

[1] Bi S, Zhang R, Ding Z, Cui S. Wireless communications in the era of big data. *IEEE Communications Magazine*. 2015;**53**(10):190-199

[2] Lin YM, Tien PL. Next-generation OFDMA-based passive optical network architecture supporting radio-over-fiber. *IEEE Journal on Selected Areas in Communications*. 2010;**28**(6): 791-799

[3] Tien PL, Lin YM, Yuang MC. A novel OFDMA-PON architecture toward seamless broadband and wireless integration. In: *Conference on Optical Fiber Communication*; San Diego, CA, USA; 22-26 March 2009; IEEE; pp. 1, 3, 22-26

[4] Zhang J, Wang T, Ansari N. An efficient MAC protocol for asynchronous ONUs in OFDMA PONs. In: *Optical Fiber Communication Conference and Exposition (OFC/NFOEC 2011) and The National Fiber Optic Engineers Conference*; Los Angeles, CA, USA; 6-10 March 2011; IEEE; pp. 1-3

[5] Wei W, Hu J, Qian D, Ji PN, Wang T, Liu X, Qiao C. PONIARD: A programmable optical networking infrastructure for advanced research & development of future internet. *Journal of Lightwave Technology*. 18 February 2009;**27**(3):233-242

[6] Wei W, Wang T, Qian D, Hu J. MAC protocols for optical orthogonal frequency division multiple access (OFDMA)-based passive optical networks. In: *Optical Fiber Communication/National Fiber Optic Engineers Conference (OFC/NFOEC 2008)*; San Diego, CA, USA; 24-28 February 2008; IEEE; pp. 1-3

[7] You W, Yi L, Huang S, Chen J, Hu W. Power efficient dynamic bandwidth allocation algorithm in OFDMA-PONs. *IEEE/OSA Journal of Optical Communications and Networking*. 2013;**5**(12):1353-1360

[8] Liu X, Effenberger F, Chand N, Zhou L, Lin H. Efficient mobile fronthaul transmission of multiple LTE-A signals with 36.86-Gb/s CPRI-equivalent data rate using a directly-modulated laser and fiber dispersion mitigation. In: *Asia Communications and Photonics Conference*; Shanghai, China; 11-14 November 2014; Optical Society of America (OSA); paper AF4B.5

[9] Andrews JG, Buzzi S, Choi W, Hanly SV, Lozano A, Soong ACK, Zhang JC. What will 5G be? *IEEE Journal on Selected Areas in Communications*. 2014;**32**(6):1065-1085

[10] Almasoudi F, Alatawi K, Matin MA. Study of OFDM technique on RoF passive optical network. *Optics and Photonics Journal*. 2013;**3**:217-224

[11] Chow CW, Yeh CH, Wang CH, Wu CL, Chi S, Lin C. Studies of OFDM signal for broadband optical access networks. *IEEE Journal on Selected Areas in Communications*. 2010;**28**(6):800-807

[12] Kanonakis K, Tomkos I, Pfeiffer T, Prat J, Kourtessis P. ACCORDANCE: A novel OFDMA-PON paradigm for ultrahigh capacity converged wireline-wireless access networks. In: *12th International Conference on Transparent Optical Networks (ICTON2010)*; Munich, Germany; 27 June-1 July 2010; IEEE; pp. 1-4

[13] Qian D, Cvijetic N, Hu J, Wang T. 108 Gb/s OFDMAPON with polarization multiplexing and direct detection. *Journal of Lightwave Technology*. 2010;**28**(4):484-493

[14] Liao HS, Chen PY, Chen WT. An efficient downlink radio resource allocation with carrier aggregation in LTE-Advanced networks. *IEEE Transactions on Mobile Computing*. 2014;**13**(10): 2229-2239

[15] Cox C. An Introduction to LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications. West Sussex: Wiley; 2014

[16] Akyildiz IF, Gutierrez-Estevez DM, Reyes EC. The evolution to 4G cellular systems: LTE-advanced. *Physical Communication*. 2010;**3**:217-224

[17] NS-3 consortium. "NS-3 network simulator," http://www.nsnam.org/, INRIA and the University of Washington [Accessed September 2017]

[18] CTTC. "LTE-EPC network simulator (LENA)," CTTC, http://iptechwiki.cttc.es [Accessed October 2017]

[19] Guan N, Zhou Y, Tian L, Sun G, Shi J. Qos guaranteed resource block allocation algorithm for LTE systems. In: *IEEE 7th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob 2011)*; Wuhan, China; 10-12 October 2011; IEEE; pp. 307-312

[20] Xu S, Xu SG, Tanaka Y. Sub-carrier sharing in OFDM-PON for 5G mobile networks supporting radio-over-fibre. In: *21st OptoElectronics and Communications Conference (OECC 2016) held jointly with International Conference on Photonics in Switching (PS)*; Niigata, Japan; 3-7 July 2016; IEEE; pp. 1-3

Chapter 7

Transport Protocol Performance and Impact on QoS while on the Move in Current and Future Low Latency Deployments

Eneko Atxutegi, Jose Oscar Fajardo and Fidel Liberal

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71779

Abstract

Transport protocols and mobile networks have evolved independently leading to a lack of adaptability and quality of service (QoS) degradation while running under the variability circumstances present in cellular access. This chapter evaluates the performance of state-of-the-art transmission control protocol (TCP) implementations in challenging mobility scenarios under 4G latencies and low delays that model the proximity service provisioning of forthcoming 5G networks. The evaluation is focused on selecting the most appropriate TCP flavor for each scenario taking into account two metrics: (1) the goodput-based performance and (2) a balanced performance metric that includes parameters based on goodput, delay and retransmitted packets. The results show that mobility scenarios under 4G latencies require more aggressive TCP solutions in order to overcome the high variability in comparison with low latency conditions. Bottleneck Bandwidth and Round-Trip Time-RTT (BBR) provides better scalability than others and Illinois is more capable of sustaining the goodput with big variability between consecutive samples. Besides, CUBIC performs better in lower available capacity scenarios and regarding the balanced metric. In reduced end-to-end latencies, the most suitable congestion control algorithms (CCAs) to maximize the goodput are NewReno (low available capacity) and CUBIC (high available capacity) when moving with continuous capacity increases. Additionally, BBR shows a balanced and controlled behavior in most of the scenarios.

Keywords: transport protocols, performance, mobility, 4G, low latency, 5G

1. Introduction

Mobile broadband (MBB) usage has risen significantly in the last years and so has done the customers' awareness regarding the quality of service (QoS). Thus, the measurement of QoS in MBB networks has become a key issue. In this regard, the proper performance of the transport layer constitutes a critical feature in order to fulfill the QoS requirements of the clients (user equipments—UE) [1, 2]. In MBB, due to the multiple sources of variability (related to the client or self-inflicted effects such as channel quality reporting, propagation and fading pattern alterations due to mobility—and related to the intrinsic features of MBB such as bandwidth sharing, modulation and so forth), the network conditions become more volatile than in fixed networks and therefore the accuracy of transport protocols to adapt their sending rate as close to the available capacity as possible is reduced, impacting on the final performance [1, 2]. Therefore, there is an urgent need to study the relation between the transport protocol performance in MBB and its impact on the actual QoS results.

We consider that the mobility is one of the biggest features that differentiates cellular access from other connectivity schemes. Besides, the wide range of moving possibilities makes every use-case distinct and independent, creating different network conditions for transport protocols. In this regard, it is important to focus the performance-based analysis of current state-of-the-art transport protocol solutions over different mobile network mobility circumstances and understand the implications of the movement in the interaction between the transport protocol and the MBB variability.

Regarding mobile networks, this work analyses different schemes in order to give a wider view of the impact that the performance of transport protocols have in the QoS: 4G scenarios and low latency scenarios targeting assisted by mobile edge computing (MEC-assisted) future 5G deployments [3]. Future 5G networks aim at allowing improved capabilities in terms of achievable capacities, modulation and end-to-end latency among other features. The reduction in the transmission latency is one of the main beneficial evolutions for the suitable performance of transport protocols. It shortens the feedback time between consecutive management decisions in the server, increasing the responsiveness of the transport layer to the fluctuations of the radio side. Therefore, we focus our second scheme in the proximity service of 5G deployments. To that end, we mimic with low latency 4G scenarios a 5G-alike service provisioning. All in all, our evaluation covers the performance of transport protocols in current and future MBB.

Considering that TCP is the predominant transport protocol on the Internet, we focus our study in the evaluation of TCP over distinct mobility circumstances over 4G latencies and low latency deployments. TCP is not a single entity but a family of different congestion control algorithms (CCAs) that manage the outstanding data of the server (clamped by the congestion window—CWND) in a different way based on pre-defined features such as throughput maximize algorithms with loss-based mechanisms, delay-aware implementations or hybrid developments. So far, despite many CCAs being available, none of them have demonstrated to both be easily deployable and appropriately face the variability of MBB fluctuation. This work selects and evaluates the performance of distinct CCAs that count on different features and implementations that in the end result in a different performance outcome in each precise network conditions.

The great success of TCP and user datagram protocol (UDP) have led to the widespread utilization of both of them, either as the selected transport solution or as a substrate to enable the so-called transport services [4–6]. These transport services are ad-hoc layers that work between the transport layer and the application layer, taking advantage of the substrate transport protocol (mainly TCP and UDP) and gaining some freedom due to its development in the user space of the operating system (OS) and additional functionalities (i.e. congestion control of QUIC over UDP). However, the utilization of TCP and UDP forces the system to stick with the infrastructural characteristics of the selected substrate transport protocol from the beginning of the transmission until it is closed. This limitation has been named "ossification" and it has three main effects: (1) it closes the opportunity to select and modify transport layer protocols at the beginning of a certain transmissions; (2) it leaves little room for transport protocol innovation; (3) it provides limited or non-existent flexibility of the application programming interface (API) [5]. This API serves as the connection point between the application layer and the transport layer. The existence of constraints in the communication between these two layers is directly translated to a standard behavior of the transport protocol with no consideration of the requirements from the application layer. In current MBB, the transport protocols misbehave due to the incapability of adapting to the actual network circumstances and the impossibility of adapting its features to the requirements from upper layers.

In a close future with the implementation of evolved transport services, the API would not only select the best transport protocol based on application requirements, but it would also consider the network status. Even though, this mechanism would require further signaling and interaction, recent advances [7] are evolving in this sense and could provide with a more complex and complete API. Taking into account that each CCA could be more suitable for certain network circumstances or application requirements than others, our work evaluates the best CCA candidate for each combination of conditions. In this regard, out of all possibilities, our analysis covers the study of TCP CUBIC, NewReno, Illinois, Westwood+ and the recently released BBR. Thus, being capable of providing hints in the complex process of improving the behavior in the transport layer and therefore in the resultant enhanced QoS.

The main findings of the chapter are the following. The mobility scenarios under 4G latencies require more aggressive TCP solutions to overcome the high variability in comparison with low latency conditions. Merely focusing on goodput: (1) although BBR provides with the best scalability, it also induces greater mean delays and lost packets; (2) in scenarios that evolve with continuous capacity reductions, Illinois shows the best adaptability to the variable conditions when the achievable capacities are high, whereas CUBIC demonstrates the same for lower bandwidth assignments. Considering the performance with a combination of goodput, delay and retransmissions, CUBIC presents a more balanced behavior with average achieved rates but greater awareness of the self-inflicted delay and retransmissions. With reduced end-to-end latencies, the most suitable CCAs are: (1) NewReno for low available capacity circumstances that moves with continuous capacity increases; (2) BBR as the most balanced CCA that allows both high bandwidth achievement and low delay and retransmissions; (3) CUBIC when scalability is required in presence of big changes between consecutive samples of assigned radio capacity and (4) similar goodput-based performance of all CCAs while moving to worse quality positions.

The chapter is structured as follows: Section 2 introduces the related work in the analysis of TCP in MBB in general and with mobility circumstances in particular. Section 3 shortly describes the utilized CCAs in the analysis. Section 4 covers the methodology with the description of the testbed and the utilized measurement and evaluation process. The analysis and results are explained in Section 5 divided by the 4G latency schemes and the low latency scenarios that mimic 5G deployments. Finally, Section 6 gathers the most important conclusions and proposes future lines.

2. Related work

When analyzing how the mobile networks' features have an effect on transport protocol behavior and therefore impact the final QoS, there are several characteristics that have to be mentioned.

2.1. Delay

It is clear that comparing mobile networks and fixed networks, the former has more variable channel conditions that could lead to achieve a degraded throughput [1, 2]. However, there are effects that from a macroscopic point of view are shared among the distinct networks. For instance, it has been proved [8, 9] that even mobile networks suffer due to the excessive buffering in intermediate queues leading to an increase in the end-to-end delays and dropped packages that severely impact the performance of TCP (bufferbloat effect). Measurements over both 3G and 4G cellular networks of four U.S. providers and Swedish networks have concluded that bufferbloat represent a problem in MBB too. Our study not only considers the achieved capacities but also the induced delay due to the possible impact that may well have as a cross-traffic.

2.2. Impact of variability

It has been demonstrated that there are differences among distinct mobile networks. A comparative work of 3.5G and 4G [10] showed that 4G networks are worse in regards to the TCP efficiency due to the superior throughput and variability. This is the higher variability, the worse scenario for TCP due to the lack of rapid adaptability. Garcia et al. [11] carried out measurements in the cellular networks of different Swedish operators so as to analyze the variation of TCP throughput and delay throughout the day. Sudden increases in traffic load leads to bandwidth variability and latency increment [2]. Therefore, TCP happens to drastically reduce its throughput. Additionally, TCP experiences timeouts many times. The timeout events are especially harmful because the CWND is reduced to one segment. In another study, Alfredsson et al. [12] proved that the variable modulation on the 4G link layer is contributing to retransmissions' increment and therefore higher delay and less throughput. Huang et al. [13] carried out a comprehensive study related to TCP throughput and latency estimation over a live long term evolution (LTE) network. In their measurements, they found out similar timeout events. Our work precisely evaluates how different variable mobility circumstances affect the performance of different CCAs in order to better understand the implications of fluctuations in the channel quality.

2.3. Uplink impact and bursty behavior

Regarding the evaluation of TCP throughput, it has been proved [14] that the uplink performance tends to degrade its performance due to scheduling policies, severely impacting on Acknowledge packet (ACK) arrival and therefore downlink injection ability. Those TCP flavors that merely depend its CWND management upon the reception of ACKs are drastically affected by these ACKs reception, also called ACK-compression. Related to this issue, it has been demonstrated [15] that modern cellular networks' traffic has a tendency to become bursty. For this reason, there can be a large variation in the actual throughput during a short period of time (varying by up to two orders of magnitude within a 10-min interval). This variability could be even harder due to the fact that mobile providers often maintain a large and individual downlink buffer for each UE, provoking high latency instability. Our analysis aims at detecting the most suitable CCA for different MBB mobility circumstances that also implicate distinct bursty conditions over the network.

2.4. Impact of speed

If mobile networks themselves suffer high variability, the channel conditions could be even more variable and challenging for TCP due to the movement of UEs. The movement leads to have distinct propagations and fading patterns over time that at the same time impact on the assigned modulation to the UE, provoking "jumps" between consecutive channel quality reports. Merz et al. [16] studied the performance of TCP in live LTE networks in mobility scenarios with speeds up to 200 km/h. They mainly evaluated the spectral efficiency depending on the modulation and the bandwidth share among the attached users to the eNodeB, together with the ability of those users to make the most of the assigned capacities. Li et al. [17] compared the performance of TCP in static positions with moving scenarios, resulting in harmful RTT spikes, massive dropped packets and eventual disconnections while on the move. Our work complements the mentioned studies by adding the evaluation of multiple state-of-the-art CCAs as well as the inclusion of different MBB schemes with 4G latencies and low latencies.

Taking into account the research and standardization momentum regarding 5G in which most of the work is yet to be fulfilled, there are few works that have considered the performance of TCP in the future 5G MBB. Pedersen et al. [18] demonstrated the potential of using different transmission-time intervals (TTI) in the eNodeB of 5G deployments depending on the metadata related to a certain channel. They showed that shorter TTIs were capable of allowing higher throughputs for short communications, whereas longer TTIs could overall benefit the performance of large transmissions. Sarret et al. [19] study the forthcoming benefit of using full duplex at the radio link layer (RLC) in comparison with the current half-duplex implementation for an improved throughput and delay. Besides, they covered the possible configurations in ultra-dense 5G deployments that could limit the envisioned rates.

Even though several research studies and proposals have reported their concerns and findings regarding the effects between mobile networks and TCP under challenging conditions, none of them have considered the analysis of distinct state-of-the-art CCAs of TCP under different mobility patterns and circumstances over 4G latencies and low latencies targeting proximity MEC-assisted provisioning in 5G.

3. Selected CCAs

The studied TCP variants fall into five categories with regard to their employed CCAs: loss-based, combined loss- and delay-based (with or without bandwidth estimation), and delay-based. As examples of loss-based CCAs, we study both NewReno [20] and CUBIC [21]. NewReno was selected due to its prevalence in research and its large implementation base, and CUBIC by the fact that it is the default CCA in Linux. The Westwood+ [22] congestion control was taken as a CCA example of a combined loss- and delay-based with bandwidth estimation technique. In many ways, TCP Westwood and its successor TCP Westwood+ laid the foundation for the work on designing a CCA that is able to distinguish between congestion and non-congestion related packet losses in wireless networks without any support from the wireless MAC layer. Also, Illinois [23] was selected as an example of a combined loss- and delay-based CCA. In contrast to Westwood+, Illinois primarily targets high-speed and long-delay networks. Finally, TCP BBR (Bottleneck Bandwidth and RTT) [24, 25] constitutes a model-based CCA that drives the congestion avoidance management based on two parameters: measured baseline RTT (delay-based) and the timing and rate of ACK packets (bandwidth estimation). A brief overview of the five studied TCP variants is given below.

TCP NewReno [20] is the basic TCP implementation that drives its CWND based on the additive increase and multiplicative decrease (AIMD) principle. In this regard, NewReno increases its CWND by one packet for each ACK reception during the Slow Start phase. Instead, during the congestion avoidance phase, it increases the CWND by one segment for each RTT. The increment is performed until a timeout period is consumed or a notification of a loss packet is received (with a triple duplicate ACK—3DUPACK). Depending on the event, NewReno would back-off differently, halving the CWND in case of 3DUPACK and establishing the CWND in one segment when a timeout is detected.

TCP CUBIC [21] uses a cubic equation during the congestion avoidance phase to manage the CWND. The closer the CWND is to the previous congestion point in terms of outstanding packets, the slower increment is applied. This function leads to a zero increment while the previous congestion point is achieved. If CUBIC does not detect congestion at that point, it increases the ramp-up pace of the CWND with a convex shape until a new loss event happens. One of the main features of CUBIC in comparison with NewReno is that the CWND is not ACK-clocked and therefore, depend less significantly in the RTT. CUBIC also introduces a modification in the Slow Start phase in order to avoid massive packet losses at the end of the ramp-up. The modification is called Hybrid Slow Start and tries to transfer the management of the CWND to the congestion avoidance phase prior to overfeed the network. To that end, two exit conditions are added: (1) if a delay increase over a predefined threshold is detected and (2) if the ACK train is lengthen. If any of the conditions is met, the Slow Start phase is left and the congestion avoidance phase would follow driving the CWND.

TCP Westwood+ [22] is a sender-side modification of TCP to allow estimating the available bandwidth by assessing the incoming ACK packets. The measured capacity serves to adjust the CWND during back-off phases after a loss event occurs. The selected CWND tries to establish the potential outstanding packets as close to the maximum capacity as possible but avoid building-up the queue of the bottleneck. To that end, the bandwidth-delay product (BDP) is calculated with the estimated bandwidth and the minimum assessed RTT.

TCP Illinois [23] is a loss-based AIMD mechanism that drives the CWND with certain knowledge of the queuing delay and buffer size of the bottleneck. This delay awareness is taken from the RTT measurements and consequently it is updated upon ACK arrival. If no excessive queuing delay is detected, the CWND would increase faster than in conditions of high induced latency. The maximum increment is established in 10 segments per RTT, while the minimum is set to 0.3. When the RTT is close to the maximum, the loss is considered as buffer overflow, whereas in low RTT the loss counts as packet corruption.

TCP BBR [24, 25] is the recently developed TCP implementation that bases its CWND management in a model of the bottleneck's BDP. It considers the estimated bottleneck bandwidth and the measured RTT in every update of the model. The estimated bottleneck bandwidth is measured by calculating the timing and rate of receiving ACKs in the sender. The calculated model determines whether the packet injection rate is below or over the capacity of the bottleneck, being able to appropriately adjust to the network requirements. Such an adjustment of the injection rate is carried out following the principle of *pacing*, either by using Fair Queue packet scheduling or the native and fall-back implementation of *pacing* developed in the transport layer. Besides the main behavioral features, BBR is handled with a four stages workflow:

- In the Startup stage BBR ramps-up as the Standard Slow Start until it detects that the obtained throughput gain is below the 25% throughout three consecutive RTTs.
- In the draining stage BBR tries to get rid of all the excessive packets in the bottleneck queue.
- In the probing bandwidth stage BBR uses an eight state cycle to cruise at different pacing rates. Throughout six states, BBR injects at the measured bottleneck's BDP rate if no change of the available capacity is detected. The other two states are a bandwidth probing phase with a 25% of the injection rate increment and a draining phase with the ability to drain the excess packets injected in the previous phase if the bottleneck does not tolerate greater throughputs.

- In the probing RTT stage, BBR re-measures the baseline RTT for a proper modeling of the network path. To this end, BBR reduces the CWND to four segments for at least 200 ms and then, re-established the previous CWND.

4. Methodology

This section presents the measurement testbed with the equipment involved in the assessment process and describes the measurement and evaluation procedure that have led to the results presented in the following analytical section.

4.1. Testbed

The LTE deployment uses a digital radio testing emulator or a LTE-in-a-box as the main equipment responsible for the LTE side. The main radio configuration parameters are the utilization of the seventh band of LTE due to its widespread usage and the availability of 100 physical resource blocks (PRB) and 20 MHz channels in order to be capable of performing at the full potential of the cell. This emulator plays among other attributes (full EUTRAN/EPC testbed) the role of the eNodeB, creating the LTE signaling to support the attachment and registration of any LTE device through a radiofrequency (RF) cable.

Figure 1 shows the experimental testbed and how the LTE-in-a-box is placed and connected to other equipment in the deployment. Apart from the LTE emulator, the testbed is formed by different parts:

A LTE UE is included with the capability of connecting to the network through the RF cable. Such connection is directly done to avoid undesired effects of the environment in the transmission.

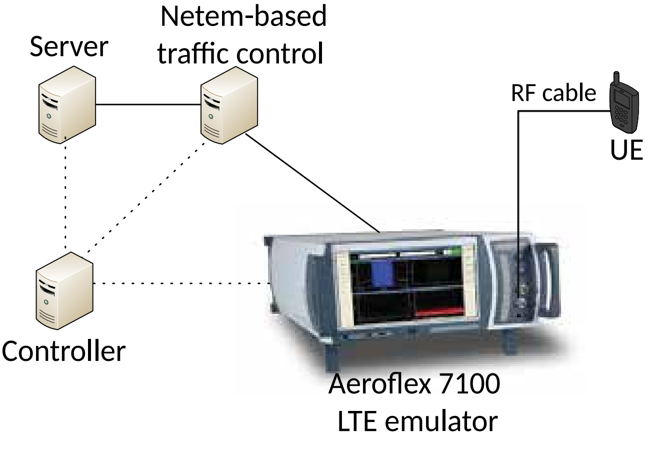

Figure 1. LTE testbed with Aeroflex 7100 LTE-in-a-box.

http://dx.doi.org/10.5772/intechopen.71779

A Linux server is placed with a 4.10 kernel that contains the recent advances in the transport layer as well as TCP BBR. Besides, the server contains the required files that are used during the experimental phase. The files are devoted to ensure long experiments or greedy sources so as to appropriately detect differences in the behavior and performance of distinct CCAs. In addition, the server is responsible for gathering the logging files created by the *ss* (socket information) and *tcpdump* (whole pcap file of each transmission) programs that later on will be used for the analysis.

One traffic-control bottleneck server is located between the end-server and the LTE-in-a-box to manage the end-to-end delay with *netem*.

A controller is used to automatize and synchronize the rest of the equipment throughout the experiments. The controller is also responsible for commanding the precise configuration that each part of the testbed should apply. For instance, the controller selects which CCA is used in the end-server, the delay of the bottleneck, the baseline signal-to-noise-plus-interference (SINR) of the channel between the LTE-in-a-box and the UE (either as a fixed value or with variations that could allow emulated mobility) or the fading pattern to be utilized in the mentioned channel.

The emulated testbed enables representative reporting and signaling with real UEs, gives the opportunity to configure the LTE-in-a-box including fading patterns and it is capable of collecting logging traces. All this features make the selected testbed realist enough to be a representation of certain real-world network circumstances with the additional control and parameterization of the measurements outcome that a close testbed provides.

4.2. Measurement and evaluation procedure

In order to experiment with 4G latency MBB scenarios as well as low latency schemes that model the proximity service of MEC-assisted 5G future deployments, we have used different delays in the network path. Due to the privacy and non-disclosure information of operators, there is little data regarding the latencies present in 4G and the ones expected in real-world deployments for 5G. Therefore, our study is based on a report [26] that shows the delay results of four operators being between 68 and 85 ms on average for 4G. Thus, we configure our 4G latency scenarios with a minimum latency of 68 ms and the low latency scheme with the lowest possible value in our testbed, 18 ms.

The emulated effect of movement is obtained by the application of two parameters: (1) a selected fading pattern in the LTE-in-a-box that would affect the channel between the eNodeB and the UE and (2) the external SINR traces that periodically command the baseline SINR. These messages are sent by the controller and applied by the emulator in order to modify the baseline SINR of the channel towards the UE.

Regarding fading, since different maximum achievable rates and variability lead to different challenges for TCP, we have applied two distinct fading patterns in regards to modeling common fading effect under mobility in real deployments. The patterns are the following:

- A mobility scenario with Extended Vehicular A model 60 (EVA60) fading model. This fading and variability tries to mimic the vehicular scenarios at 60 km/h, which is a common limitation in rural roads.
- A mobility scenario with High Speed Train (HST) 300 (HST300) fading pattern. The selected fading models the signal fluctuation of current highspeed trains at 300 km/h.

Figure 2 shows the distribution of CQIs while applying the combination of a baseline SINR of 20 dB and the selected fadings. It is clear that both patterns depict a variable behavior even under static baseline SINR circumstances. However, both fading patterns result in a completely different environment for TCP. EVA60 demonstrates a variable behavior in high CQI values (high capacities), whereas HST300 shows even more fluctuation in lower CQI values (lower achievable rates).

Apart from the variability that the fading traces create in the radio link, in order to run under different mobility patterns, the Controller in the testbed would periodically command the baseline SINR of the channel to the emulator. Our experiments evaluate the responsiveness and suitability of the selected CCAs in two simplified mobility conditions:

- Forward movement or out-cell: this mobility pattern evaluates the performance of different flavors of TCP with a UE moving from the eNodeB towards a constantly worsened radio channel conditions. Therefore, the applied mobility traces would start from the baseline SINR of 20 dB and would go progressively worsening until 5 dBs are reached. The speed of such transition is determined by the scenario under study and thus, the 60 km/h mobility scenario spends five times the time is needed to perform the same at 300 km/h. The main idea is to evaluate TCP under variable conditions that progress with an average continuous capacity reduction.
- Backward movement or in-cell: this mobility pattern evaluates the performance of the selected CCAs with a UE moving in the other way around. The mobility traces would start by applying a baseline SINR of 5 dBs to the channel and the quality would progress increasing in accordance to the selected scenario and its modeled speed. The idea is to assess the performance of TCP in variable channel conditions when the quality of the channel tends to continuously increase its available capacity.

We understand that the simplification of movement into backward movement and forward movement does not directly represent the mobility circumstances in the real-world. However, we stand that many movement patterns can be divided or split into the aforementioned cases.

In regards to the utilized traffic, greedy sources are employed with no cross-traffic. The decision of performing with isolated flows is based on the better understanding of the impact that different mobility patterns and fading traces have, together with the better detection of the CCAs' adaptability.

In order to analyze NewReno, CUBIC, Westwood+, Illinois and BBR in the selected mobility scenarios under 4G latencies and low latencies that model the potential MEC-assisted service

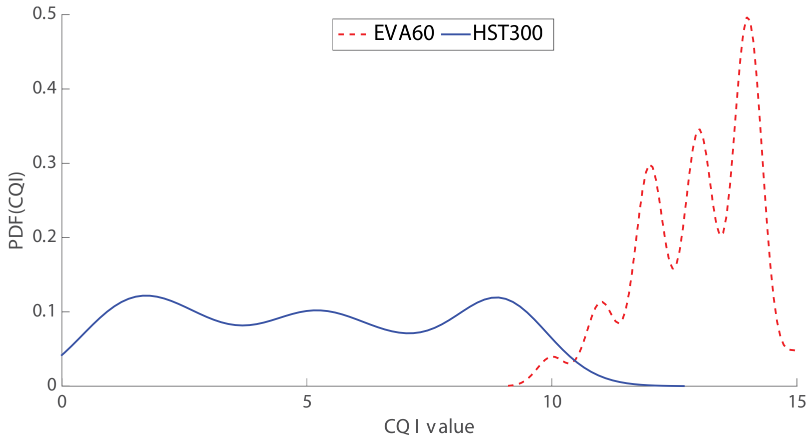

Figure 2. CQI distribution while combining a baseline SINR of 20 dBs and the selected fading patterns.

provisioning delay in 5G, we consider different metrics in the evaluation process. Instead of merely measuring the performance only based on the goodput performance, two different approaches are followed:

1. Criterion 1—pure goodput performance in order to assess the adaptability of CCAs and which are the rates capable of achieving.

2. Criterion 2—a performance metric that includes not only the goodput samples but the delay as well as the retransmission events. This way, a single value is able to include several important fields of the performance of a certain CCA. Thus, the evaluation could consider a more balanced parameter that does not rely on a single performance side, avoiding the cyclic dependency of TCP (i.e. in many situations the achievement of greater goodput—positive fact, also means the injection of greater delay—negative impact). In order to build a metric that considers all abovementioned parameters, we decided to create the following one:

$$A = \frac{K}{Kt} \tag{1}$$

$$B = \frac{Dmin}{D} \tag{2}$$

$$C = \frac{BDP - MSS * R}{BDP} \tag{3}$$

$$\alpha = 1; \beta = 1; \gamma = 1 \tag{4}$$

$$Pm = \frac{(\alpha * A) + (\beta * B) + (\gamma * C)}{\alpha + \beta + \gamma} \tag{5}$$

The first parameter (Eq. (1)) measures how much out of the available capacity (Kt) is reached in a precise sample (K). The second one (Eq. (2)) indicates the growth of current delay (D) considering the baseline delay of the transmission ($Dmin$). The third one (Eq. (3)) takes into account out of the current BDP (BDP), how many bytes are wasted in retransmissions (number of retransmissions—R times the maximum segment size—MSS). α, β and γ (Eq. ((4)) are parameters to weight the importance of the three principal. These parameters could be configured for the precise requirements of a certain application (critical with losses or critical with delays among other options). In our evaluation (Eq. (5)), our performance metric (Pm) considers every main parameter equally important to get a balanced performance.

The evaluation process aims at deciding the most appropriate options among selected CCAs depending on the selected mobility use-case and application requirement (evaluation criterion). The analysis and evaluation is linked with the current and state-of-the-art possibilities that are being opened in the transport layer so as to select the most suitable protocol at the beginning of every transmission [7]. Our work contributes to the better selection of CCAs under 4G latency and low latency MBB mobility scenarios.

5. Analysis

This section covers the overall performance and most suitable selection of TCP under mobility in 4G latencies and under low latency scenarios. In order to better explain the analysis, the section is divided into two subsections: (1) Section 5.1 covers the analysis regarding the performance of TCP under 4G latencies in mobility scenarios, together with the evaluation and selection of most TCP candidate considering goodput requirements and performance metric requirements; (2) Section 5.2 performs the same analysis and evaluation in low latency mobility scenarios. Each subsection covers three main steps in the explanation, it presents the overall performance, the goodput-based evaluation and the assessment based on the performance metric.

Since throughout the analysis we will use a performance metric that also considers the injected delay and number of retransmissions, the representation of the overall performance will take into account three parameters: mean goodput, mean delay and mean retransmissions per time slots of 100 ms. In order to avoid complicated graphs that may well be misinterpreted, the overall performance is depicted in spider plots of three axis, one per performance parameter. Each figure comprises four subplots: the first two subplots on the top depict the results for forward movement pattern, whereas, the bottom line shows another two subplots for backward movement pattern. The two subplots per line are representative of the different speed and fading conditions: from the left to the right, the scenarios at 60 and 300 km/h.

Supported by the spider plot of the overall performance and adding a table that gathers the mean values and the average confidence intervals of goodput in each mobility scenario, the most appropriate CCA is selected. In this sense, the selection may well be utilized by requirement of the applications to pick among all CCAs the option that best maximizes the achieved rates in a precise mobility scheme.

Once selected the best candidates for mobility scenarios merely based on goodput, it is important to carry out a similar task but evaluating the performance of distinct TCP flavor from a point of view that gathers more information about the outcome and possible side-effects. To that end, a figure will depict the empirical cumulative distribution function (ECDF) results of the performance metric in all scenarios. The distribution of subplots is the same as before with forward movement scenarios in the first row and backward movement ones in the bottom line. Each column represents a different speed and fading combination, being from the left to the right the scenarios at 60 and 300 km/h, respectively.

5.1. TCP performance with mobility under 4G latencies

This section covers the analysis under 4G latencies of the selected movement patterns of five CCAs. **Figure 3** shows the spider plot results in mobility scenarios for CUBIC (straight line), NewReno (dotted line), Westwood+ (straight thick line), Illinois (dash-dotted line) and BBR (dashed line).

Out of all the results in **Figure 3**, the most important ones are the following:

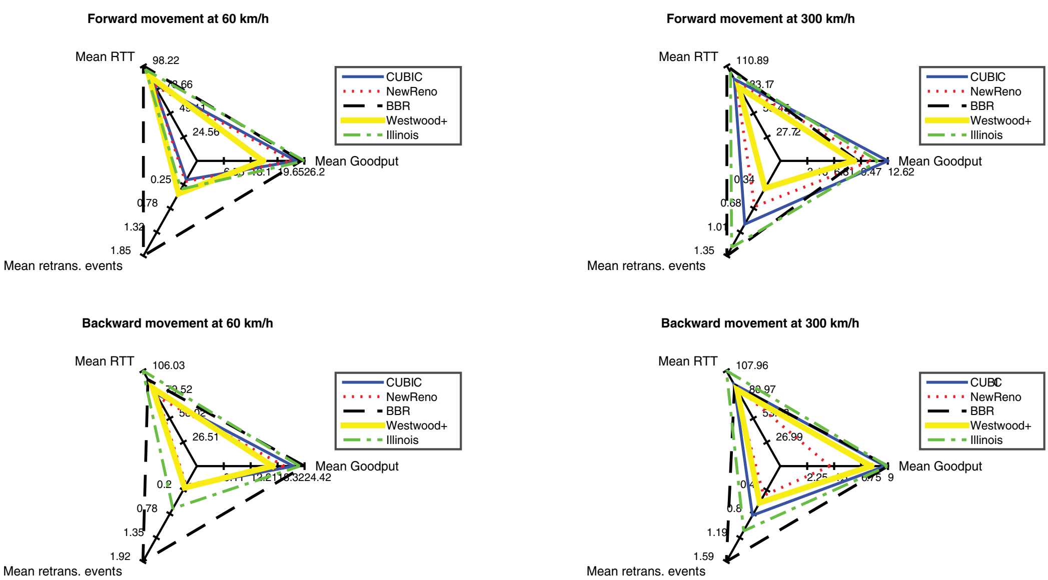

Figure 3. Performance spider plot of five CCAs while moving at different speeds under 4G latencies: forward movement on top; backward movement on bottom line.

- In forward movement pattern: We find two different cases: (1) when the variability happens at high capacities (scenario at 60 km/h) since the movement itself regardless the CCA under use allows having almost throughout the whole experiments packets in-flight, the mean achieved goodput is very similar in most of CCA cases. The best candidates are Illinois and BBR in terms of goodput but they also induce more delay and retransmissions (more significant with BBR) than NewReno or CUBIC; (2) at 300 km/h where the mean fluctuation is harder due to the superior speed, the self-inflicted effects of delay and retransmissions severely impact the goodput performance, dropping the goodput performance of BBR and Illinois and being clearly surpassed by CUBIC; (3) as detected in Ref. [27] Westwood+ suffers due to the excessive reduction in the back-off application after Slow Start, leading to an underuse of the radio capacity.
- In backward movement pattern: (1) both Illinois and BBR has shown better scalability than CUBIC in all scenarios achieving greater goodput rates, but as a drawback, inducing more delay in the network and suffering more retransmissions; (2) the scenario itself due to its low available rates at the beginning of the transmission helps appropriately perform weak CCAs such as Westwood+. In this sense, the deficiency of Westwood+ is minimized, leading to better results in terms of goodput while the delay is kept low; (3) considering goodput, apart from NewReno that demonstrates to underperform in variability circumstances when scalability is required under 4G latencies, the rest of the candidates achieve similar (not equal) results. The scenario at 60 km/h shows bigger differences among the CCA candidates due to the greater capacities present in the scenario, allowing more aggressive CCAs adapt better to the available bandwidth.

We have detected cases in which a similar goodput is achieved but significantly more delay and retransmissions are suffered. These examples, that suppose a difficult performance trade-off to analyze, are the foundation for the evaluation of the protocols based on different points of view in the performance in order to appropriately select the best candidate for each network circumstances but also considering the application requirements.

The overall performance has depicted the goodput performance as one of the parameters in the spider plot. Now instead, **Table 1** covers the performance of the goodput showing the actual average values of the transmission throughout each mobility scenario, together with the average confidence interval as a representation of the differences between independent tests.

Table 1 shows that depending on the scenario a proper CCA selection could allow the achievement of greater capacities. The best practises regarding the goodput-based evaluation are the following under 4G latencies:

- In backward movement pattern (right side of the table), BBR shows great variability in the behavior. Even with that, it is able to scale better than any other candidate. This ability could be of a great value in other MBB scenarios in which scalability is required, either as a necessary feature or as a response to available bandwidth increments.

Context	Forward movement		Backward movement	
	Mean	Mean CI	Mean	Mean CI
60 km/h with EVA60 under 4G latencies				
CUBIC	24.18	6.59	21.87	6.89
NewReno	22.95	7.95	19.87	7.16
BBR	25.43	18.7	24.42	17.54
Westwood+	16.08	8.01	17.16	7.68
Illinois	26.2	10.7	23.35	14.84
300 km/h with HST300 under 4G latencies				
CUBIC	12.62	7.81	8.76	5.51
NewReno	10.63	7.58	4.15	2.89
BBR	9.09	10.21	9	6.76
Westwood+	8.61	4.13	7.58	4.28
Illinois	11.48	6.8	8.95	5.84

Table 1. Goodput-based evaluation of the selected CCAs in mobility scenarios under 4G latencies.

- In forward movement pattern (left side of the table), the two different scenarios require a distinct treatment and therefore CCA candidate. The best practises in forward movement pattern under 4G latencies are the following:

   1. Illinois is selected in scenarios with big bandwidth jumps between consecutive RTTs due to its aggressiveness to handle such variability but also due to its delay awareness. The combination of features provides the best and most consistent (less variation in the behavior than the second candidate according to the results—BBR) goodput performance.
   2. CUBIC is picked in variable scenarios with smaller changes between consecutive available bandwidth samples. The aggressive features of CUBIC are suitable to handle great fluctuation in smaller capacities but do not make the most when it comes to bigger bandwidth jumps.

Table 1 reveals that in forward movement, an appropriate CCA selection could provide with a great gain in terms of goodput even using a movement pattern that makes easier the task of TCP due to the avoidance of starvation events in the bottleneck buffer.

Once selected the best candidates for mobility scenarios under 4G latencies merely based on goodput, it is important to carry out a similar task but evaluating the performance of distinct TCP flavor from a point of view that gathers more information about the TCP performance. To that end, **Figure 4** depicts the ECDF results of the performance metric in all scenarios. The

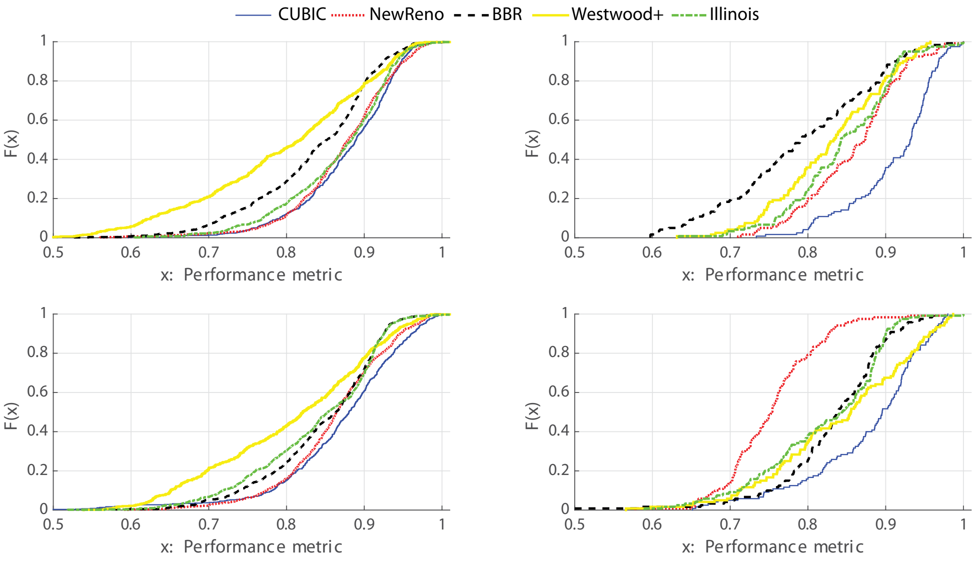

Figure 4. Comparison of five CCAs performance metric while movement at different speeds under realistic 4G latencies: forward movement on top; backward movement on bottom line.

distribution of subplots is the same as before with forward movement scenarios in the first row and backward movement ones in the bottom line.

Figure 4 contrasts with previous goodput-based evaluation in many ways. It shows that sometimes the TCP performance depends on the proportion between parameters, suffering a performance cycle that impacts negatively certain performance fields when others are improved and vice-versa. In general in forward movement, **Figure 4** demonstrates that the performance enhancements are not that clear in one CCA over the other. Maybe one CCA is able to perform better in terms of goodput but at the same time suffering from delay and retransmitted packets. Therefore, taking into account the three selected performance factors, it is not that easy to decide whether we should use one congestion control or the other. If the decision of the CCA is taken upon the evaluation of this performance metric, in the scenarios with similar outcome, the final CCA would be picked based on other circumstances. For instance, the decision is taken with: (a) developer preference; (b) the by-default CCA prevails over the others; (c) if at the point of selection there is one CCA already established and it is among the group with similar outcome, the CCA could remain selected. In contrast and still selecting the most appropriate CCA in forward movement pattern, the 300 km/h scenario is clearly dominated by CUBIC. CUBIC not only achieves a great goodput, but shows a more general and extensive good performance.

Regarding the results in backward movement scenarios, there is a clear pattern that reflects that CUBIC is more appropriate. The outcome contrasts the evaluation based on goodput that suggests the selection of BBR, proposing to pick CUBIC as a candidate that improves the overall performance based on the three selected performance aspects: goodput, delay and impact of retransmissions.

http://dx.doi.org/10.5772/intechopen.71779

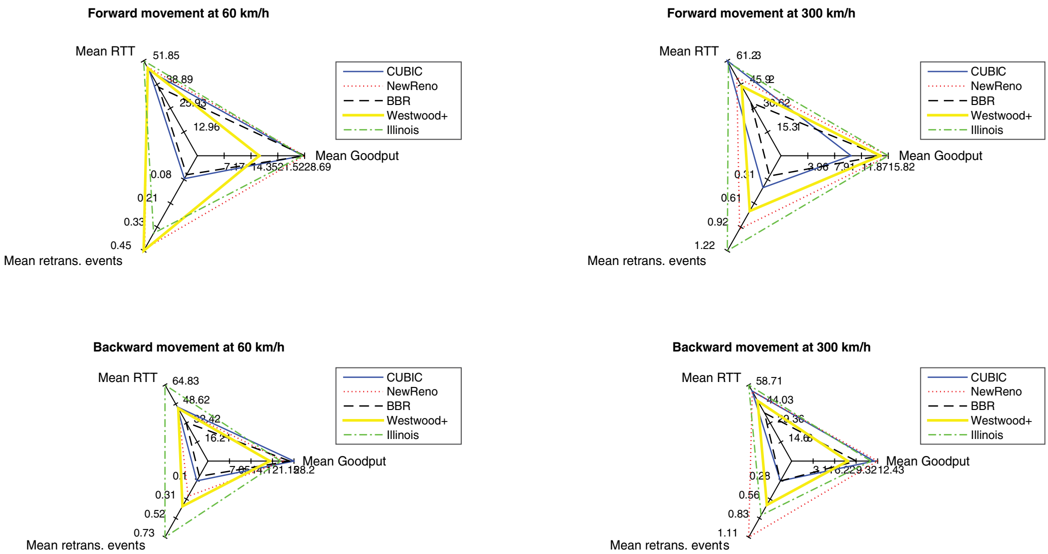

Figure 5. Performance spider plot of five CCAs while moving at different speeds under low latencies: forward movement on top; backward movement on bottom line.

5.2. Overall performance under mobility under low latency

Once explained the performance of the selected CCAs over the mobility scenarios under 4G latencies and having evaluated the best TCP candidate considering the goodput and the performance metric, this subsection covers the analysis under low latencies. **Figure 5** shows the spider plot results in mobility scenarios for CUBIC (straight line), NewReno (dotted line), Westwood+ (straight thick line), Illinois (dash-dotted line) and BBR (dashed line).

The most important details in **Figure 5** are the following:

- As a general performance and comparing with 4G latencies, Illinois still suffers due to an excessive delay and number of retransmissions, whereas BBR shows a more controlled behavior with a good performance in some scenarios in terms of goodput but also avoiding massive delay and retransmitted packets.
- In forward movement pattern: (1) there still exist huge similarities among CCA from the goodput's point of view. Under low latency circumstances, even the CCAs that in principle are less aggressive (i.e. NewReno in both scenarios or Westwood+ at 300 km/h), demonstrate to perform very similar to more aggressive candidates, thus achieving a great goodput performance but suffering from excessive delay and retransmissions; (2) moving at 300 km/h where the achievable capacities are low, aggressive solutions such as CUBIC obtain worse results than others due to their injection pattern. Even though the mean retransmission events are lower than Illinois or NewReno, the periodicity is different. In each "big" retransmission event of Illinois or NewReno, many retransmitted packets are involved, whereas in the case of CUBIC the "big" retransmission events are more but with less packets involved. Thus, CUBIC due to its aggressiveness, finds

more times "big" retransmission events (increasing the suffered mean RTT), reducing in more occasions the CWND and therefore, being less capable of achieving the available capacity.

- In backward movement pattern: (1) Westwood+ and Illinois are not able to perform better than CUBIC or NewReno in any case. The former primarily due to its poor ability to scale. The latter, due to its self-inflicted effects, suffering by far the greatest delay; (2) in the case of BBR, as mentioned earlier, the behavior under low latency looks more conservative. In this regard, in backward movement pattern that requires scalability, in terms of goodput BBR performs worse than CUBIC in the mobility scenario at 60 km/h and underperforms in the scenario at 300 km/h earning very poor performance results.

Once again and even more noticeable than with 4G latencies, we have detected cases in which a similar goodput is achieved but significantly more delay and retransmissions are suffered by different CCAs. Those examples require an evaluation of the protocols based on different application-layer requirements so as to appropriately select the best candidate whose performance matches the application requirements and network conditions.

Table 2 only covers the representation of the goodput with average values of the transmission for each mobility scenario and the average confidence interval for different independent tests.

Table 2 demonstrates that depending on the scenario, the selection of the most appropriate CCA is capable of allowing better performance, achieving greater capacities. All in all, under low latency conditions, the most suitable CCA regarding the goodput-based evaluation are the following:

- In forward movement, CUBIC is damaging in scenarios with drastic fluctuation in low available bandwidths comparing with other candidates and its aggressiveness looks as if does not properly fit for such scenarios. Due to the low end-to-end latency and taking into account that forward movement is an easier movement pattern to handle by TCP (because it allows having almost every time packets in-flight), we can barely decide one specific CCA in each scenario. The selection should be carried out among a group of CCA that have reported a very similar outcome in terms of goodput.

  1. NewReno and Illinois have shown a great performance in variable scenarios with low available capacities (mobility scenario at 300 km/h).

  2. Under variable conditions with big capacities and big jumps between resource assignments, all CCAs but Westwood+ have proven to achieve very similar performance. Comparing with 4G latency scenarios, these similarities clearly come due to the reduction in the end-to-end delay.

  3. In backward movement, even with similarities among the CCAs, it is still clear that one or two candidates prevail over the others. Variable scenarios are better handled by CUBIC when the capacities are higher due to its superior aggressiveness, whereas

NewReno improves the performance of variable scenarios with low available bandwidths. These results demonstrate that, yet, there is value in the performance of by-default CCA (CUBIC) and classic CCA (NewReno) in current and more importantly, future networks with low latency scenarios.

Once we have evaluated and selected based on goodput the best candidates for low latency mobility scenarios, it is crucial to evaluate the performance of the TCP flavors from the point of view of the performance metric. **Figure 6** depicts the ECDF results of the performance metric in all scenarios.

Taking into account that the current evaluation considers three different aspects of the overall performance, the results in **Figure 6** widely differ from the outcome merely based on goodput. There is a clear pattern that confirms BBR as the best candidate in forward movement scenarios. Previous evaluation has shown that based on goodput in all scenarios, BBR is within the group of best performers. Its capacity to reduce the delay close to the baseline delay in a movement pattern that strongly suffers due to the contrary, makes BBR the best candidate considering the performance metric.

In backward movement pattern, there are two clear candidates, one per each speed and fading combination:

Context	**Forward movement**		**Backward movement**	
	Mean	**Mean CI**	**Mean**	**Mean CI**
60 km/h with EVA60 under low latencies				
CUBIC	28.18	8.84	28.21	9.02
NewReno	28.68	9.4	24.89	10.36
BBR	28.69	8.29	26.84	8.41
Westwood+	16.57	10.83	20.3	12.27
Illinois	28.19	9.1	23.81	11.78
300 km/h with HST300 under low latencies				
CUBIC	10.33	5.98	11.91	5.4
NewReno	15.79	5.86	12.43	4.97
BBR	14.66	5.38	9.35	5.32
Westwood+	14.66	4.76	8.11	5.58
Illinois	15.82	5.48	11.46	4.92

Table 2. Goodput-based evaluation of the selected CCAs in mobility scenarios under low latencies.

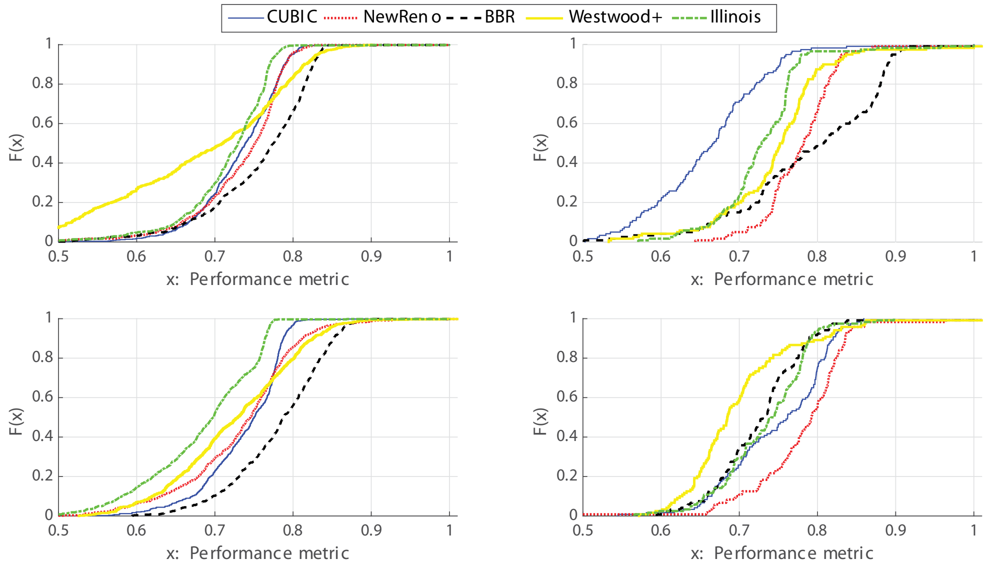

Figure 6. Comparison of five CCAs performance metric while movement at different speeds under low latencies: forward movement on top; backward movement on bottom line.

1. BBR is able to better handle the required scalability under variable fading conditions with big available capacities and jumps between consecutive assignment samples. As under 4G latencies, BBR demonstrates great scalability. In addition, under low latency circumstances BBR is capable of better managing the induced delay, achieving a considerably better performance than the rest of the TCP implementations in the commented network use-case.

2. The combination of low latency together with low available bandwidths makes the performance playground very suitable for the selection of NewReno, showing great performance both in achieved goodput and based on the performance metric.

Overall the results in **Figure 6** show two important facts in comparison with the mobility scenarios under 4G latencies: (1) the performance of BBR shows a consistent good performance in terms of goodput. The main difference is that in low latency scenarios the behavior of the CCA is more controlled and is able to take full advantage of its features, achieving great goodput while the delay is kept close to the baseline latency. This outcome shows that, the shorter response times (RTT) in each packet, the better for the accurate estimation of BBR in mobility scenarios; (2) based on the performance metric, it is clear that the reduction in the end-to-end latency makes Illinois perform worse in comparison with other candidates. While in 4G latencies, Illinois is eligible in many mobility circumstances that take advantage of it aggressiveness, the same scenarios under shorter latencies do not require its aggressiveness. Actually, this feature only increment the injected delay as well as the number of retransmitted packets.

6. Conclusion

After evaluating the QoS of selected mobility scenarios, based on goodput and the performance metric, under both 4G latencies and low latency that mimic the potential delays in 5G networks, we have been able to select the most appropriate CCA for each situation. The selected CCAs are gathered in **Table 3** with an asterisk for the cases in which more than one CCA could be picked.

The mobility scenarios under 4G latencies require more aggressive TCP solutions to overcome the high variability with an increased BDP in comparison with low latency conditions. With reduced end-to-end latencies that model the proximity service provisioning in 5G, the most suitable CCAs are a compendium of classic NewReno for low available capacity circumstances that move in-cell, BBR as the most balanced CCA that allows both high bandwidth achievement and low delay and retransmissions, CUBIC when scalability is required in presence of big changes between consecutive samples of assigned radio capacity and equality in goodput-based outcome in out-cell movement. In this sense, in our pseudo-5G mobility scenarios, it is important to highlight the usability of even weak CCAs such as NewReno in future low latency deployments. The explanation of the selected TCP candidates and their implications in the mobility scenarios is followed CCA-wise.

Illinois has shown an aggressiveness that is suitable for application with goodput requirements in both movement patterns in 4G latency scenarios, but only partially appropriate for forward movement under low latency conditions. The reduction in the end-to-end delay increases the responsiveness of TCP in general and Illinois in particular. This effect makes Illinois suffer greater mean delays and retransmitted packets due to the aggressive injection and ramp-up, leading to deteriorate results (i.e. achieving poor results in backward movement).

BBR has demonstrated good performance in terms of goodput under 4G latencies. However, it also induced a long delay and provokes a great number of transmitted packets, demonstrating an unbalanced behavior in terms of the performance metric. In low latency scenarios instead, BBR has a more controlled and balanced behavior, achieving a sufficiently good performance in goodput but also preserving a low delay.

Context	**Low latency**		**4G latencies**	
	Goodput	**Metric**	**Goodput**	**Metric**
60 km/h with EVA60				
Forward	*	BBR	Illinois	*
Backward	CUBIC	BBR	BBR	CUBIC
300 km/h with HST300				
Forward	*	BBR	CUBIC	CUBIC
Backward	NewReno	NewReno	BBR	CUBIC

Table 3. Wrap-up of most appropriate CCAs for each mobility scenario and QoS metric.

Westwood+ has demonstrated a poor performance in majority of the scenarios due to its aggressive back-off policy. However, in scenarios with low achievable capacities (at 300 km/h) Westwood+ is capable of slightly bridging the performance gap with other CCAs.

NewReno, under low latency, provides with a TCP implementation able to achieve close to the maximum available capacity, being especially significant its performance and scalability in backward movement pattern with low achievable rates.

CUBIC has shown a balanced performance with low available bandwidth under 4G latencies. Apart from that, under low latency circumstances, CUBIC is appropriate for forward movements with variable big available bandwidths and scenarios that depend upon scalability.

The future work will progress in three main lines. Firstly, the analysis will be extended with the measurement of similar mobility scenarios in a selection of real-word use-cases. Secondly, the work would re-measure and compare the evolution of BBR in terms of adaptability and management of losses. Finally, we will evaluate pseudo-random mobility scenarios that combine both forward and backward movement as well as static positions.

Acknowledgements

This work is partially funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644399 (MONROE) through the open call project MARiL and by the Spanish Ministerio de Economía y Competitividad (MINECO) under grant TEC2016-80090-C2-2-R (5RANVIR). The views expressed are solely those of the author(s).

Author details

Eneko Atxutegi*, Jose Oscar Fajardo and Fidel Liberal

*Address all correspondence to: eneko.atxutegi@ehu.eus

University of the Basque Country (UPV/EHU), Bilbao, Spain

References

[1] Koksal CE, Jamieson K, Telatar E, Thiran P. Impacts of channel variability on link-level throughput in wireless networks. In: Joint International Conference on Measurement and Modeling of Computer Systems, ser. SIGMETRICS '06/Performance '06; 2006. New York: ACM; 2006. pp. 51-62

[2] Nguyen B, Banerjee A, Gopalakrishnan V, Kasera S, Lee S, Shaikh A, Van der Merwe Jacobus J. Towards understanding TCP performance on LTE/EPC mobile networks. In: 4th Workshop on All Things Cellular: Operations, Applications & Challenges; 2014. ACM; 2014. pp. 41-46

[3] European Telecommunications Standards Institute (ETSI). Mobile Edge Computing (MEC); Technical Requirements [Internet]. 2016 [Updated: 2016]. Available from: http://www.etsi.org/deliver/etsi_gs/MEC/001_099/002/01.01.01_60/gs_MEC002v010101p.pdf [Accessed: 2017]

[4] The Chromium Projects. QUIC, a multiplexed stream transport over UDP [Internet]. Available from: https://www.chromium.org/quic [Accessed: 2017]

[5] Grinnemo KJ, Jones T, Fairhurst G, Ros D, Brunstrom A, Hurtig P. Towards a flexible internet transport layer architecture. In: 2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN); 2016;**2016**:1-7

[6] Jones T, Fairhurst G, Perkins C. Raising the datagram API to support transport protocol evolution. In: Workshop on Future of Internet Transport (FIT 2017); 2017. 2017

[7] Hurtig P, Alfredsson S, Bozakov Z, Brunstrom A, Evensen K, Grinnemo K-J, Fosselie Hansen A., Rozensztrauch T. A NEAT approach to mobile communication. In: The ACM SIGCOMM 2017 Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2017); 2017. ACM Press; 2017

[8] Jiang H, Liu Z, Wang Y, Lee K, Rhee I. Understanding bufferbloat in cellular networks. In: 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design; 2012. New York. ACM; 2012. pp. 1-6

[9] Alfredsson S, Del Giudice G, Garcia J, Brunstrom A, De Cicco L, Mascolo S. Impact of TCP congestion control on bufferbloat in cellular networks. In: World of Wireless, Mobile and Multimedia Networks (WoWMoM); 2013. IEEE; 2013. pp. 1-7

[10] Garcia J, Alfredsson S, Brunstrom A. A measurement based study of TCP protocol efficiency in cellular networks. In: 2014 12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt); 2014; 2014. pp. 131-136

[11] Garcia J, Alfredsson S, Brunstrom A. Examining TCP short flow performance in cellular networks through active and passive measurements. In: 5th Workshop on All Things Cellular: Operations, Applications and Challenges, ser. AllThingsCellular '15; 2015; ACM; 2015. pp. 7-12

[12] Alfredsson S, Brunstrom A, Sternad M. Cross-layer analysis of TCP performance in a 4G system. In: Software, Telecommunications and Computer Networks (SoftCOM); 2007; IEEE; 2007. pp. 1-6

[13] Huang J, Qian F, Guo Y, Zhou Y, Xu Q, Mao ZM, Sen S, Spatscheck O. An in-depth study of lte: effect of network protocol and application behavior on performance. In: ACM SIGCOMM 2013 Conference on SIGCOMM; 2013; ACM; 2013. pp. 363-374

[14] Zhang L, Okamawari T, Fujii T. Performance evaluation of end-to-end communication quality of LTE. In: Vehicular Technology Conference (VTC Spring); 2012; IEEE; 2012. pp. 1-5

[15] Xu Y, Wang Z, Leong WK, Leong B. An end-to-end measurement study of modern cellular data networks. passive and active measurement. Lecture Notes in Computer Science. 2014;**8362**:34-45

[16] Merz R, Wenger D, Scanferla D, Mauron S. Performance of LTE in a high-velocity environment: A measurement study. In: 4th Workshop on All Things Cellular: Operations, Applications, & Challenges, ser. All Things Cellular '14; 2014; New York. ACM; 2014. pp. 47-52

[17] Li L, Xu K, Wang D, Peng C, Xiao Q, Mijumbi R. A measurement study on TCP behaviors in HSPA+ networks on high-speed rails. In: 2015 IEEE Conference on Computer Communications (INFOCOM); 2015; 2015. pp. 2731-2739

[18] Pedersen KI, Niparko M, Steiner J, Oszmianski J, Mudolo L, Khosravirad SR. System level analysis of dynamic user-centric scheduling for a flexible 5G design. In: 2016 IEEE Global Communications Conference (GLOBECOM); 2016; 2016. pp. 1-6

[19] Sarret MG, Berardinelli G, Mahmood NH, Mogensen P. Impact of transport control protocol on full duplex performance in 5G networks. In: 2016 IEEE 83rd Vehicular Technology Conference (VTC Spring); 2016; IEEE; 2016. pp. 1-5

[20] Henderson T, Floyd S, Gurtov A, Nishida Y. The NewReno Modification to TCP's Fast Recovery Algorithm [Internet]. 2012 [Updated: 2012]. Available from: https://tools.ietf.org/html/rfc6582 [Accessed: 2017]

[21] Ha S, Rhee I, Xu L. CUBIC: A new TCP-friendly high-speed TCP variant. SIGOPS Operating Systems Review. 2008;**42**(5):64-74

[22] Mascolo S, Casetti C, Gerla M, Sanadidi MY, Wang R. TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links. In: 7th Annual International Conference on Mobile Computing and Networking; 2001; New York. ACM; 2001. pp. 287-297

[23] Liu S, Başar T, Srikant R. TCP-Illinois: A loss and delay-based congestion control algorithm for high-speed networks. In: 1st International Conference on Performance Evaluation Methodolgies and Tools, ser. Valuetools '06; 2006. New York: ACM; 2006

[24] Cardwell N et al. BBR: Congestion-based congestion control. Communications of the ACM. 2017;**60**(2):58-66

[25] Cardwell N, Cheng Y, Yeganeh S, Jacobson V. BBR Congestion Control [Internet]. 2017 [Updated: 2017]. Available from: http://www.ietf.org/internet-drafts/draft-cardwell-iccrg-bbr-congestion-control-00.txt [Accessed: 2017]

[26] fiercewireless. 3G/4G wireless network latency: How did Verizon, AT&T, Sprint and T-Mobile compare in Q3 2015? [Internet]. 2016 [Updated: 2016]. Available from: http://www.fiercewireless.com/special-report/3 g-4 g-wireless-network-latency-how-did-verizon-at-t-sprint-and-t-mobile-compare-q3 [Accessed: 2017]

[27] Atxutegi E, Liberal F, Grinnemo K-J, Brunstrom A, Arvidsson Å, Robert R. TCP behaviour in LTE: impact of flow start-up and mobility. In: IFIP Wireless and Mobile Networking Conference - WMNC; 11-13 July 2016; Colmar, France

Chapter 8

Network Coding-Assisted Retransmission Scheme for Video-Streaming Services over Wireless Access Networks

Aleš Švigelj and Melisa Junuzović

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71784

Abstract

Video-streaming services, such as Internet protocol television, promising the delivery of multimedia contents over wireless access networks to clients whenever and wherever, are becoming more and more popular. However, scarce radio resources, lossy characteristics of wireless links and high bandwidth demands pose the never-ending challenges for provisioning of real-time streaming services over wireless networks in a timely and reliable manner. Furthermore, a wireless channel may suffer from interference and multipath fading, which may cause random packet losses. In addition, wireless link layer does not provide a retransmission mechanism for multicast/broadcast traffic. This would significantly impact the clients' quality of experience of streaming services. Traditional unicast retransmission solutions improve client's quality, at the bandwidth expense, because every lost packet must be retransmitted separately. This chapter presents and practically evaluates a retransmission scheme for video-streaming services over last mile wireless networks. It is based on network coding techniques that increase the overall performance by means of reducing the number of physical transmissions, in comparison to traditional unicast retransmission approach, resulting in reduced bandwidth consumption. Thus, the Internet service providers can increase the number of clients over the same infrastructure or, alternatively, offer more services to the clients.

Keywords: retransmission, wireless network, broadcast, network coding, streaming, IPTV

1. Introduction

The provision and efficient delivery of high-quality real-time streaming services, such as Internet protocol television (IPTV) over wireless networks (e.g., Wi-Fi) in the last mile, introduces never-ending challenges. Easy and fast deployment and relatively low deployment costs of Wireless IEEE 802.11 networks, where one antenna covers several different clients/groups of clients, stimulate the usage of Wi-Fi technology in underdeveloped countries and rural areas. However, the bottleneck in the content delivery is due to spectrum limitations in wireless medium shifted to the last mile. Thus for ISPs, it is of paramount importance to increase the last mile access performance as in such cases, the bandwidth reduction means better services for existing clients, or alternatively, additional clients over the same infrastructure. Thus in this chapter, we are addressing the efficient and quality delivery of video-streaming services such as IPTV using the IEEE 802.11 wireless broadcast network in the last mile [1].

By using IEEE 802.11 wireless network in the last mile, the video content can be delivered using unicast or broadcast/multicast mechanism. Due to the fact that unicast mechanism, which delivers the content to clients individually and supports retransmissions (and back off) that assure reliable content delivery, is a preferred solution in currently deployed systems. On the other hand, multicasting can be used instead of unicasting in order to reduce the bandwidth. The bandwidth can be significantly reduced in particular, when multiple clients watch the same video stream, as the server, instead of sending multiple packets to different clients, has to send only one packet. All broadcast packets are delivered to all clients only under ideal wireless channel conditions. However, the packet losses frequently occur in the practical wireless environment; thus, some sort of retransmission mechanism is necessary to ensure reliability. In order to support reliability and/or reduce bandwidth consumption, different retransmission schemes have been proposed, where some of them also consider network coding (NC).

Since the pioneering work in network coding [2], numerous papers appeared on this subject and significant progress has been made in applying NC to different networks. NC has already been successfully used to increase throughput as shown in practical deployment in [3], where a reliable and scalable live streaming solution is presented. It is based on wireless multicast with real-time network coding in hyper-dense Wi-Fi spaces. A timely delivery scheme that uses a minimum amount of feedback from the clients is at the core of this approach. It generates coded repair packets, which are useful to a large number of clients. Packet loss estimation is used in the scheme, which is able to operate well with a very limited amount of feedback. Besides, all coded packets are linear combinations of original packets over the Galois field of size 2. Nevertheless, they are focusing on satisfying different clients' requests regarding the same flow, while we are focusing on satisfying different clients regarding different flows, using different approach of NC and feedbacks.

In [4], for the particular case of wireless mesh network, a network coding and scheduling scheme for transmitting several video streams over a wireless mesh network is proposed. Their main outcome is that in a network coding capable wireless network, the transmission of video streams should be optimized also, and even more important than for network throughput, for video quality. All approaches mentioned in [4–6] are designed for intersession network coding, by combining different flows. They compute a packet that must be decoded by the next hop of the head of the sender's output queue, based on a first in first out (FIFO) management for the requests of each client. However, there is no such restriction in our proposal and all client requests are taken into consideration, which significantly changes the complexity and the nature of the coding problem.

http://dx.doi.org/10.5772/intechopen.71784

Retransmission schemes based on NC are discussed in [7–9]. Their key idea, in order to combine different lost packets with network coding to achieve retransmission, is to use the feedback of lost packets information. Their approaches by broadcasting combined packets effectively save the number of transmissions and advance the efficiency of transmissions. However, in practical real-time streaming applications, those approaches have their own drawbacks and limitations.

Depending on the NC application, the implementation affects different open systems interconnection (OSI) layers. In this chapter, we propose transmission scheme that is integrated in the application layer and is adapted for video-streaming applications such as IPTV. It can be used for increasing network throughput of mobile and static wireless networks. We also evaluate its performance using practical Wi-Fi test bed that comprises a streaming server and multiple Wi-Fi clients.

The rest of the chapter is organized as follows. In this section, we introduced the video-streaming challenges in wireless medium using IPTV. Section 2 discusses the packet loss recovery techniques in streaming wireless networks. Section 3 proposes concept of transmission scheme for reliable video-streaming service using network coding and gives an overview about implementation details. Practical evaluation of the proposed solutions is described in Section 4, and finally, Section 5 concludes the chapter.

1.1. Wireless medium challenges

The main advantages of wireless networks over wired counterparts include ease of deployment and mobile clients support. However, the unique characteristics of video data amplify the difficulty of video transmission and thus impose a number of challenges. Some of the main challenges of wireless networking and their impact on video communication are discussed and highlighted in the following.

- Shadowing and multipath fading: they are common wireless effects. Shadowing is caused by obstacles between the transmitter and the receiver that attenuate signal power through absorption, reflection, scattering, and diffraction. It occurs over a relatively large time scale. On the other hand, multipath fading is due to multipath propagation: different path signals are added constructively or destructively. Multipath fading results in rapid fluctuation of signal amplitude within the order of a wavelength. The presence of shadowing and multipath fading results in time-varying channel conditions and location-dependent packet erasures. They cause burst errors in the form of multiple lost packets.
- Limited bandwidth: wireless networks are limited in capacity. Although, the 802.11 products are advertised as having a high data rate. Still, "protection" mechanisms such as binary exponential back off, rate adaptation, and protocol overheads cut the throughput by 50% or more. Moreover, due to the shared nature of the wireless medium, the actual bandwidth available to individual clients can even be much lower, posing a great obstacle for providing high bandwidth video services.
- Interference: wireless medium is shared among multiple clients; thus signals, which arrive at a receiver from other concurrent transmissions, although attenuated, cause the

interference for the receiver. It is a common effect in WLANs because they operate in the unlicensed 2.4/5 GHz industrial, scientific, and medical (ISM) frequency band. Interference affects the quality of a wireless link, and consequently, its error rate and achievable capacity.

1.2. Video-streaming challenges

Video services and applications have become the driving force in the development and widespread deployment of wireless broadband access technologies and high speed local area networks. Convenient and cable-free connectivity can be achieved by using wireless local area networking (WLAN) technologies. Using the wireless networks, it also can provide mobility to portable TVs.

As conventional service architectures, network structures and protocols have not been designed to consider the high data rate and real-time transmission requirements of digital video, they lack to provide a robust medium distribution.

Video-streaming applications have distinctive QoS requirements, such as high bandwidth requirement, delay sensitiveness, and loss tolerance. We list the challenging QoS issues as follows:

- Bandwidth: transmission of video sequences typically has a minimum bandwidth requirement in order to achieve acceptable presentation quality. Therefore, supporting the delivery of video over time-varying wireless links could be very unreliable. The challenge lies in keeping the quality degradation to a level that is hardly noticeable or tolerable while utilizing the wireless resources efficiently.
- Delay: in contrast to data transmission, which is usually not subjected to strict delay constraints, video-streaming requires bounded end-to-end delay. Each video frame needs to arrive at the client to be decoded and displayed, before its play out deadline. Otherwise, it is useless. If the video packet does not arrive on time, the play out process will have to be temporally paused, which is annoying to human eyes and deteriorates the overall streaming quality. Consequently, video-streaming applications are usually known to be very sensitive to delay.
- Packet loss: video-streaming technology is tolerant to a certain level of loss, since the visual quality will still be acceptable if the packet loss ratio is kept below a certain threshold. However, loss of packets can potentially make the presentation displeasing to human eyes, especially when some of the key video frames are lost, which could make the presentation impossible.
- Unreliability: Wi-Fi does not provide a retransmission mechanism for broadcast/multicast transmissions (with multiple clients). Wi-Fi multicast traffic could normally experience a packet loss rate of 5%. As a loss of even a single video packet can result as an error that propagates for many video frames, this is a serious problem. Thus, it is quite normal that multicast video applications fail completely when they operate on a Wi-Fi network although they work fine on a wired network.

http://dx.doi.org/10.5772/intechopen.71784

1.3. IPTV

Internet protocol television (IPTV) is a system that is used to deliver the Internet television services across the Internet protocol (IP) infrastructure. IPTV integrates voice, video, and data delivery into a single IP framework, to enable interactive broadcasting services to the subscribers. It promises significant advantages for both service providers and clients.

In the past few years, IPTV has gained an unremarkable growth rate. IPTV services are becoming very popular among telecommunication companies, most of all because they can provide television programs anytime and anywhere. IPTV can support broadcast and unicast services, such as broadcasting of TV channels, Video on demand (VoD) or time-shifted television. Unlike conventional broadcasting systems, IPTV broadcasts will not be restricted by the limited number of channels in the broadcast/radio spectrum.

Besides, IPTV will provide its clients with the opportunity to access and interact with a wide variety of high quality on demand video content over the Internet. Since IPTV is considered as a real-time broadcast service over the Internet, the success of the IPTV service depends on the quality of experience (QoE) perceived by the end clients.

The IPTV system has to support a wide range of transmission channel characteristics (i.e., varying packet loss rates and transmission delays) between the content providers and the end clients. The characteristics of video traffic as well as the high-quality requirements of IPTV broadcast impose strict requirements on transmission delay. IPTV framework has to provide mechanisms to satisfy the stringent delay, jitter, and bandwidth requirements over lossy transmission channels.

The architecture of IPTV is composed of five main parts as depicted in **Figure 1**: (i) head network, (ii) core network, (iii) distribution network, (iv) access network, and (v) customer network.

Transport network includes the core network, distribution network, and access network. IPTV service provider transports the IPTV streams to client end by using all entities of the transport

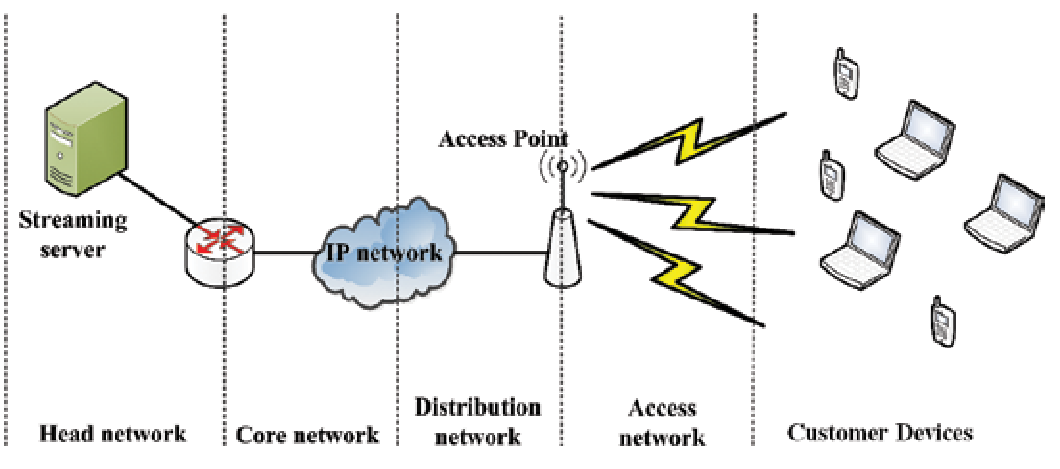

Figure 1. IPTV architecture.

network. Among all these entities, major bandwidth restricting entity in transport network is the access point (AP) because of limited buffer size of queues. To avoid bandwidth restriction, wired access links such as cable and digital subscriber line (DSL) are preferred by IPTV service providers.

Using WLAN for IPTV distribution improves the mobility and flexibility of IPTV network deployment; however, it has its challenges as described in previous sections and it can impact the clients' quality of experience when watching an IPTV program.

Error correction schemes such as forward error correction (FEC) and/or retransmissions can be used to improve reliability of an IPTV program. Both of them are explained in more detail below.

2. Packet loss recovery

Considering the earlier mentioned strict quality requirements of video, strong error recovery techniques are required to protect the content that is delivered to end clients and recover from the occasional packet losses. As it can be seen from the previous examples, there are two possibilities either to prevent packet loss or, alternatively, when packet loss does happen, reduce the noticeable effects of packet loss or restore the missing packets.

In other words, one wants to provide a mechanism to recover from packet loss or to have an error resiliency mechanism that reduces errors due to packet loss. Numerous techniques for providing error resiliency against packet loss can be divided in two categories:

- Techniques that provide recovery for packet loss.
- Techniques that try to reduce the impact or effect of the packet loss.

Two different techniques aimed to recover packet losses during the delivery of IPTV services are FEC and retransmission. The FEC approach operates by adding redundant information to the data at the server, while the retransmission approach recovers from packet losses by requesting retransmission from the server.

2.1. Forward error correction (FEC)

An FEC-based error recovery protocol uses redundant information to allow the client to correct packet losses. With this redundant information, the clients can recover from packet losses nearby the client. The amount of data that can be reconstructed is determined by the amount of loss and on the amount of redundant data. As an FEC-based recovery mechanism is suitable in networks that only allow unidirectional traffic (e.g., satellite broadcast) or environments where the latency from the client to the server is relatively high (e.g., cellular networks), it does not require any feedback from the server to the client. It can be applied from the physical layer up to the application layer according to the OSI reference model. For IPTV streaming applications, FEC can be offered at network layer, transport layer, or application layer FEC. A disadvantage of a FEC protection mechanism is that the applied FEC protection scheme might be insufficient (too weak) for certain clients, while at the same time can be superfluous for some other clients, thus being a waste of bandwidth.

2.2. Retransmissions

Packet retransmission is the technique of retransmitting packets that are considered lost. A packet retransmission requires communication between the client and server, as the client needs to explicitly or implicitly request the packet retransmissions from server. The transmission control protocol (TCP) is a well-known protocol using packet retransmission. Bi-directional communication is required at packet retransmission mechanism in order to allow the client to indicate packet loss and to identify which packet needs to be retransmitted. It is typically achieved by applying sequence numbers. To indicate the loss of packets, two types of messages can be used:

- Acknowledgment (ACK) messages can be used to implicitly indicate the loss of a packet by acknowledging the reception of one packet or a sequence of packets. As the missing packet will never be acknowledged, these messages can then be used to implicitly determine packet loss (e.g., TCP uses ACK messages).
- Negative acknowledgment (NACK) messages are used to explicitly indicate that one or more packets were not received. As feedback from client to server is kept to a minimum, NACK messages are used in constrained networks with many clients per server.

A packet retransmission mechanism is adaptive to variable network conditions. When there is no loss in the network, there will be no packet retransmission, hence no additional required bandwidth. Because packet retransmission mechanisms introduce delay, they are only suitable for applications with nonstrict delay requirements.

Using packet retransmissions for (live) IPTV services has also the drawbacks as IPTV streams are not likely to get adapted to congestion. Thus, if congestion occurs during transport, the packet retransmissions may contribute to the network congestion if the packet retransmission is enabled, thus further lowering bandwidth available for IPTV stream delivery.

2.3. Network coding

In the past few years, network coding (NC) has become popular in both wired and wireless networks as a mechanism for increasing the network throughput. Proposed by Ahlswede et al. in [2], it is a technique for increasing the throughput of the network by encoding multiple packets into the same packet, either from the same or from different streams.

Using the NC, the broadcast nature of the wireless medium is exploited. In the wireless physical medium, all data transmitted by the source is in fact inherently broadcasted to all destinations, even when it is not intended. For no extra cost, other receivers will receive packets not addressed to them. In the case where a packet is lost at the receiver, other receivers may receive it successfully. Keeping those packets by receivers for some time enables NC to reduce the number of packets to be transmitted.

NC reduces the number of transmissions by replacing the conventional store-and-forward paradigm in relaying scenarios with a process-and-forward approach. Instead of forwarding packets in the same form as they are received, the intermediate nodes combine the received outgoing packets using an algebraic function such as XOR. All nodes store all the received packets for possible packet decoding purposes. As a result, a typical NC mechanism is composed of two main operations [6, 10]:

- Coding outgoing packets on intermediate nodes.
- Decoding incoming packets upon their reception on nodes.

The main concept of NC is presented through examples and is also sketched in **Figure 2**. Let us consider the situation where destination node D_1 can only hear transmissions from relay node R and source S1. Similarly, destination D_2 can only hear transmissions from source S_2 and relay node R. S1 has a packet p_1 for D_2 and S_2 has a packet p_2 for D_1. S_1 and S_2 send their packets separately to R, using two transmissions. In the first example (i.e., **Figure 2** (without NC)), the relay node R cannot transmit two packets at the same time. Thus, again two transmissions are used to transmit packets p_1 and p_2.

Referring to the **Figure 2** (with NC), it shows the case of the same network but using NC. The relay node R has enough processing power to code two packets, that is, the R node performs a coding operation, for example, linear operation over the p_1 and p_2; thus creating a coded packet of the same size as largest of p1 and p2. Node R can now by sending only one packet forward both p_1 and p_2 to its destination nodes. D_1 can decode the coded packet as it is familiar with the content of p_1 and similarly D_2 can also decode coded packet as it knows the content of p_2. Instead of using four transmissions, the broadcast nature of the media has been embraced and only three transmissions were needed.

The slight overhead that can be seen here is the buffer space requirement as well as additional processing in encoding and decoding the packet. With the advancement in solid states like high speed processors, and high speed and large capacity memories, these overheads are well taken care of. The bandwidth is limited especially in wireless networks. Using NC, we can efficiently utilize the bandwidth of the network.

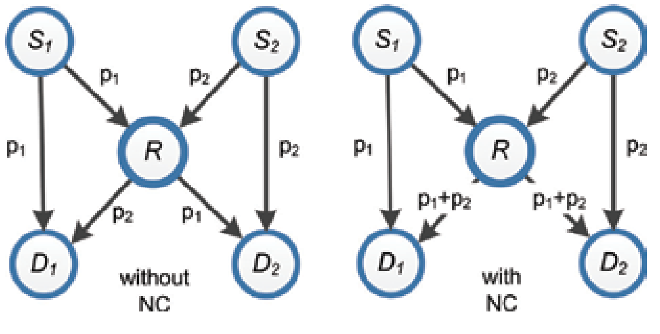

Figure 2. Network coding example.

3. Network coding-assisted retransmission scheme

3.1. Concept

Consider the system setup illustrated in **Figure 3**. System is intended for transferring multiple video streams to multiple clients. In depicted example, streaming server is streaming two channels (streams *A* and *B*) to *Client* 1 and *Client* 2, respectively. Packets that are not successfully delivered in the first transmission are retransmitted. Clients signal the server on the not received packets using signalization, for example, in our case with signalization.

Figure 4 shows similar system. The difference is in the retransmission concept. In the first system, every packet that has not been received is retransmitted individually (e.g., five retransmissions). The second system codes multiple packets together and generally uses fewer transmissions (e.g., three).

For the system depicted in **Figure 4**, the design parameters were selected as follows:

- Only packets that require retransmission are transmitted.
- Packets are transmitted as late as possible, that is, after the retransmission timeout (RTO) or opportunistically.

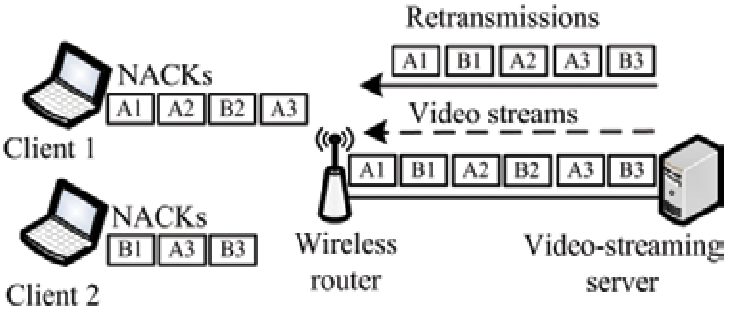

Figure 3. Retransmissions using traditional retransmission scheme.

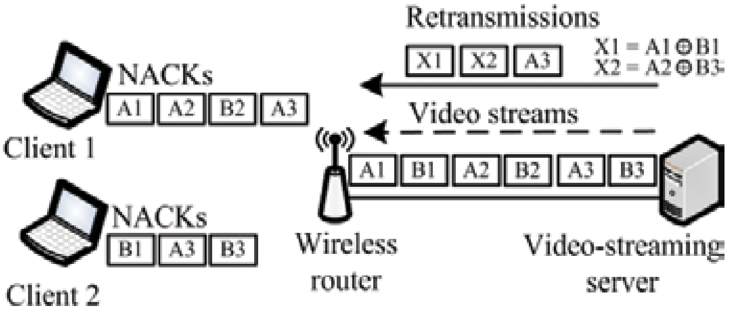

Figure 4. Retransmissions using NC-assisted retransmission scheme.

- All coded packets must be decodable on all the clients.
- Packets not received by any of the clients are sent out as they are, that is, not coded.

The streaming server keeps track of packets received for each of the clients. Information about unsuccessful received packet is provided through the negative acknowledgement (NACK) packet. Current status of the received packets at individual clients is presented in a transition table (e.g., **Table 1(a)**). *M* rows in the table correspond to packets that have not been received by at least one of the clients, and *N* columns correspond to the clients. Packets that have been received by all clients are not presented in the table. Packets with the lowest index have been present in the table for the longest period. There are three different states that indicate packet status:

- *State* 1: packet successfully received by the client.
- *State* 0: packet intended for the corresponding client not received means that packet needs to be retransmitted.
- *State* 2: packet not intended for the corresponding client not received means that packet does not need to be retransmitted. This state is required in the coding procedure as we explained below.

Packet status is represented in the transition table for the time needed to receive feedback from all the clients as explained in next section. Transmission and coding processes are based on the coding table (e.g., **Table 1(b)**) that is obtained from transition table and where only packets that require retransmissions are represented. As soon as packet status is noted in the coding table, the packet can be opportunistically transmitted. In coding table, two states are used to describe the packet status of the clients. Namely, not received packet is indicated by state 0, while packet received on the corresponding client is indicated by state 1. However, in the coding table, we are not presenting the packets that do not need to be retransmitted as they were not transcript from transition to coding table.

NC algorithm for making decisions on which packets to code together is presented in Algorithm 1. The algorithm is called after native packet is broadcasted. First, it checks if the first

(a) Transition table

	C_1	C_2
A_1	0	1
B_1	1	0
A_2	0	1
B_2	2	1
A_3	0	2
B_3	1	0

(b) Coding table

	C_1	C_2
A_1	0	1
B_1	1	0
A_2	0	1
A_3	0	1
B_3	1	0

Table 1. Transition and coding tables.

packet has exceeded *RTO*. Second, algorithm inspects the statuses of the first packet on the clients using the coding table (that is first row) and decides if it is codable. Packet is considered codable, if it has been received at least by one of the clients. Otherwise, it is transmitted as is. Last, the algorithm tries to code the packet with as many packets as possible (even the ones that have not reached the *RTO*) but prioritizes packets that have been waiting for a longer period.

Definition 1. Algorithm considers two packets codable, if coded packet can be decoded by all the clients.

In practice, this means that all sums over the corresponding packet columns from the coding table are not less than the number of coded packets $N_{CODED} - 1$. As shown in [5], client can decode coded packet if it has previously received at least $N_{CODED} - 1$ native packets coded in the encoded packet.

Algorithm 1

```
if (P1 exceed RTO)
    if (P1 is codable)
        k = 1;
        coded_packets [k] = &P1;
        temp_coding = coding_table [first_row];
        coding = temp_coding [];
        for (m = 2; m <= M; m++)
            codable = 1;
            for (n = 1; n <= N; n++)
                temp_coding[n] = temp_coding[n] + coding table [m, n];
                if (temp_coding [n] < k)
                    codable = 0;
                    temp_coding = coding;
                    break;
                end if
            end for
            if (codable)
                k++;
                coding = temp_coding;
```

```
                coded_packets [k] = &P_m
            end if
        if (k == N)
            break;
        end if
    end for
end if
if (k > 1)
    encode packets from coded_packets
    sent encoded_packet;
else
    sent P_1 uncoded;
end if
    update coding_table;
end if
```

Example 1: Let us use the example from **Table 1 b** to explain the algorithm in practice. When algorithm is initiated, it first checks if the *RTO* for the first packet A_1 has exceeded. Then, it checks for the packet A_1 if it is codable. Since these conditions are fulfilled, algorithm examines if the second packet B_1 is codable with A_1. Referencing to the *Definition* 1, packets are codable, that is, sum of their corresponding columns is not less than k. In this example, only two packets can be coded together, hence the algorithm stops the coding procedure for packet A_1. Packets A_1 and B_1 are coded together and sent out using one transmission. If there were more clients, the algorithm would look for other coding opportunities trying to code more packets together. Further, when *RTO* is reached for packet A_2 algorithm, tries to code it with the next packet in the coding table A_3. Packets do not meet requirement from *Definition* 1, hence algorithm matches A_2 against the next packet, that is, B_3. In this case, packets are codable, thus, sent out in one transmission as together coded packet. Since there are no more packets left to code it to, lastly, packet A_3 is sent out as is. However, in such cases, where there is only one packet left in the coding table, is rarely encountered in practice. Simultaneously with packet transmissions, there are requests for retransmissions for new packets received. Thus, new packets are coming in the coding table to which A_3 would be matched for new coding opportunities.

In a given *Example 1* using proposed approach, five retransmitted packets are transmitted using only three transmissions. Even higher gains can be obtained in cases with more clients.

http://dx.doi.org/10.5772/intechopen.71784

3.2. Implementation details

This section describes implementation details of the mechanisms required on the server and client.

3.2.1. Client side

On the client, *signalization*, *overhearing,* and *packet decoding* are the main processes and are described in greater detail in the following:

- *Signalization*: *NACK* packets are the only stand-alone packets used for signalization purposes and inform server in which packets have not been received on the clients. Two types of *NACK* packets are foreseen: hard *NACK* and soft *NACK*. Hard *NACK* is a packet sent by the client that has missed a packet from a stream currently in use (i.e., marked with 0 in the transition table). Soft *NACK* is a packet sent by the client that has not received packet for stream that is not intended for it (i.e., marked with 2 in the transition table). Soft *NACKs* only inform the server on the status of the received packets, though the packet not received is not required by the client. *NACK* packets are sent to the server using unicast mechanism, as it is a reliable mechanism and packets are unlikely to get lost.
- *Overhearing*: all clients listen to all the transmissions, even the ones not intended for them and store all packets in the packet pool, for decoding purposes.
- *Decoding*: with every received packet, the client checks if it is native or coded packet. In the case when a coded packet is received, the process checks the packet pool. If enough packets have been received (and stored in the packet pool), the client has enough information and it decodes the coded packet. Using previously stored packets and the *XOR* operation against coded packet, a native packet, previously not received is obtained [6].

3.2.2. Server side

Broadcasting, *NACK handling,* and *coding* processes take place on the server side, namely:

- *Broadcasting*: all packet transmissions from the server to clients are carried out using broadcast.
- *NACK handling*: for every *NACK* packet that server receives updates in transition table are made. Packet status is recorder in transition table only for a short time in order to gather *NACKs* for the packet from all the clients.
- *Coding*: after every broadcast of native packet server searches for coding opportunities in the coding table. If an opportunity is found, first packets are coded together with as many packets as possible sent out in one transmission to all clients. Otherwise, it is not coded than sent out as is.

3.2.3. Overhead estimation

NC-assisted retransmission scheme brings an additional overhead into the network in terms of additional packets and headers. They are depicted in **Figure 5**. All packets have additional

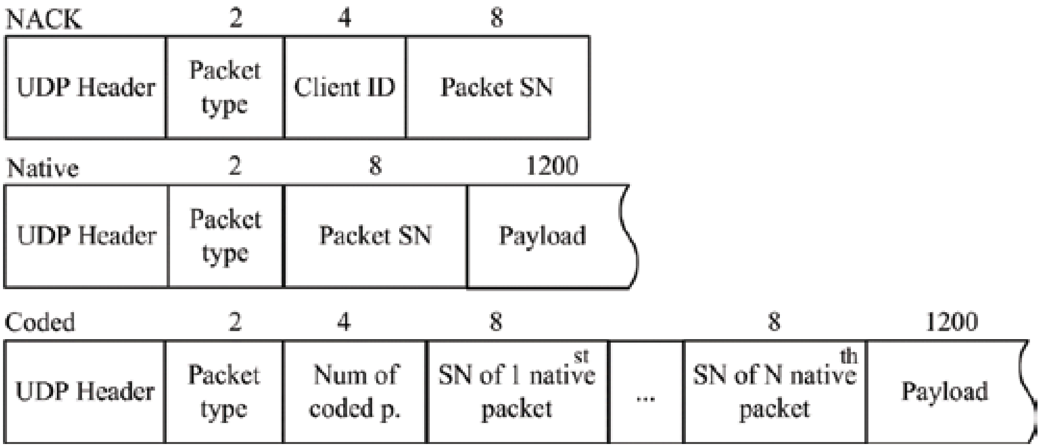

Figure 5. Packet headers.

headers such as *packet type* field indicating that a packet is signalization (*NACK*), native, or coded packet (2 bytes). Native packet has additional *sequence number* (8 bytes) of a packet used for identification. In the case of a coded packet, *sequence number* for each of the encoded packet is given. Also, coded packet carries information on how many packets are coded in the encoded packet (4 bytes). Native packet has *Payload* (1200 bytes) field where video content is placed. Coded packet payload is the same as the native packet and it is composed of *XOR-ed* payloads of every encoded native packet.

Signalization (*NACK*) packet contains the network coding unique information such as: *Client ID* (4 bytes) field, which identifies the sender of the client.

4. Practical evaluation

4.1. Wi-Fi test bed

In order to evaluate the proposed scheme, we deployed IEEE 802.11 network test bed consisted of streaming server, access point, and multiple Wi-Fi clients. Server and clients were laptops with different hardware, for example, different wireless interfaces. For the access point, we used wireless router Linksys WRT54GL. Server was connected to the wireless router via Ethernet interface while clients were connected via WLAN interface. In order to perform tests, we used a third party firmware (i.e., DD-WRT) in our wireless router, because the original firmware was not capable of transmitting a multicast channel through it. Wireless router configuration was fixed during the experiments. Packets size was constant (i.e., 1210 bytes). Inter-arrival time between packets was also constant (i.e., 6 ms), resembling video-streaming application. Results were presented for the shortest time intervals to observe steady state conditions. Two client setups, mobile and static were used for the evaluation and are presented in the next sections.

4.1.1. Mobile client setup

In this scenario, experiment was set up with server and six clients, where each client position has changed once, during the experiment.

Server streamed one UDP 608 MB stream. If it streamed one packet every 6 ms (inter-arrival time between packets) that would be approx. 166 packets per second. The size of a single packet is 1200 bytes. Thus, the data rate was 1.525 Mbps.

The experiment started with all clients placed in the room with the router. Every 7 min, one client at a time was relocated to the nearby room. By moving clients to the nearby room, we wanted to observe how delivery probabilities were changing during the time. We assumed that delivery probabilities would decrease, due to wall obstacles between the server and the clients. Further, obstacles had caused packet loss, which we supposed would affect delivery probability of single clients in the nearby room.

Next, with decreased delivery probabilities, we expect more coding opportunities and higher bandwidth reduction, which we will discuss later in the result section.

4.1.2. Static client setup

In the second scenario, where clients were static and located in one room, four experiments were carried out, each with different number of clients (4, 6, 8, and 10). Server streamed one UDP stream in total of 122 MB and 1.525 Mbps data rate. Value of data rate is calculated as in for mobile client setup. Using this architecture setup, we wanted to observe if the number of clients will affect improvements in the bandwidth.

We expect to reduce more of the bandwidth, because with more clients we assume more coding opportunities, which means fewer transmissions and therefore better bandwidth reductions.

4.2. Evaluation metrics

Several performance metrics were used to evaluate the performance of the proposed scheme. All the presented results were observed over the sample period T_S (i.e., 20 s).

N_N is the number of native packets sent. N_R notes the number of retransmitted native packets. N_{RNC} is the number of transmissions performed to retransmit N_R packets, where $N_R \geq N_{RNC}$ if network coding is used. N_{Di} presents number of successfully received packets at i-th client, while N_C presents the number of clients.

Bandwidth reduction B_R is calculated as the proportion of difference of the N_R and N_{RNC} packets to the sum of N_N and N_R packets. We use B_R to show the share of bandwidth we reduced by using proposed approach.

$$B_R = \frac{N_R - N_{RNC}}{N_N + N_R} \cdot 100\% \tag{1}$$

Client delivery probability DP_i is defined for individual clients as:

$$DP_i = \frac{N_{D_i}}{N_{N_i}} \cdot 100\% \tag{2}$$

Delivery probability DP is an average of client delivery probabilities:

$$DP = \frac{1}{N_C} \sum_{i=1}^{N_C} DP_i \tag{3}$$

Client gain G_i is defined for ith client as:

$$G_i = 1 - \frac{N_{Ri}}{N_{RNC_i}} \cdot 100\% \tag{4}$$

Gain G is average client gains:

$$G = \frac{1}{N_C} \sum_{i=1}^{N_C} G_i \tag{5}$$

Coding table size (CTS) refers to the number of packet-statuses on the clients noted in the table (note that only packet that requires retransmission is present there). Size is sampled periodically, after broadcasting native packet.

C_{AVG} is the average number of native packets sent in one retransmission (average number of encoded packets).

4.3. Experimental results

Results are presented individually for two client's setups that we explained above, that is, mobile and static. Mobile client's results give more detail results of described performance metrics above, while static client's results just discuss the number of clients depended of the gain.

4.3.1. Mobile client results

Bandwidth reduction B_R of proposed approach is presented in **Figure 6**. Straight lines in the figure indicate transition periods at which clients were relocated and divide figure in equal rectangles. For example, in the second rectangle from the left, five clients were located in the router room and one was in the room nearby. From this figure, we can observe that B_R increases over the time. The results show that in the given testbed, we have been able to save 6% of total consumed bandwidth per average. The value in observed time interval ranged between 4 and 8%.

DP is shown in **Figure 7** and we can see that it is over time lowering. This is expected because we were moving clients away from the router over the time.

http://dx.doi.org/10.5772/intechopen.71784

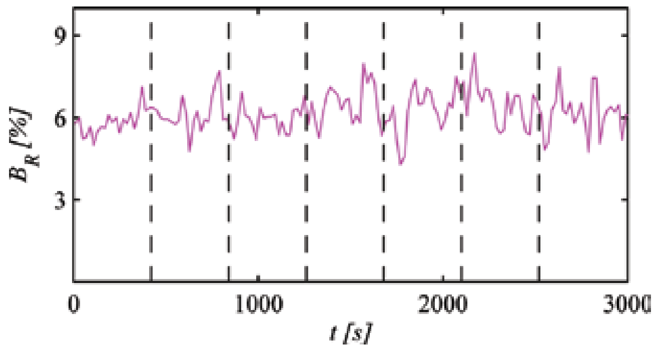

Figure 6. Bandwidth reduction B_R [%] in dependency of time t [s].

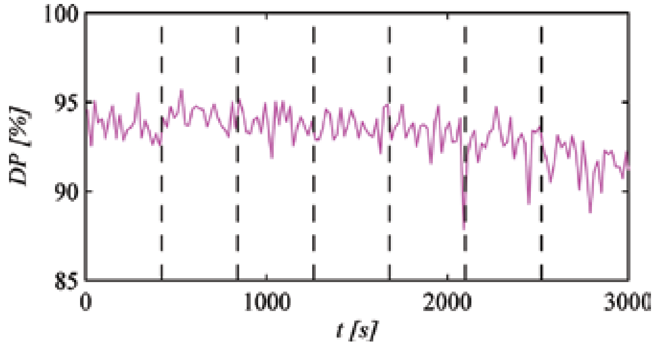

Figure 7. Delivery probability *DP* [%] in dependency of time *t* [s].

In the following, we show relations between average numbers of coded packets C_{AVG}, the number of transmissions performed to retransmit N_R packets, and gain *G* depicted in **Figure 8**. If C_{AVG} falls, N_{RNC} increases, and *G* decreases. For example, let us examine graphs in the time interval between 2100 and 2180s where very apparent case of described situation is shown. If more native packets are coded together, fewer packets need to be transmitted, thus *G* is increased. For example, for t = 2111 s, C_{AVG} = 1.831, N_{RNC} = 350, and G = 45%. Contrarily for t = 2130 s, C_{AVG} = 1.23, N_{RNC} = 1278, and G = 19%.

In the next step, we investigate *DP* versus the *G* as depicted in **Figure 9**. *DP* values range from 80 to 96%, while the corresponding values for *G* are from 18 to 46%. The majority of points from graph are located in the area with high *DP* and high *G*. There are several points that are far away from the cluster.

4.3.2. Static clients results

We investigate how the proposed scheme performs with different number of clients (i.e., 4, 6, 8, and 10). Results are shown in **Figure 10** for B_R. Observed interval is between 0th and 250th. Results show that highest B_R values are observed when the number of clients is the highest

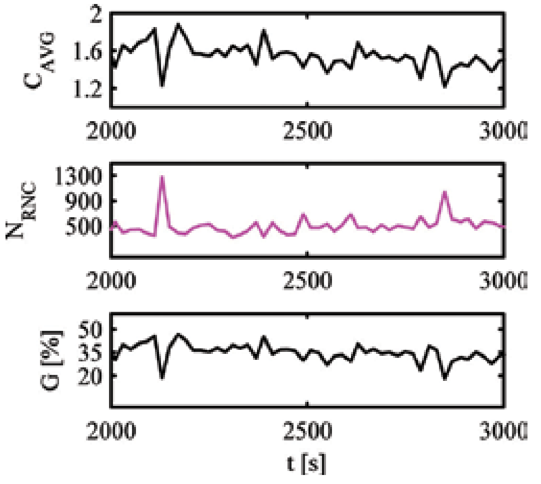

Figure 8. Average number of coded packets CAVG, coding table size CTS, and gain *G* [%] over time.

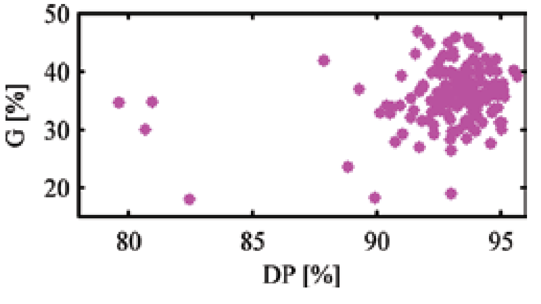

Figure 9. Gain in *G* [%] dependency of delivery probability *DP* [%].

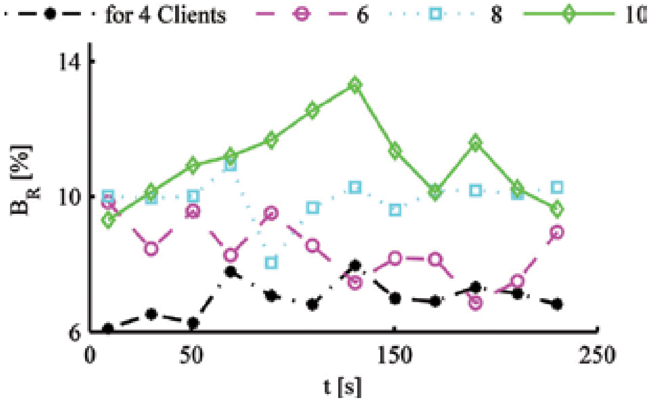

Figure 10. Bandwidth reduction B_R [%] over time for different number of clients.

http://dx.doi.org/10.5772/intechopen.71784

(e.g., average B_R values for 4, 6, 8 and 10 clients are: 7, 9, 10, and 13%, respectively). When the number of clients increases, B_R improves at the same time, that is, when the number of clients is higher, there are more packets that get lost, hence there are more entries in the coding table CTS and coding algorithm is able to find more coding opportunities.

5. Conclusion

In this chapter, the approach for improving bandwidth in the last mile wireless broadcast network using network coding was investigated theoretically and practically. We have evaluated the performance of our approach for video-streaming services, such as IPTV, by designing and implementing a specific system solution, for a lossy wireless scenario. Our proposed schemes combine different lost packets from different clients in such a way that all clients are able to recover their lost packets with one transmission by the server. In particular, network coding is used to code and decode packets, as the new way of transmitting packets through the network. Instead of just store-and-forward technique, the server is now able to code packets in a more clever way, and transmit them over the broadcast network.

The implementation of such architecture, in a real scenario, was shown to be feasible. We tested the proposed approach with two different scenarios. It was found that the proposed approach in both cases decreases the number of transmitted packets. As compared to traditional unicast retransmission approach, where every packet is retransmitted separately, the reduction in bandwidth using the proposed network coding assisted retransmission scheme was up to 15%. This is a significant result and the developed scheme should be of great interest to be implemented in throughput demanding systems. In addition, results show that proposed approach saves bandwidth in various situations where users have different terminal equipment and different link quality. We observed the highest gains with high number of clients (i.e., 10 clients), which is also the case in the envisioned application where even higher number of different concurrent streams is anticipated.

Author details

Aleš Švigelj* and Melisa Junuzović

*Address all correspondence to: ales.svigelj@ijs.si

Jožef Stefan Institute, Ljubljana, Slovenia

References

[1] Junuzović M, Alič K, Švigelj A. Wireless broadcast transmission scheme for reliable video-streaming service. WSEAS Transactions on Communications. 2015;**14**:256-266

[2] Ahlswede R, Cai N, Li S-YR, Yeung RW. Network information flow. IEEE Trans. on Information Theory. 2000;**46**(4):1204-1216. DOI: 10.1109/18.850663

[3] Ferreira D, Costa RA, Barros J. Real-time network coding for live streaming in hyper-dense WiFi spaces. IEEE Selected Areas in Communications. 2014;**32**(4):773-781. DOI: 10.1109/JSAC.2014.140409

[4] Seferoglu H, Markopoulou A. Video-aware opportunistic network coding over wireless networks. IEEE Journal on Selected Areas In Communications. 2009;**27**(5):713-728. DOI: 10.1109/JSAC.2009.090612

[5] Katti S, Rahul H, Hu W, Katabi D, Médard M, Crowcroft J. XORs in the air: Practical wireless network coding. IEEE/ACM Transactions on Networking. 2008;**16**(3):497-510. DOI: 10.1109/TNET.2008.923722

[6] Alič K, Pertovt E, Švigelj A. Bearing-opportunistic network coding. International Journal of Computers, Communications & Control. 2015;**10**(2):154-164. DOI: 10.15837/ijccc.2015.2.457

[7] Nguyen D, Tran T, Nguyen T, Bose B. Wireless broadcast using network coding. IEEE Transactions on Vehicular Technology. 2009;**58**(2):914-925. DOI: 10.1109/TVT.2008.927729

[8] Xiao X, Lu-Ming Y, Wei-Ping W, Shuai Z. A wireless broadcasting retransmission approach based on network coding. In: ICCSC 2008. 4th IEEE International Conference on Circuits and Systems for Communications; 26–28-05-2008. Shanghai, China. Shanghai: IEEE; 2008. pp. 782-786. DOI: 10.1109/ICCSC.2008.171

[9] Fang W, Liu F, Liu Z, Shu L, Nishio S. Reliable broadcast transmission in wireless networks based on network coding. In: IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS); 10–15-04-2011; Shanghai, China. Shanghai, China: IEEE; 2011. pp. 555-559. DOI: 10.1109/INFCOMW.2011.5928875

[10] Alič K, Švigelj A. Self-adaptive practical opportunistic network-coding procedure for static wireless mesh networks. Ad-hoc & Sensor Wireless Networks. 2017;**36**(1–4):87-105

Atmospheric Attenuation of the Terahertz Wireless Networks

Milda Tamosiunaite, Stasys Tamosiunas,
Mindaugas Zilinskas and Gintaras Valusis

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72205

Abstract

The increase of data traffic, a demand for high-speed reliable mobile networks and congested frequency bands raised both technological and regulatory challenges. Therefore, the fifth-generation mobile network (5G) is being developed. Recently, researchers have focused on a very promising terahertz (THz) band (frequencies from 100 GHz to 30 THz), which will allow fast transmission of huge amounts of data. However, transmission distance is limited due to atmospheric attenuation, as THz waves undergo significant absorption by water vapor and oxygen molecules in the atmosphere. Moreover, THz waves are very vulnerable by precipitation. Furthermore, the path of the propagating waves changes due to variations of the atmospheric refractive index. Nevertheless, the THz networks could be perfect candidates for fiber-to-THz bridges in difficult-to-access areas. The aim of this chapter is to present the possibilities and challenges of the THz networks from a point of view of atmospheric attenuation. The results show that simulations of the atmospheric attenuation using real-time data are a powerful tool that should complement technological basis, as it will help to foresee possible failures, extend transmission distance and improve reliability of the THz and other high-frequency broadband wireless networks.

Keywords: broadband wireless communications, terahertz wireless networks, atmospheric attenuation, rain attenuation, refractive index

1. Introduction

Data traffic in wireless communications has been increasing rapidly over recent decades. Edholm's law (resembling the well-known Moore's law for transistors) states that wireless data rates have doubled every 18 months over the last three decades and are quickly approaching the capacity of wired communication systems [1]. Furthermore, the Internet traffic is expected to reach over 130 exabytes per month by 2018 [2]. The speed limits are based on the Shannon theorem [3], which states that, for a given average signal power, the channel bandwidth limits the maximum data rate that can be attained with a sufficiently low error rate.

The most desirable spectrum to satisfy the wireless needs is the frequency band of 1–10 GHz, whereas propagation losses due to atmospheric absorption and precipitation play a secondary role, and in many cases may generally be neglected[1] [4]. However, the band up to 10 GHz is highly congested. Despite the efforts to improve its efficiency, it is difficult to locate a sufficient amount of free bandwidth. Since current networks are no longer able to cope with the forthcoming load (i.e., Internet of Things (IoT), self-driving cars and so on), recent research has focused on the THz technology, which has a lot of potential owing to a large bandwidth capability, highly directive beams, obtained with relatively small antennas, and, consequently, low transmitter power requirement. Although transmission distance will be limited due to severe effects in the atmosphere, the THz networks are perfect candidates for short-distance communications, such as fiber-to-THz bridges in difficult-to-access areas.

The aim of this chapter is to present the possibilities and challenges of the THz networks from a point of view of atmospheric attenuation. The chapter consists of three parts. The first part is dedicated to the review of the possible applications and basic concepts of the THz wireless networks. In the second part, the main mechanisms of the atmospheric attenuation are presented. This part also includes a brief overview of the basic principles of radio system design and recommendations of the International Telecommunication Union (ITU-R). The third part is based on real-time measurements and simulations using meteorological data.

2. Terahertz wireless networks

2.1. The terahertz band

The terahertz (THz) band (also often called the T-rays or „terahertz gap”) consists of electromagnetic waves in the frequency region of 100 GHz–30 THz.[2] One terahertz (1 THz = 10^{12} Hz) corresponds to a vacuum wavelength of 0.3 mm. As terahertz radiation begins at a wavelength of 1 mm and proceeds into shorter wavelengths, it is sometimes called the submillimeter band. The THz band is also called the “terahertz gap,” as it divides the microwaves and infrared waves or, in other words, the electronics and optics (**Figure 1**). Both electronics and optics contributed to the development of the THz technology, as photonics has led the way to the realization of many important THz devices, such as the development of the quantum cascade laser (QCL), while electronics contributed with solid-state electronic devices, such as resonant tunneling diodes (RTD) [5].

[1]However, the application cannot be so straightforward. In order to ensure a reliable connection, the attenuation of the atmosphere should be evaluated starting from 1 GHz or even lower frequency.
[2]100 GHz corresponds to 0.1 THz. Perspectives for telecommunications are expected in the region up to 10 THz.

http://dx.doi.org/10.5772/intechopen.72205

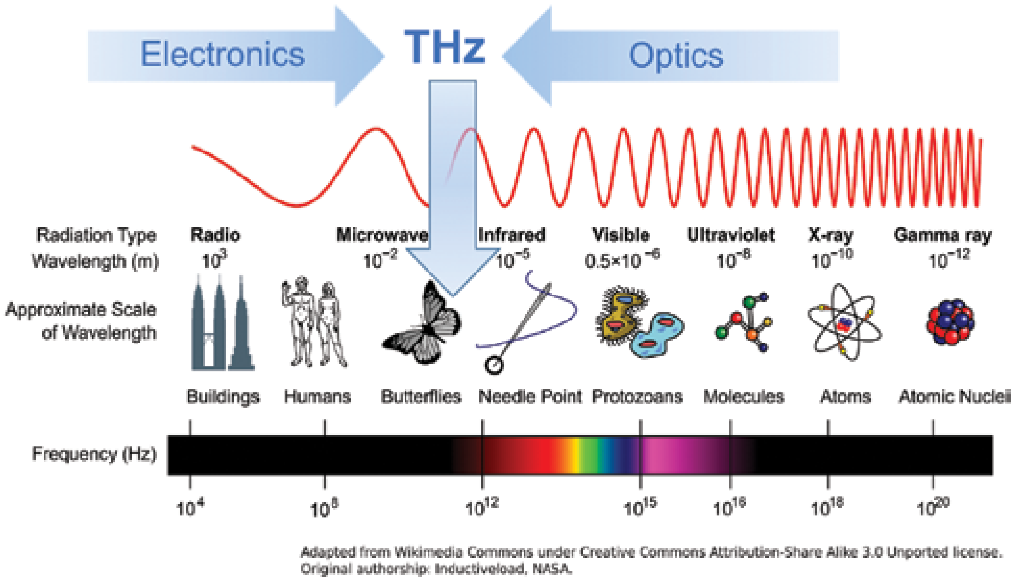

Figure 1. The terahertz (THz) gap. Both electronics and photonics contributed to the development of the THz technology.

The THz band, combined with recent technological innovations, offers wide application possibilities from nondestructive medical imaging, quality control, security issues (i.e., detection of weapons, explosives and narcotic substances, as every material has a distinct signature in its THz spectra [6]), communications and many more. The communication sector can benefit from THz technology in many ways, starting from wireless communications and high-speed data processing, to satellite communications.

2.2. The concept of ultrabroadband THz wireless networks

According to [7], there are several reasons that motivate the use of the THz band for ultrabroadband communication networks. First of all, wireless technologies below 0.1 THz (100 GHz) are not able to support terabit per second (Tbps) links, as available bandwidth limits the achievable data rates (however, compact wireless technologies above 10 THz are not able to support Tbps links). Furthermore, the THz band offers a much larger bandwidth, which ranges from tens of GHz up to several THz, depending on the transmission distance, and it opens the door for the variety of applications that demand ultrahigh data rates. In 2012, a group of researchers have smashed the record for wireless data transmission in the terahertz band. The data rate was 20 times higher than the best commonly used Wi-Fi standard [8]. However, due to technological limitations, the best performance of the THz wireless network is shown under indoor conditions. One of the main challenges is a very high path loss at THz band frequencies, which limits the communication distances[3] [7]. The appropriate frequency band should be determined by the application [9], as the higher the operational frequency,

[3]In addition to all the challenges, the THz Band is not yet regulated. However, the beginning development is accompanied by standardization activities addressing the lower THz range. According to the proposal of the IEEE 802.15.3 group, the frequency range 252–325 GHz is defined as the Thz PHY, and the links using Thz PHY are called THz links.

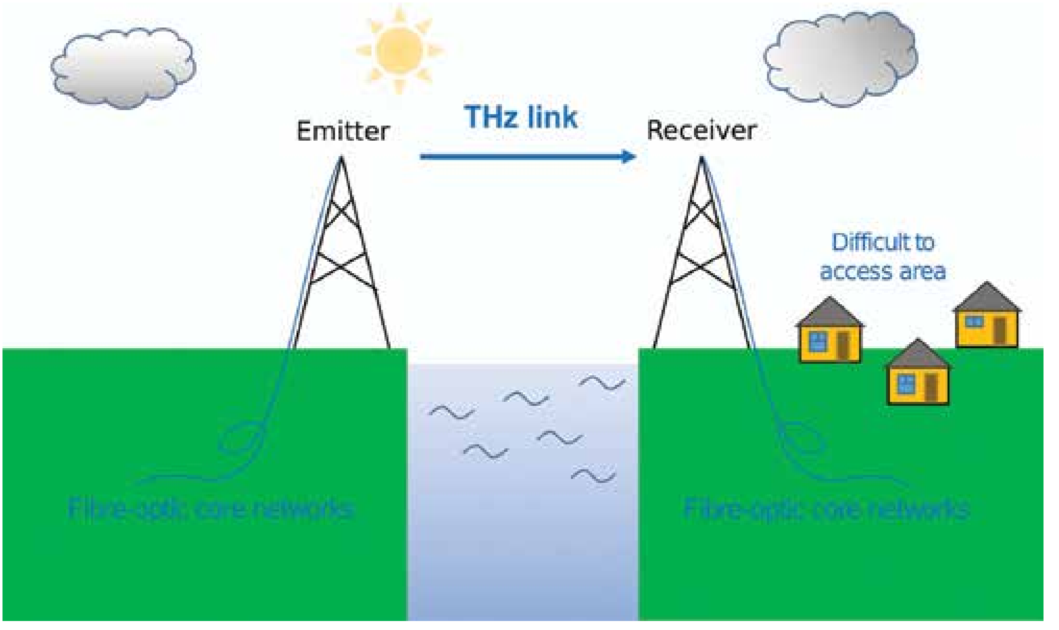

Figure 2. The concept of fiber-to-THz radio bridges for connection between a fiber-optic link and a THz link.

the shorter the distance (~100–150 GHz for long distance (~1–10 km), <350 GHz for medium distance (~100 m–1 km) and <500 GHz for indoor (~10–100 m) communications) which THz wave can travel without crucial distortions. Bearing in mind the distance problem, THz communications could benefit from the well-developed fiber-optic links. The wireless-over-fiber technology was introduced in [10]. In fact, owing to low losses in the optical fiber cables for long-haul transmission, wireless-over-fiber is a very promising technology for the difficult-to-access areas (**Figure 2**). In addition, the basic principles of the 5G technology can be adapted, such as the massive Multiple Input, Multiple Output (MIMO) antennas integrated into base stations [11] and densely deployed small cells [12] that are able to maintain a decent level of signal quality in areas with high density of mobile users [13]. Other solutions are also possible. For example, in [14], a concept of mirror-assisted wireless coverage system is suggested. The smart antennas are utilized with a number of moveable dielectric mirrors that act as reflectors for the THz waves, creating a virtual line of sight.

However, as was concluded in [15], the transmission capacity of a system with a given transmission power depends critically on channel properties like path losses, antenna misalignment and interference due to reflection and scattering.

3. Propagation modeling (atmospheric attenuation)

The design techniques of the new telecommunication links take into account a series of various factors that are comprehensively explained by Freeman [4] and regulated by the International Telecommunication Union (ITU-R). If a beam would propagate in free space without atmosphere, the path would be a straight line. However, the atmosphere is always present. In this chapter, the atmospheric effects on propagation are briefly overviewed in terms of the THz communications. A more detailed description of the atmospheric effects is provided in [16].

3.1. Free-space loss vs. actual atmosphere

A free-space path loss (FSPL) is inevitable attenuation in the line-of-sight path through free space. In general case, the free-space loss rises with the square of the carrier frequency and link distance (i.e., 100 dB for a 10-m THz link at a carrier frequency of 300 GHz [17]). In actual atmosphere, there are additional sources of atmospheric loss that affects the amplitude, phase and polarization of electromagnetic waves. Generally, these effects could be predicted. The prediction procedures consist of few steps, and the meteorological data (rain rate, temperature, humidity, etc.) are required.

3.2. Gaseous absorption

Due to the presence of gas and water molecules in the atmosphere, electromagnetic wave travels more slowly, and part of its energy is scattered or absorbed. Some molecules present in a standard atmosphere are excited by the specific frequencies in the THz band. An excited molecule internally vibrates, and as a result of this vibration, part of the energy of the propagating wave is converted into kinetic energy and is simply lost. The necessary parameters to characterize the different resonances for different molecules are collected, contrasted with real measurements, and compiled into HITRAN database [18]. Considering the standard atmospheric composition (mostly molecular oxygen and nitrogen), the gaseous attenuation, γ_g (in dB/km), up to 0.350 THz (accurate model) and 1 THz (general model) can be evaluated using procedure described in [19]. For distances below 1 m, molecular absorption loss is considered to be almost negligible. When transmission distance exceeds 1 m, molecular resonances become significant.

3.2.1. Water vapor

More than any other region of the electromagnetic spectrum THz region is very vulnerable by water molecules. Being a polar molecule with a nonlinear molecular orientation, water displays a strong absorption line for nearly all of its rotational modes [20]. Therefore, it is worth to mention water vapor, which, being minor constituent of the atmosphere, together with oxygen, is the main contributor to gaseous attenuation. Hence, the gaseous attenuation, γ_g, is a sum of attenuation due to water vapor and oxygen, and the path loss due to atmospheric absorption is estimated as multiplication of gaseous attenuation and path length, d,

$$A_G = \gamma_g d. \tag{1}$$

In the molecular absorption spectrum, there are several regions of relative transparency between the absorption peaks. Those regions are called transmission windows. Under standard atmospheric conditions, transmission windows are present at about <0.300, 0.330–0.370, 0.390–0.440, 0.625–0.725, and 0.780–0.910 THz (**Figure 3**). When the frequency exceeds 1 THz, the radio wave undergoes significant absorption by water vapor and oxygen molecules in the atmosphere [9]. In [21], an accurate terahertz time-domain spectroscopy (THz-TDS) characterization of water vapor from 0.2 to 2 THz was reported. The results agreed with the previously predicted and measured attenuations for the weak water lines, but showed more attenuation for the relatively transparent windows between these lines.

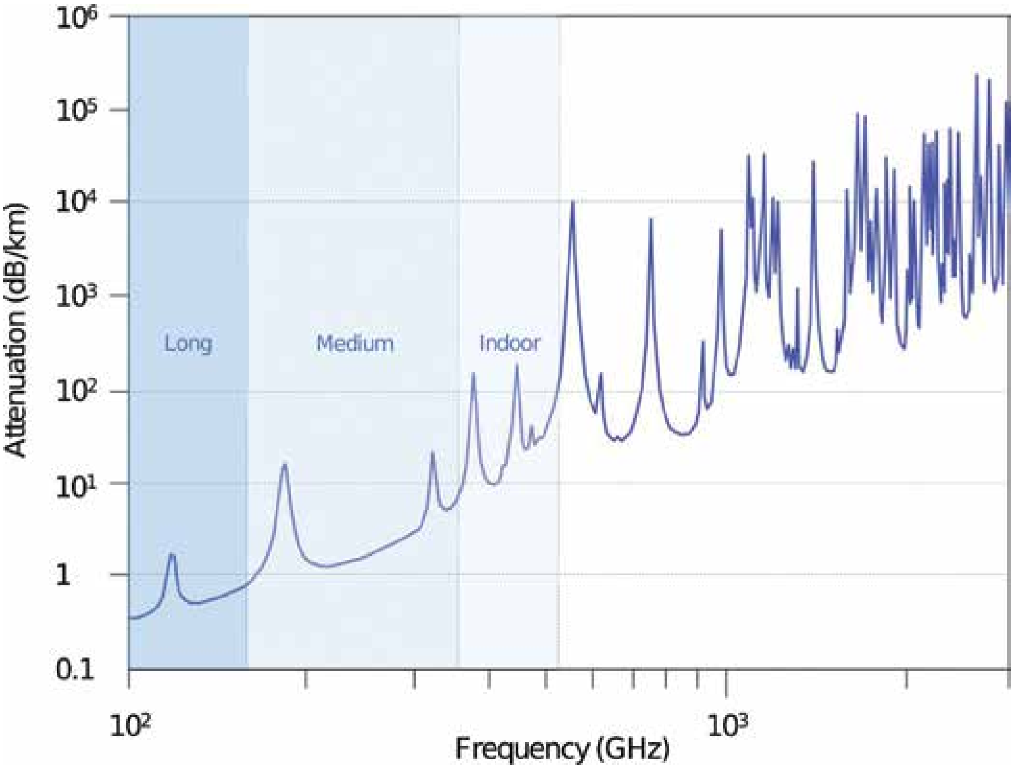

Figure 3. Transmission distance in the THz band is limited by the atmospheric attenuation. Water vapor absorption spectra indicate several transmission windows (the regions of relative transparency between the resonance peaks); when the frequency exceeds 1 THz, the radio wave undergoes significant absorption.

3.3. Precipitation

3.3.1. Rainfall

The effect due to hydrometeors, especially rain, is the most severe. The main mechanisms of the attenuation due to rainfall are absorption and scattering. When an incident electromagnetic wave passes over an object that has dielectric properties different from the surrounding medium, some energy is absorbed (this energy heats the absorbing material) and some is scattered (the smaller the scatterer, the more isotropic it is in direction with respect to the wavelength of the incident energy) [4]. It is a common practice to express the excess attenuation due to rainfall as a function of precipitation rate, which depends on liquid water content and the fall velocity of the raindrops, which in turn depends on drop size distribution (DSD). There are various raindrop size distributions; the most popular is the Laws and Parsons distribution. However, dealing with raindrop size distribution is a complex task, because the shapes of actual raindrops are far from the usually depicted spherical or tear-drop shapes, as larger drops become flattened and eventually break into smaller droplets [22].

One of the most accepted methods of dealing with excess path attenuation due to rainfall is an empirical procedure based on the approximate relation between excess path attenuation A (in dB) and the rain rate R (in mm/h) [23]:

$$A_{dB} = a\,R^{b}, \tag{2}$$

where a and b are functions of frequency. Horizontally polarized waves suffer greater attenuation than vertically polarized waves, so there are different a and b values for vertical and horizontal polarization. The values of a and b can be found in [23].

For frequencies above 100 GHz, the attenuation is considered to be about 10 dB/km in the case of heavy-rain conditions (R = 25 mm/h). However, empirical methods must be applied with caution. Most often, the rain rate of the rainfall is calculated using averaged hourly, daily or even annual data. For path design above 10 GHz, where the path availability better than 99.9% is required, such averaged statistics are not sufficient, as several weeks of light drizzle will affect the overall long-term path availability much less than several good short-lived downpours [4]. In fact, such downpours are cellular in nature and appear in so-called rain cells. To trace such downpours, "one-minute" rain rate values, $R_{(1\,\mathrm{min})}$ (in mm/h), are required. Since data with such accuracy are rarely collected, the "one-minute" rain rate values (integration time is 1 min) can be estimated from various models [16]. However, those models are also based on averaged data; therefore, in territories of varying climate conditions, it is suggested to specify the conversion formula in accordance with the peculiarities of the climate. As the practice shown, the actual "one-minute" rain rate values tend to be higher than the ITU-R suggested values, and it is the main reason for the unexpected telecommunication failures during downpours.

3.3.2. Other precipitation

There are other meteorological phenomena. In example, hail causes only a small attenuation due to rain. The effect of snow depends on temperature, flake size, and water content. Fog and clouds can be treated as very light rains (small droplets or/and ice crystals). In this chapter, other precipitation sources will not be discussed. More information can be found in [4, 16].

3.4. Variations of the atmospheric refractive index

In the atmosphere, the propagating radio beam encounters variations of the atmospheric refractive index. Due to these variations, the ray-path becomes curved, and the fading occurs. The radio refractive index is defined as the ratio of the velocity of propagation of a radio wave in free space to the velocity in a specified medium. At standard atmosphere conditions near the Earth's surface, the radio refractive index has a value of approximately 1.0003. The atmospheric refractive index, n, can be calculated by the following formula:

$$n = 1 + N \times 10^{-6}, \tag{3}$$

where N is the atmospheric refractivity.

According to the procedure described in [4], the atmospheric refractivity is expressed by:

$$N = \frac{77.6}{T}\left(P + 4810\,\frac{e}{T}\right), \tag{4}$$

where P is atmospheric pressure (in hPa), e is water vapor pressure (in hPa) and T is absolute temperature (in K). This expression may be used for all radio frequencies (with the error less than 0.5% for frequencies up to 100 GHz).

The water vapor pressure is given by

$$e = \frac{H\,e_s}{100}, \tag{5}$$

$$e_s = a\exp\left(\frac{bt}{t+c}\right), \tag{6}$$

where H is relative humidity (in %), t is temperature (in °C), e_s is saturation vapor pressure (in hPa) at the temperature t (in °C), and the coefficients a, b, c are $a = 6.1121$, $b = 17.502$, $c = 240.97$ (for water, valid between −20°C and +50°C, with an accuracy of ±0.20%) and $a = 6.1115$, $b = 22.452$, $c = 272.55$ (for ice, valid between −50°C and 0°C, with an accuracy of ±0.20%).

4. Simulations using real-time data

The main challenge of the radio system design is the reliable source of data. When dealing with atmospheric attenuation, it is very important to gather as reliable as possible metrological data. Unfortunately, data of such accuracy are rarely available. In this part, simulations based on real-time data are presented.

4.1. Rain rate integration time problem

As mentioned above, the mostly available meteorological data are not sufficient for the radio system design, as averaged meteorological data suppress the peak values of significant events, such as downpours. Therefore, 1-min integration time should be used if available.

In **Figure 4**, the data of the heavy-rain event are presented. During 6-h lapse, the rain intensity differed from light to heavy rain, and 37 mm of rain was accumulated until midnight (marked as Precip. Accum.). At the same time, the rain intensity was measured (R-value (10 min)). For technical reasons, the measurements were carried out in 10-min intervals. In addition, 1-min rain rate value, $R_{(1\,\mathrm{min})}$, was calculated, using the Moupfouma and Martin method [24]:

$$R_{(1\,\mathrm{min})} = \left(R_{(\tau_{\mathrm{min}})}\right)^{0.987\tau^{0.061}}, \tag{7}$$

where $R_{(\tau_{\mathrm{min}})}$ is the rain rate value (in mm/h) measured in lapse of time τ (in minutes).

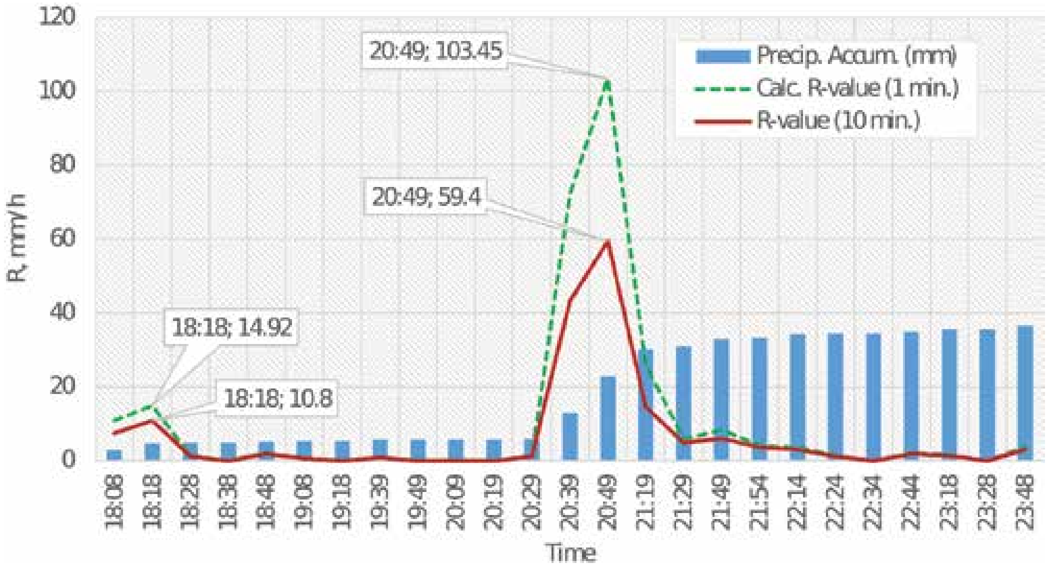

Figure 4. Heavy-rain event. The peak values of the rain rate at the different aggregation times are presented. According to measurements, the maximum R-value was $R = 59.4$ mm/h at 20:49; the maximum calculated 1-min R-value was almost double, R (1 min)=103.45 mm/h. Both of these values are higher than the ITU-R suggested value $R = 35$ mm/h.

According to measurements, the maximum R-value was $R = 59.4$ mm/h at 20:49. However, calculated 1-min R-value was almost double, $R_{(1min)} = 103.45$ mm/h. Both of these values are higher than the ITU-R suggested value $R = 35$ mm/h [25].

In **Figure 5**, calculated values of rain attenuation at 0.1 THz (or 100 GHz; the minima of the THz range) electromagnetic waves are presented. As can be seen, the results are different for 1- and 10-min integration times. One-minute rain rate value reveals the peak attenuation being 32.27 dB/km for horizontally polarized electromagnetic waves and 31.55 dB/km for vertically polarized ones. The values calculated using 10-min integration time are lower (22.11 and 21.02 dB/km, respectively). However, when the rain rate starts to fall, the integration time and polarization become less important, as the results in all cases are nearly similar.

In **Figure 6**, the same rain event is presented. The difference is that the R-values were calculated using more widely accessible averaged rainfall data with integration time of 1 h and 6 h (standard). In these cases, the downpours are suppressed and delayed ($R = 24.6$ mm/h at 21:19 according to hourly data; $R = 6.56$ mm/h at 23:18 according to 6-h average data).

As can be seen in **Figure 7**, the impact of the suppressed and delayed R-values on the attenuation of the propagating 0.1 THz electromagnetic waves is significant and the results are distorted in comparison with the **Figure 5**. The low attenuation values become higher, and the peak values are lower (12.13 dB/km for horizontally polarized waves in comparison with 32.27 dB/km in **Figure 5**). The 6-h (standard) R-values (see inset in **Figure 7**) determine very low and delayed attenuation values.

The same calculations were performed for electromagnetic waves with operating frequency of 0.3 THz (or 300 GHz; in many cases, this frequency is considered as a realistic candidate

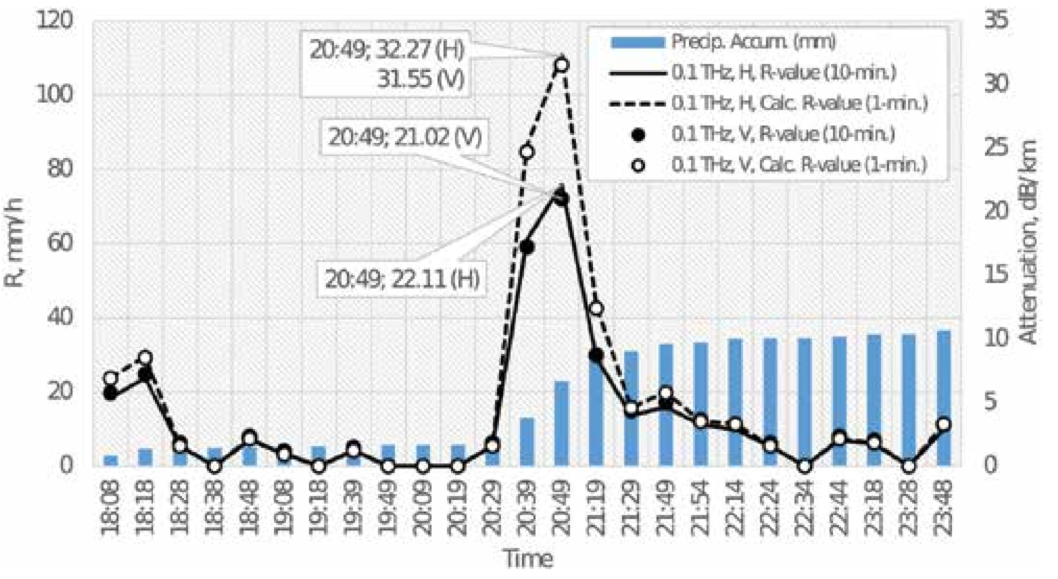

Figure 5. The calculated values of rain attenuation at 0.1 THz using real-time *R*-values. The results differ with integration time. One-minute rain rate value reveals the peak attenuation being 32.27 dB/km for horizontally polarized electromagnetic waves and 31.55 dB/km for vertically polarized ones when the rain rate is 103.45 mm/h.

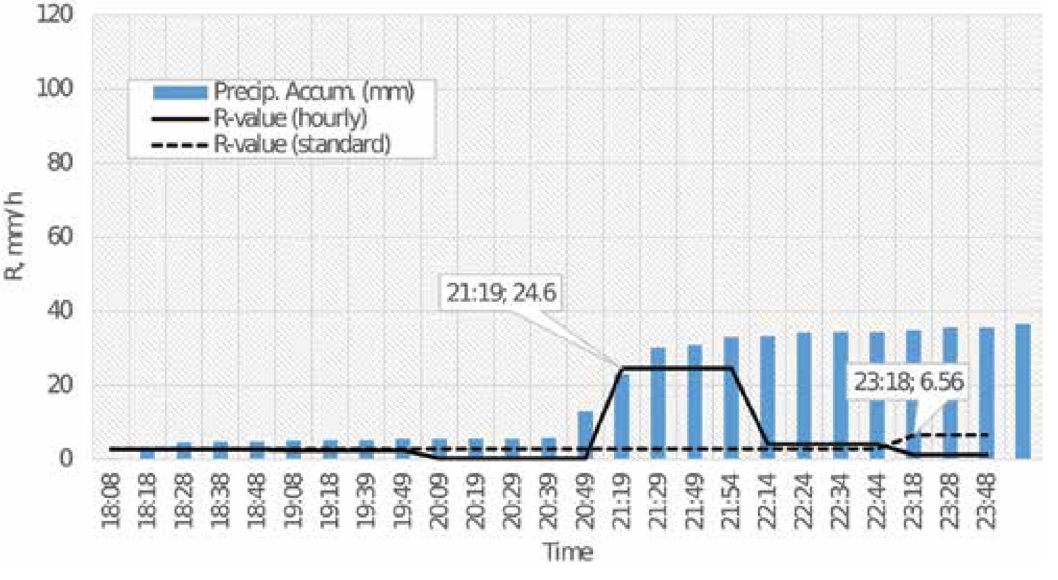

Figure 6. The *R*-values calculated using averaged hourly meteorological data. The downpours are suppressed and delayed in comparison with **Figure 4**.

for the THz communications). The results are presented in **Figure 8**. As can be seen, the tendencies are the same as for 0.1 THz. However, for 0.3 THz, the calculated attenuation values are smaller than for 0.1 THz (according to model, for frequencies higher than 0.1 THz, the rain attenuation slightly decreases) (**Figure 9**).

http://dx.doi.org/10.5772/intechopen.72205

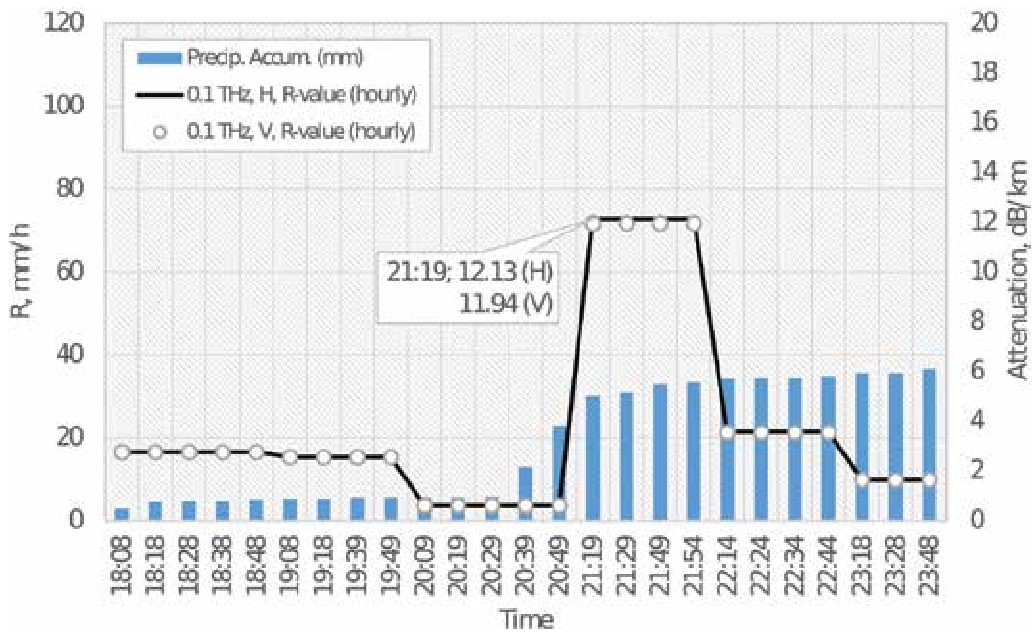

Figure 7. The calculated values of rain attenuation at 0.1 THz using averaged *R*-values. The results are distorted, as the low attenuation values are higher and the peak values are lower.

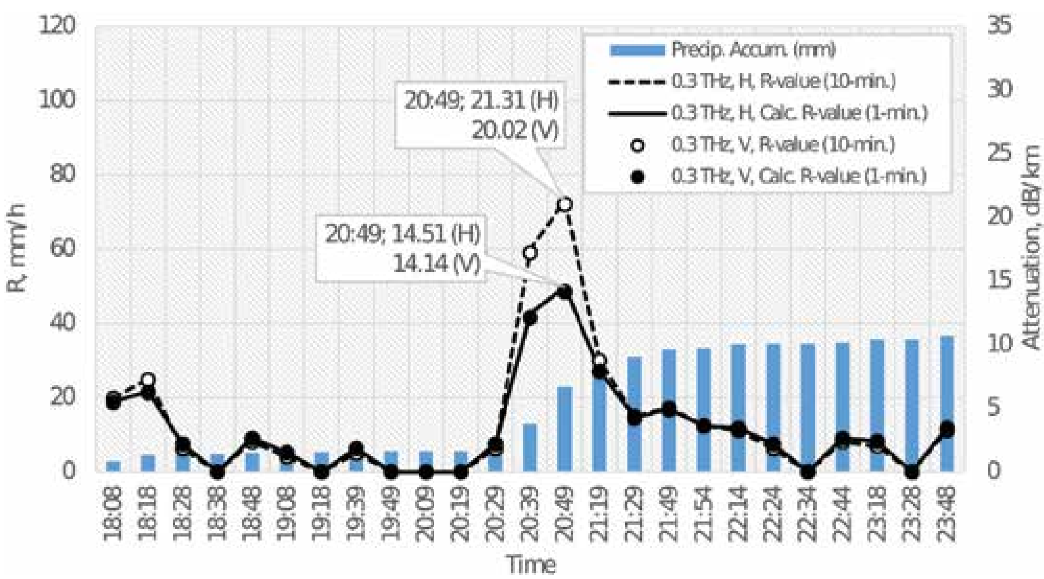

Figure 8. The calculated values of rain attenuation at 0.3 THz. For 0.3 THz, the attenuation values are smaller than for 0.1 THz, but the tendencies are the same.

4.2. Atmospheric refractivity

The atmospheric refractivity, *N*, defines the curvature of the ray path due to which the fading occurs. Since the *N*-value is frequency independent (see formula (4)), the main peculiarities

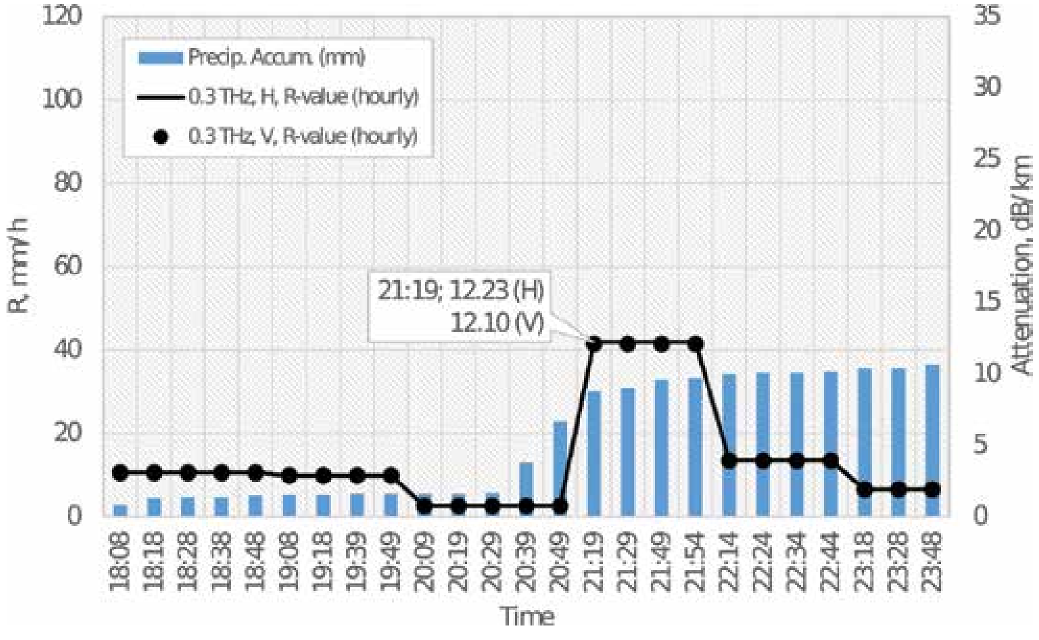

Figure 9. The calculated values of rain attenuation at 0.3 THz using averaged *R*-values. The results are distorted, as the low attenuation values are higher and the peak values are lower.

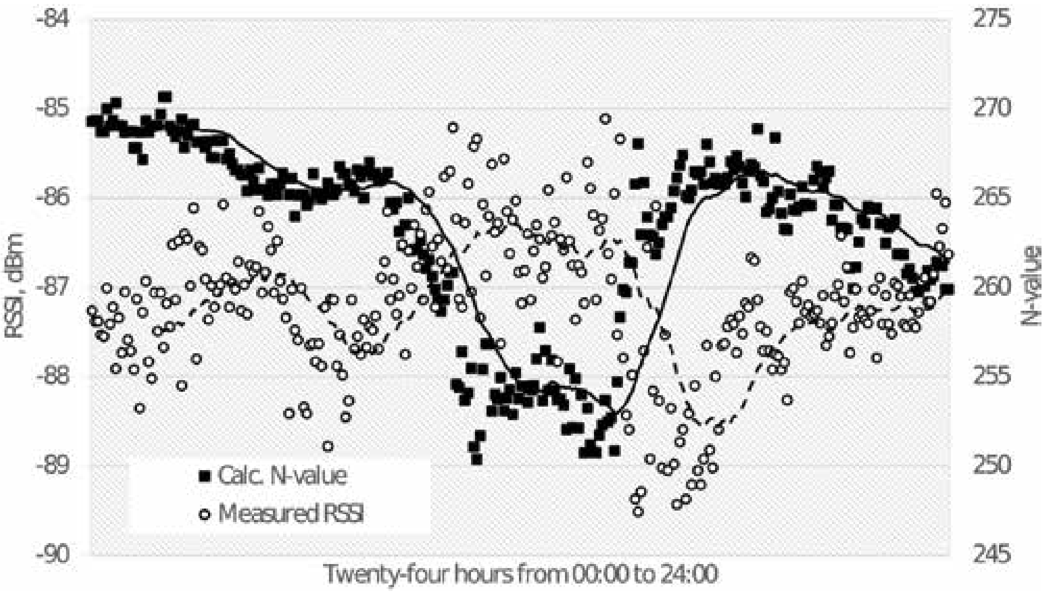

Figure 10. Variations of radio refractivity compared to measured signal strength. Changes of the signal strength occurred at same time as the variations of the radio refractivity. Since the *N*-value is frequency independent (see formula (4)), the main peculiarities can be identified using measurements performed for lower operating frequencies.

can be identified using measurements performed at lower operating frequencies. In **Figure 10**, the actual measurements of the signal (working frequency 3.5 GHz) are presented. The measurements were carried out in the area of weak radio coverage. The measurements are compared to the *N*-values, calculated using (4) formula and real-time meteorological data, which was measured at the same place. As can be seen, changes of the signal strength occur at same time as the variations of the radio refractivity (please note that the scale is chosen for the convenience of viewing the regularities).

5. Conclusions

In this chapter, the challenges of the very promising ultrabroadband terahertz (THz) wireless networks in terms of the atmospheric attenuation and atmospheric refractivity are presented. The main mechanisms of the atmospheric attenuation are briefly discussed. The simulations, based on the real-time measurements and reliable meteorological data, are presented as well. As most of the atmospheric attenuation mechanisms are quite well predictable or can generally be neglected, the causes of the hardest-to-predict failures are the events of heavy rain (the THz waves are vulnerable by the water molecules) and variations of atmospheric radio refractivity. These simulations are focused on the lower THz band, but could also be applicable for other operating frequencies.

Simulations show that in the events of heavy rain the actual peak values of the rain rate can be twice as high as the peak values calculated using meteorological data, collected with the 10-min integration time. As a result, averaged meteorological data gives inaccurate and distorted results, as the peak values are suppressed and delayed.

The curvature of the ray-path is determined by the atmospheric refractivity. Since the formula of atmospheric refractivity is frequency independent (for frequencies up to 0.1 THz, the error is less than 0.5%), the main peculiarities can be identified using measurements performed for lower operating frequencies. Some initial measurements are presented. The results show that the changes of the signal strength occur at same time as the variations of the atmospheric radio refractivity. Those variations might be very influential in the areas of the weak coverage.

Simulations of the atmospheric attenuation using real-time data are a powerful tool that should complement technological basis, as it will help to foresee possible failures, extend transmission distance and improve reliability of the THz and other high-frequency broadband wireless networks.

Author details

Milda Tamosiunaite[1]*, Stasys Tamosiunas[2], Mindaugas Zilinskas[2,3] and Gintaras Valusis[1,2]

*Address all correspondence to: milda.tamosiunaite@ftmc.lt

1 Center for Physical Sciences and Technology, Vilnius, Lithuania

2 Faculty of Physics, Vilnius University, Vilnius, Lithuania

3 Communications Regulatory Authority of the Republic of Lithuania, Vilnius, Lithuania

References

[1] Cherry S. Edholm's law of bandwidth. IEEE Spectrum. 2004;**41**(7):58-60. DOI: 10.1109/MSPEC.2004.1309810

[2] Cisco. The Zettabyte Era: Trends and Analysis [Internet]. [Updated: June 7, 2017]. Available from: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html [Accessed: 20-09-2017]

[3] Shannon CE. A mathematical theory of communication. The Bell System Technical Journal. 1948;**27**(3):379-423, 623-656. DOI: 10.1002/j.1538-7305.1948.tb01338.x

[4] Roger L. Freeman. Radio System Design for Telecommunication. 3rd ed. Wiley-IEEE Press; 2007. 912 p

[5] Tonouchi M. Cutting-edge terahertz technology. Nature Photonics. 2007;**1**:97-105. DOI: 10.1038/nphoton.2007.3

[6] Kawase K, Ogawa Y, Watanabe Y, Inoue H. Non-destructive terahertz imaging of illicit drugs using spectral fingerprints. Optics Express. 2003;**11**(20):2549-2554. DOI: 10.1364/OE.11.002549

[7] Akyildiz IF, Jornet JM, Hana C. Terahertz band: Next frontier for wireless communications. Physical Communication. 2014;**12**:16-32. DOI: 10.1016/j.phycom.2014.01.006

[8] Ishigaki K et al. Direct intensity modulation and wireless data transmission characteristics of terahertz-oscillating resonant tunnelling diodes. Electronics Letters. 2012;**48**(41):582. DOI: 10.1049/el.2012.0849

[9] Nagatsuma T, Ducournau G, Renaud CC. Advances in terahertz communications accelerated by photonics. Nature Photonics. 2016;**10**:371-379. DOI: 10.1038/nphoton.2016.65

[10] Koenig S et al. Wireless sub-THz communication system. Nature Photonics. 2013;**7**:977-981. DOI: 10.1038/nphoton.2013.275

[11] Hoydis J et al. Massive MIMO in the UL/DL of cellular networks: How many antennas do we need? IEEE Journal on Selected Areas in Communications. 2013;**31**(2):160-171. DOI: 10.1109/JSAC.2013.130205

[12] Ge X et al. 5G ultra-dense cellular networks. IEEE Wireless Communications. 2016;**23**(1):72-79. DOI: 10.1109/MWC.2016.7422408

[13] Akyildiz IF. Enabling next generation small cells through femtorelays. Physical Communication. 2013;**9**:1-15. DOI: 10.1016/j.phycom.2013.04.001

[14] Barros MT, Mullins R. Integrated terahertz communication with reflectors for 5G small-cell networks. IEEE Transactions on Vehicular Technology. 2016;**66**(7):5647-5657. DOI: 10.1109/TVT.2016.2639326

[15] Kleine-Ostmann T, Nagatsuma T. A review on terahertz communications research. Journal of Infrared, Millimeter, and Terahertz Waves. 2011;**32**(2):143-171. DOI: 10.1007/s10762-010-9758-1

[16] Tamosiunaite M et al. Atmospheric attenuation due to Humidity. In: Zhurbenko V, editor. Electromagnetic Waves. InTech; 2011. DOI: 10.5772/21430

[17] Internation Telecommunication Union. P.525: Calculation of free-space attenuation [Internet]. 2016. Available from: https://www.itu.int/rec/R-REC-P.525/en [Accessed: 01-09-2017]

[18] HITRAN database. Available from: http://hitran.org/ [Accessed: 15-09-2017]

[19] Internation Telecommunication Union. P.676: Attenuation by atmospheric gases [Internet]. 2016. Available from: https://www.itu.int/rec/R-REC-P.676/en [Accessed: 01-09-2017]

[20] Brown E. Fundamentals of terrestrial millimeter-wave and THz remote sensing. International Journal of High Speed Electronics and Systems. 2011;**13**(4):1-106. DOI: 10.1142/S0129156403002125

[21] Yang Y, Shutler A, Grischkowsky D. Measurement of the transmission of the atmosphere from 0.2 to 2 THz. Optics Express. 2011;**19**(9):8830-8838. DOI: 10.1364/OE.19.008830

[22] Villermaux E, Bossa B. Single-drop fragmentation determines size distribution of raindrops. Nature Physics. 2009;**5**:697-702. DOI: 10.1038/nphys1340

[23] International Telecommunication Union. P.838: Specific attenuation model for rain for use in prediction methods [Internet]. 2005. Available from: https://www.itu.int/rec/R-REC-P.838/en [Accessed: 01-09-2017]

[24] Moupfouma F, Martin L. Modelling of the rainfall rate cumulative distribution for the design of satellite and terrestrial communication systems. International Journal of Satellite Communications and Networking. 1995;**13**(2):105-115. DOI: 10.1002/sat.4600130203

[25] International Telecommunication Union. P.837: Characteristics of precipitation for propagation modelling [Internet]. 2017. Available from: https://www.itu.int/rec/R-REC-P.837/en [Accessed: 05-09-2017]

Ka-Band HTS System User Uplink SNIR Probability Models

Liping Ai and Hermann J. Helgert

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.75920

Abstract

Ka-band High-throughput-satellite (HTS) systems reuse frequency bands in spot beams for much higher system capacity and better spectrum efficiency. They however are prone to intra-system co-color interference and so suffer from the channel signal-to-noise-plus-interference ratio (SNIR) degradation. This chapter presents the development of the uplink SNIR probability models for Ka-band spot beam HTS systems. The models are applicable to different Ka-band propagation channel conditions of statistical significance. Its use of collective representation to model traffic variation of co-color beams captures the statistics of traffic variation and allows feasibility and variety of use case representation. The analytical approach complements known studies and fills in the blank of the use cases of urban and mobile users. The models can be used for system design performance estimation and prediction. It features computation time and memory savings in numerical implementation.

Keywords: spot beam, signal-to-noise-plus-interference ratio (SNIR), probability model, co-color interference (CCI), Ka-band

1. Introduction

Communications satellites have been transitioning from coverage to capacity steadily in the past three decades because of, mainly, the ever increasing demand for high data rates in broadcast and Internet access, and the big data of Internet of Things (IOT). Going higher in spectrum is one way to address the capacity demand. The first dedicated Ka-band experimental advanced communications technology satellite (ACTS) developed and operated by National Astronomy and Space Administration (NASA) concluded the Ka-band precipitation attenuation models and spot beam hopping via on-board switching. Its many experiments laid out the foundation for the current high throughput satellite (HTS) systems; most of them operate in Ka-band.

A HTS system is a user spot beam system with frequency and polarization (color) reuse applied among spot beams [1]. The higher frequency bands and color reuse increase system capacity or throughput measured by data rate drastically compared to that of the traditional broad beam systems in lower frequency bands without color reuse. Depending on the service area, served population and service type, the later also increase the spectrum efficiency many times. The highest capacity Ka-band HTS system today is ViaSat2 which offers a system throughput of over 270 Gbps.

A variety of uncertainties exist in the complex Ka-band HTS systems including signal propagation channel conditions from a user terminal to satellite, the channel noise, and the varying co-color interference (CCI) power levels. The uncertainties also differ under different system architectures or configurations [2]. They affect the system performance and are studied using probability and statistics before a complex HTS system is constructed. These studies are aimed at assessing system quality of service (QoS) performance matrix against the design specifications and are only feasible via simulations or numerical implementations for a complex system.

The current HTS systems are dominantly geostationary systems for its many advantages [2]. The distinct disadvantage of a HTS system is its many co-color interferers (CCIs) resulted from the color reuse. Moreover, Ka-band systems suffer much higher precipitation attention and noise at higher operating frequency bands. Before DVB-S2, site diversity and power control were the only regularly used fade mitigation methods; link availability at 99.9% is very expensive next to unaffordable. With the use of DVB-S2 capable of covering a link budget gap of more than 15 dB and other digital fade mitigation techniques, uninterrupted operation today has become affordable. To select the right combinations of the available digital technologies, the system designers first estimate the system offered signal-to-noise-plus-interference ratio (SNIR) in a system design as part of the estimation of the other system parameters in simulations or numerical implementations of analytical studies at the early stage of a system design.

The SNIR is the fundamental QoS parameter of a communications system [3]. It determines the achievable capacity with known resources and its complementary cumulative distribution leads to link availability directly. In a HTS system, it varies randomly because of the uncertainties of the signal channel, the CCI channels, the uncertain traffic patterns of the CCIs and the random varying noise in the signal channel for a given system layout and a given satellite antenna design. The SNIR variations also defer in different operational scenarios and application types. Therefore to study the user uplink SNIR of an individual channel or a beam average in a HTS system, we first define the operational scenarios within which the SNIR applies. Each application scenario is often also called use case.

In the remainder of this chapter, we start with a discussion of the major prior works known and relevant to our study in Section 2, follow up with system assumptions in Section 3 with a touch of the use cases and proceed to present the user uplink SNIR probability model in Section 4 applying probability theory to the Ka-band slant path channel to satellite together with the collective representation. Then in Section 5, two specific use cases are explored to show the applications of the SNIR models. Proceeding to Section 6, sample SNIR numerical implementation results of the model system described in Section 3 are evaluated. Finally we conclude our study of the SNIR models and the sample applications in Section 7.

2. Relevant prior work

Several publications reported the research in SNIR estimation for system capacity prediction with iterative and interactive design of the physical layer and the system parameters for a given HTS system with intra-system CCI. In reports by European Space Agency (ESA), a comprehensive HTS system level simulation was described with a model system of 43 spot beams covering Europe and each of the 43 beams is uniformly populated with users [4–7]. The traffic patterns are proportioned and classified for different traffic types with typical traffic probability models. The traffic is fed to each user in a beam one at a time according to the traffic models which is modulated and transmitted to satellite uplink with real system parameters (transmitter EIRP, receiver antenna G/T and etc.). The system adopts multi-frequency time division multiple access (MF-TDMA) system access scheme where a user chooses the channel with the highest power for its traffic delivery which renders one CCI per co-color beam. Typical Ka-band satellite antenna radiation patterns in closed analytical formulas are used for co-color transmission power and signal-to-interference ratio (SIR) distribution estimation. The channel attenuations are assumed fixed. The simulations measure the CCI, SIR, bit-error-rate, call drop rate, link data rates and system capacity at varied system design parameters and traffic load conditions. Two CCI distributions measured at the satellite receiver point are reported in the study, one with the signal beam at the edge of the service coverage area and the other in the center of the service coverage area. The CCI distribution under loaded system condition is symmetric for center located signal beam, resembling a non-central chi-squared (χ^2) probability density function (pdf). It is skewed for the edge located signal beam. The SNIR is not simulated but bench mark calculated.

In another representative HTS system level study, beam average capacity is estimated with the physical layer resource utilization optimization in a loaded system with varied system parameters including transmission powers and reuse size [8]. The system optimization and performance assessment simulation uses the parameters (transmitter EIRP, receiver antenna G/T and etc.). The goal of the study is to improve beam average capacity via DVB-S2X physical layer specification. The total CCI seen by the satellite receiver is simply assumed half-Gaussian distributed in their resource allocation optimization formulated as a mathematical min max problem. The study also reports CCI in χ^2 distribution for the MF-TDMA use case in their Menlo Carlo simulations.

When these studies were made, user uplink traffic is predominately short calls such as web browsing requests, satellite news gathering and email messages. The co-color user powers received by the geostationary satellite are treated as constants in clear-sky (CS) line-of-sight (LOS) channel conditions [4–6, 8]. However, with the introduction of ever increasing number of different data types and operational scenarios such as urban mobile users holding steady calls, the channel multipath variations are to be included which requires the probability representations of the satellite received powers as well as a collective representation of the total CCI power variation as will be shown later. Our study addresses these use cases.

3. System assumptions

A HTS system typically assumes a system architecture shown in **Figure 1**. It consists of a geostationary satellite relaying the communications traffic in the sky, typically several to about ten gateways and tens of thousands to several hundreds of thousands of users served by the spot beams.

A four co-color reuse with user beams on the right hand side is shown in the figure where total bandwidth for user spots is divided into four sub-bands and each is used by a spot beam marked in one of the four numbers. Every four spot beams numbered one through four form a cluster. The total bandwidth is reused by clusters. The advantage of capacity increase and disadvantage of the intra-system CCI of the HTS system can be deducted from the figure readily. The CCI results because the real antenna radiation patterns do not cease beyond the 3 or 4 dB defined beam edge. In theory, the spots are hexagons seamlessly attach each other or circles with triple cross points for every three adjacent spots as shown in the figure. The circles mark the 3 or 4 dB edge of the spots down from the power of the centers of the spots. In real systems, the spots trace ellipsis contours of different sizes with each contour line indicating a power level. Two channels are defined for the typical digital broadcast and Internet access applications. By DVB-S2 convention, the forward channel is from a gateway via satellite to a user while a reverse channel is from a user via satellite to a gateway. Although Ka-band tropospheric precipitation brings much higher attenuation to the slant path signals, the shorter wavelengths of Ka-band have also made high EIRP directional antenna possible with the

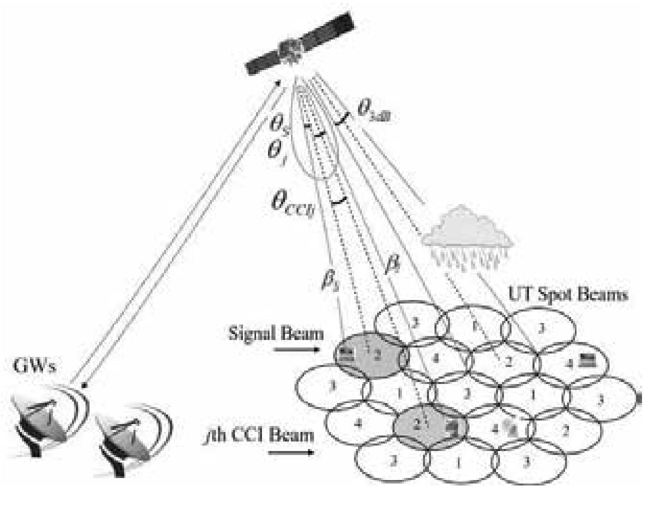

Figure 1. HTS system architecture.

satellite and the gateways. Because gateways typically operate in different frequency bands from that of users and the sites for gateways can be arranged carefully and in advance, interference among gateways uplink is not a performance limiting factor. Our study focuses on the uplink of the reverse channel with users where the color reuse interferences among the spot beams limit the system performance. For its many advantages with geostationary satellite systems, multi-frequency time-division-multiple-access (MF-TDMA) is selected for the model system and it renders one interferer per co-color beam.

The HTS model system which will be used for the system SNIR distribution numerical estimation in this study is a Ka-band system covering US and parts of its borders with 101 spot beams and frequency reuse three as shown in **Figure 2**. The user uplink operates in 30 GHz band. The elliptic spots in the system are in average 350–470 km in minor when illuminated by the geostationary satellite at 99° west with the ITU simple Ka-band satellite antenna radiation pattern at 3 dB apertures of 0.5° side-to-side [9]. The centers of the spot beams are selected a prior according to the service coverage sub-areas. We also assume that there are 1000 users per co-color beam. Each user uses the same user station directional antenna. The user uplink signal received by the satellite is interfered by CCIs because of the imperfect satellite antenna radiation pattern.

The majority of the user reverse channel traffic on the right of **Figure 1** consists of short burst transmissions including emails, network access requests or sales transactions for the second generation Ka-band HTS system applications [10]. This scenario has been studied by simulations and numerical evaluations [4–8]. In this use case, the assumption that the satellite received powers from co-color user transmissions are constants is justified because the signal and CCI power levels vary little in short time frames. Two current application scenarios arise, however, with user reverse channels that require steady transmissions. They are real-time

Figure 2. A user spot beam model system.

news gathering return channel defined in DVB-S2 and teleconferencing. Same applies to future systems with Internet-of-Things (IOT) applications such as WiGRID data. All involve multi-path channels typically.

Because of the urban multi-path environments of the links for the two scenarios, the signal and/or co-color interferers' power levels received by satellite antennas vary over time and in space. Therefore they must be treated as random variables [10]. We investigate the two scenarios as one use case and assess its SNIR performance with an analytical approach which results in simple numerical implementation and statistical accuracy.

Figure 3 illustrates the HTS user reverse channel model with major channel impairments of statistical significance [10]. The SNIR is measured at the reference point of the receiver shown in **Figure 3(b)** in this study.

4. User reverse channel SNIR probability model

With reference to **Figure 1** and **3**, we can formulate the user reverse uplink SNIR as:

$$\mathrm{SNIR} = \frac{\beta_s \epsilon_s(\alpha_s) G_s(\theta_s)}{N + \sum_{j=1}^{n} \epsilon_j(\alpha_j) G_j(\theta_j) \beta_j \mu_j} \tag{1}$$

where the Greek symbols can be read by **Figure 1** except ϵ_j, α_j, N and μ_j. We let μ_j be the traffic factor taking a value between 0 and 1 for the fraction of the time of the interferer traffic presence while the signal is in steady transmission, N the channel noise measured at the reference point of **Figure 3(b)** and $\epsilon_j(\alpha_j)$ the user reverse channel transmitter gain in the direction to satellite away from its nadir. It is clear from **Figure 1** that $\epsilon_j(\alpha_j)\beta_j$ is the satellite antenna received signal power while $G_j(\theta_j)$ is the satellite receiver antenna radiation pattern gain in the direction of incoming signal or interferer. The summation term represents the total co-color interference.

Assume that the user terminal antenna radiation patterns are low in EIRP because of their limited sizes, then $\epsilon(\alpha)$s vary little and Eq. (1) can be simplified to

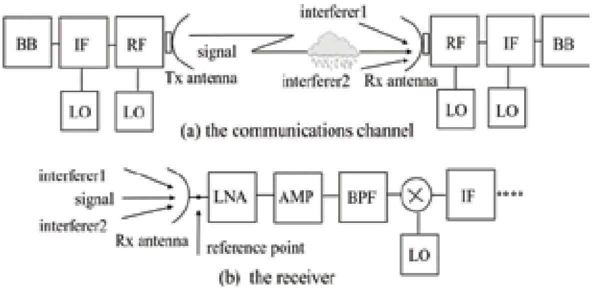

Figure 3. The HTS system uplink transceiver and communications channel.

$$\mathrm{SNIR} = \frac{\beta_s(G)G_s(\theta_s)}{N + \sum_{j=1}^{n} G_j(\theta_j)(G)\beta_j\mu_j} \tag{2}$$

and further simplified to [10]

$$\mathrm{SNIR} = \frac{\beta_s G_s(\theta_s)}{N + \sum_{j=1}^{n} \tau_j\beta_j} \tag{3a}$$

$$= \frac{\beta_s G_s(\theta_s)}{N + \tau\tau_0 a \sum_{j=1}^{n} \beta_j} \tag{3b}$$

by lumping the interferer traffic factors and gains and relocate them outside of the summation. We let τ to represent the traffic factors μ_j collectively while τ_0 satellite antenna angular filtering of the CCIs $G_j(\theta_j)$ collectively. For a mobile user in the urban or open space environment or a stationary user in the urban environment, the user reverse link transmission power varies over time because of the existence of multi-paths. The Rice distribution has long been used for such a propagation channel [10, 11]. It is well known that the Rican distribution is non-central chi-squared (χ^2) distributed in power which can be readily used for the random variable β in clear-sky (CS) line-of-sight (LOS) channel conditions. This χ^2 channel model is applicable for steady transmissions over time in multi-path channel conditions such as urban and mobile users' reverse channels including the future mobile users in autonomous vehicles. It has been shown that the power of SNIR as a random variable in Eq. (3) has a probability density function (pdf) in closed analytical form [10]:

$$f(z)_{S/(den)} = \int_0^\infty y f_{Signal}(yz) f_{den}(y)\, dy \tag{4}$$

where

$$f_{den} = f_{Icllr+N} = \left(\int_0^\infty f_I\left(\frac{x}{y}\right) f_{cllr}(y) \left|\frac{1}{y}\right| dy\right) * f_N \tag{5}$$

is the pdf of the sum of noise random variable and the total CCI random variable. The later has a pdf

$$f_I = f_{\chi^2}(y_{cs}) = \left(\frac{1}{2}\right) e^{-(y_{cs}+\lambda)/2} \left(\frac{y_{cs}}{\lambda}\right)^{\left(\frac{n_1}{2}-\frac{1}{2}\right)} I_{n_1-1}\left(\sqrt{\lambda y_{cs}}\right) \tag{6}$$

where * in Eq. (5) denotes convolution; $\lambda = \sum_{i=1}^{n_1} \lambda_i$, λ_i denotes each interferer's χ^2 distribution parameter for CS LOS channels and steady CCI transmission [10, 12]. For the burst short transmission in the same CS LOS channel, constant power received at the satellite can be assumed [4–8, 10]. In this scenario, the CCI pdf in Eq. (5) can be represented by a delta function

$$f_I = \delta(x - I_t) \tag{7}$$

where I_t is the CCI power of short calls received by the satellite receiver, and

$$f_N(x) = n_0 e^{-n_0 x} \tag{8}$$

is the power pdf of the additive-white-Gaussian-noise (AWGN) with noise power density measured at the reference point of **Figure 3** [3, 10].

The pdf f_{cllr} denotes the collective representation of τ and τ_0 in Eq. (3b) as a scaled probability density function to model the satellite received CCI power variations in the user reverse channels. This is possible because the total CCI power seen by the satellite is typically much less than that of the signal. The collective representation will be shown to produce fairly accurate descriptions of typical system operational scenarios. The separation of the nominal CCI power levels and their channel conditions also facilitates the total CCI power pdf representation owing to channel conditions as a single valid pdf shown in Eq. (6) and Eqs. (8–12) [10].

Precipitation is another important channel condition for Ka-band [10, 13]. For a signal and n_2 CCI channels of precipitation, we find the signal power pdf as the function of the rain attenuation model developed by ITU for individual channels [10, 13].

$$f_s(x) = f_{LN}(10(\log(x)))\left(\frac{10}{\ln(10)}\right)\left(\frac{1}{x}\right) \tag{9a}$$

$$= f_{LN}(A_{ps})\left(\frac{10}{\ln(10)}\right)\left(10^{A_{ps}/10}\right) \tag{9b}$$

where A_P is the ITU rain attenuation random variable in dB. The ITU model predicts annual rain attenuation in dB as lognormal distribution. We then use moment function and Gaussian-Hermit transformation to find the pdf of the sum of n_2 CCI power random variables by equating the means and variances of the lognormal distributions [3, 10, 12]. They give us the total power pdf of n_2 CCI through precipitation slant paths

$$f_I(x) = f_{LN}(10(\log(x)))\left(\frac{10}{\ln(10)}\right)\left(\frac{1}{x}\right) \tag{10a}$$

$$= f_{LN}(A_{ptI})\left(\frac{10}{\ln(10)}\right)\left(10^{A_{ptI}/10}\right) \tag{10b}$$

by equating the moment generating functions and applying the Gaussian Hermit transformation:

$$(-1)^{n2}\prod_{j=1}^{n2}\left(\sum_{n=1}^{k}\left(\frac{w_n}{\zeta\sqrt{\pi}}\right)e^{-sI_j(a_n)}\right) = -\sum_{n=1}^{k}\left(\frac{w_n}{\zeta\sqrt{\pi}}\right)e^{-sI(a_n)} \tag{11}$$

where

$$I_j(a_n) = b\left(10^{-Apj/10}\right) = b\left(10^{-\exp\left(\frac{(\sqrt{2}a_n\sigma_j+\mu_j}{\zeta}\right)/10}\right) \text{ j} = 1, 2, \ldots\ldots, \text{n}_2 \tag{12}$$

$$I(a_n) = n_2 b\left(10^{-Ap/10}\right) = n_2 b\left(10^{-\exp\left(\frac{(\sqrt{2}a_n\sigma+\mu}{\zeta}\right)/10}\right) \tag{13}$$

where the (σ_j, μ_j) are known. The (σ, μ) in Eq. (13) can be solved by setting two real values for parameter s in Eqs. (10b–11) [10, 13]. The pair is the lognormal distribution parameters for A_{ptl} in Eq. (10b). The justification for the model selection and development can be found in [10].

5. Collective representation

A quick examination of Eq.(5) shows that the user reverse channel SNIR model consists of noise probability model, the signal and CCI channel models and the collective representation f_{cIIr} of the distribution of the total CCI power filtered by traffic activity and the satellite antenna radiation pattern. For all co-color users in multi-path urban or mobile channel conditions of Eq. (6) or Eqs. (8–12), a Beta collective representation in Eq. (3b) represents well the typical 24 h traffic pattern classification of full, heavy, average or light loads for the simplified random user spatial distribution scenarios because all CCI call holding time fall into interval between zero and one when normalized with respect to the signal call holding time [10]. The constant one assumed by the collective traffic τ represents the fully loaded CCI condition. The Beta distribution has a pdf:

$$f_{Bt}(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} \tag{14}$$

Figure 4 shows a typical traffic pattern set in Beta collective representation for the simplified user spatial distribution use case [10]. Let the simplified user location distribution scenario or use case described above is use case one, then the total CCI power filtered by the satellite antenna radiation pattern in Eq. (3b) is represented collectively by a Beta probability distribution function for use case one [10]. In general, depending on the application scenario or use case, it is collectively represented by a combination of probability distribution functions. This is shown later in this section with fully loaded systems of use cases two and three.

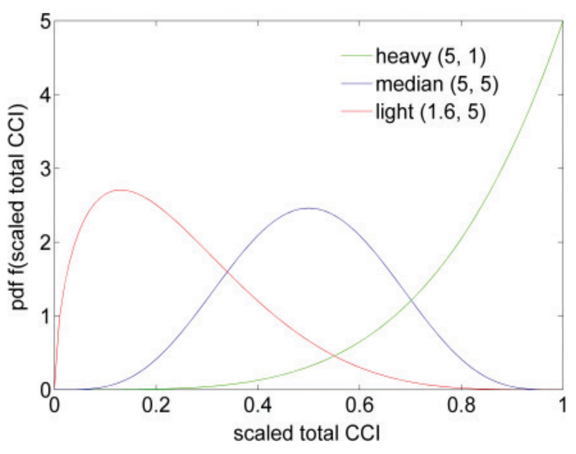

Figure 4. CCI traffic Beta collective representation.

The user reverse channel calls in current HTS systems are most often short calls such as emails or network access requests originated from the randomly located stationary users with CS LOS channels in the co-color beams. In such calling environment, the signal power loss along the slant path to satellite varies little and a constant loss is justified. Another typical use case scenario would be a few urban long holding time calls in the midst of the short stationary calls during the peak hours in the co-color beams of the model system of **Figure 2**. The long calls could be video teleconference calls from an office of an organization header quartered in the downtown area of a city with satellite offices in difference cities of similar urban environments or a news gathering van in an urban area reporting an event during peak traffic hours when the system is loaded. By adaptively applying the models in Section 3 to these two typical use cases, namely, the all short calls use case two and the mixed calls use case three in the model system of **Figure 2** with several further assumptions, we investigate the scenario collective representation in CS LOS channel conditions. First, we expand our original simplified user location assumption to 1000 users uniform or/and Gaussian randomly located in the co-color spot beams. Second, the system is assumed fully loaded. One co-color call at a time from one of the 1000 random user locations is made to the satellite continuously from all co-color beams. This equates the CCI traffic factor in Eq. (3b) to 1 and puts the variation of the total CCI power across the co-color beams received by satellite at the receiver reference point of **Figure 3** dependent on the satellite antenna radiation pattern angular filtering of the incoming CCIs. Finally, the second green beam in the right most column green beams at the edge of the service coverage area in **Figure 2** is designated as the east signal beam (ESB). The rest 32 green beams constitute CCIs to the ESB signal at the satellite receiver. Together, the green 3 reuse system is denoted as ESBg3.

In use case two of all short calls for ESBg3, the total CCI power distribution over time is obtained by summing the 32 CCI gains seen by the satellite antenna at any time instance over equally weighted 1000 time intervals. This distribution at uniform user location distribution can be collectively modeled as a linear combination of two scaled Beta distributions of different parameters:

$$f_{ccis} = f_{cllr} = a_1 f_{Beta1} + b_1 f_{Beta2} \tag{15}$$

For the same system but mixed call use case three, it is a combination of an exponential and a Gaussian distribution:

$$f_{ccis} = f_{cllr} = a_2 f_{exp} + b_2 f_{norm} \tag{16}$$

The parameters a_1, b_1, a_2, b_2, λ of f_{exp}, (μ, σ) of f_{norm}, (α_1, β_1) of f_{Beta1} and (α_2, β_2) of f_{Beta2} vary with the type of user location distribution, the satellite antenna radiation pattern and the number of interferers. Clearly, the *a*s, and *b*s are constrained by $a + b = 1$. **Figures 5** and **6** show two use cases ESBg3 typical short call CCI power distributions and their collective representations by Eqs. (15–16). The collective approximation can be refined with additional or/and different combinations. The parameters for the two figures are tabulated in **Table 1** of Appendix A.

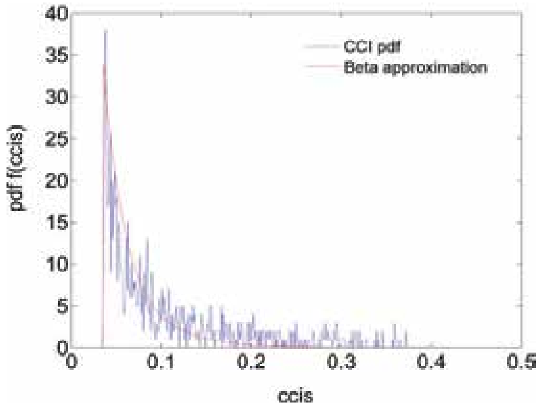

Figure 5. Beta collective approximation of CCI distribution in use case two.

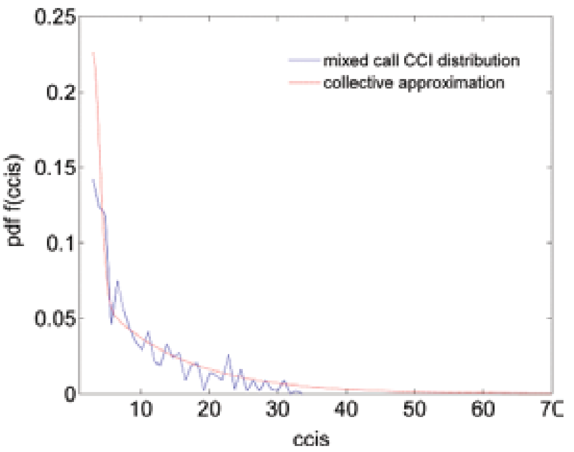

Figure 6. Combined exponential and normal collective representation of mixed call CCIs.

Both use cases in **Figures 5** and **6** illustrate the averaged total CCI power temporal variation when the equally probable random short calls at the co-color user reverse channels progress at full system load. A more realistic scenario of use case two is that the user short call holding times also vary randomly. This scenario can be modeled by weighting the 1000 user call times by a random number generator for each co-color beam which also embodies the spatial CCI variation of that beam. A simple CCI spatial variation approximation has been made in use case one [10]. Clearly, depending on the intended SNIR distribution estimation, the total nominal CCI power variation represented collectively by the random variable τ_0 in Eq. (3b) can be modeled in a variety of ways.

6. Sample applications

In this section, we assess Ka-band HTS system user reverse channel SNIR performances in use cases defined above and present the sample assessment results using the model system of **Figures 2** and **3** and the assumptions made progressively in Sections 3 through 5. They serve to capture the typical geostationary Ka-band HTS system characteristics essential in system design. We also show the wide range applicability and scenario representation flexibility of the models. A central signal green reuse 3 system (CSBg3) is defined with the signal beam being at the center of the service coverage area at the intersection of the fifth column and third row of the green reuse beams in **Figure 2** similar to ESBg3. We exam:

- Use case two CCI distributions characteristics and comparison
- Traffic distribution and user location impact on the SNIR distribution by use case one
- Beam average SNIR distributions and comparison for use case two
- Individual short call SNIR pdfs in all short call scenario of use case two
- Steady call SNIR distributions in a mixed call environment of use case three

The green reuse beam system of **Figure 2** is simulated as shown in **Figure 7** for use. The light green beam shows the ESB whereas the CSB sits in the middle among the five dark green beams. They are used with user case two and three results presented below. Gaussian user location distribution is used for **Figure 7** whereas both Gaussian and uniform user location distribution are used for SNIR performance evaluation. We choose the centers of the user location distributions at the co-color beam centers. The beam centers are determined according to the service demand of the area. For user case one, the simplified user random location scheme presents one user per co-color beam within the beam coverage area [10].

Continuing on CCIs, we find that the CCI of CSBg3 is heavier than that of ESBg3 in **Figure 8** and the typical co-color beam CCI spiky distributions in **Figure 9** for use case two. Clearly, the

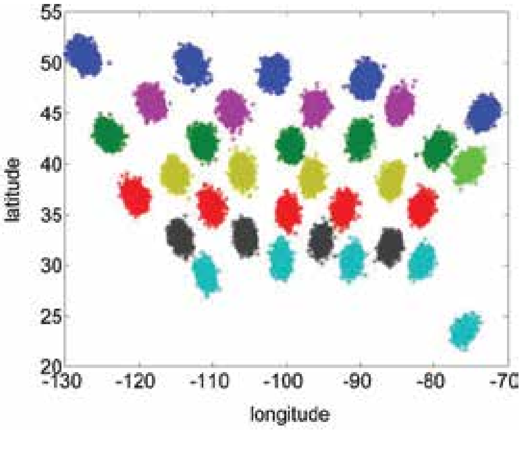

Figure 7. ESBg3 Gaussian co-color beams simulation illustration.

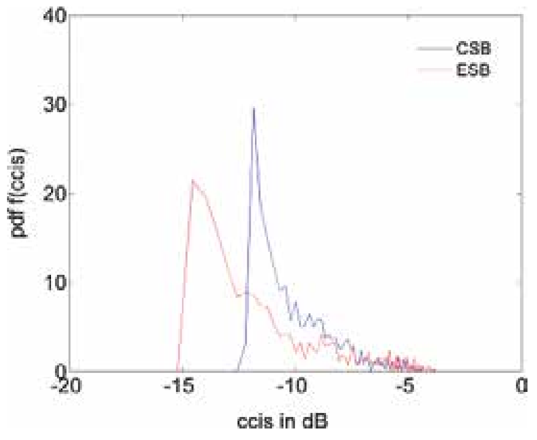

Figure 8. Total CCI distribution comparison.

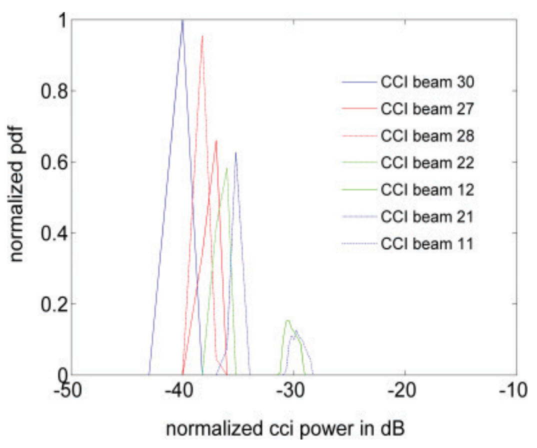

Figure 9. Individual beam CCI distribution.

average 2.5 dB more CSBg3 CCI results because the CSB is surrounded four sides by the close-by first 2 tier CCIs whereas the ESB faces them in three sides only. The individual beam CCI distributions are shaped like spikes within 3 dB because of the many non-linear transformations from the user location Uniform distribution in contrast to the signal beam power distribution which approximates a Uniform.

The traffic distribution impact and typical co-color beam user location random variation effects on ESBg3 user reverse channel SNIR performance for use case one are shown in **Figures 10** and **11**. The traffic distribution in **Figure 4** is used for **Figure 10** which results in up to 1 dB difference at peaks. This is because in use case one, the CCI gain reduction with Gaussian satellite antenna radiation pattern is much more pronounced than traffic load variation in the system. **Figure 11** further shows that the simplified CCI user random location in use case one renders SNIR distribution approximately 1 dB that of all co-color users at the beam centers.

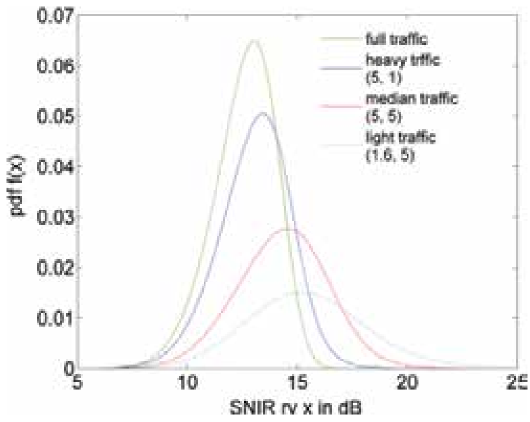

Figure 10. Traffic impact (CSBr4).

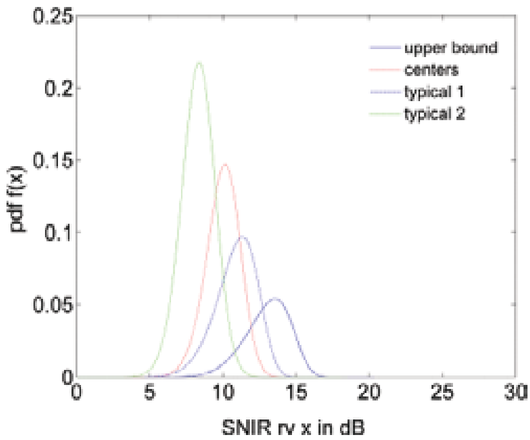

Figure 11. ESBg3 user location effect.

Therefore, the all center user SNIR distribution is a fair approximation of the beam SNIR distribution average [10].

For use case two of all short calls Uniform or Gaussian distributed in each and every co-color beams, only the beam average signal-to-inference ratio (SIR) or SNIR using average noise power spectrum density is measurable and meaningful. In **Figures 12–14**, we compare the beam average SNIR distributions in CS LOS channel conditions with fixed channel power loss:

$$f_{SNIR} = f\left(\sum_{j=1}^{1000} \frac{S\left(x_j,\ \ y_j\right)}{n_0 + \sum_{k=1}^{32} I_k\left(x_{j,k},\ \ y_{j,k}\right)}\right) \tag{17}$$

where f denotes the frequency counts of the SNIR values calculated at 1000 users in the parenthesis.

http://dx.doi.org/10.5772/intechopen.75920

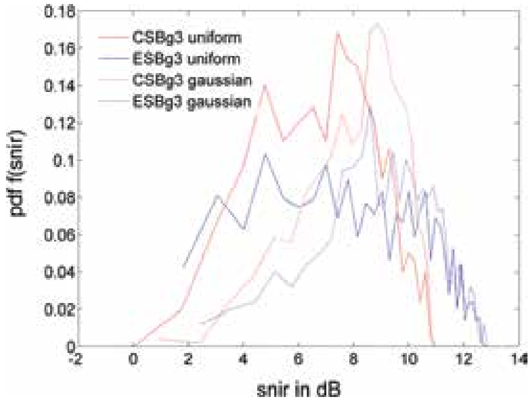

Figure 12. A beam average SNIR distribution comparison.

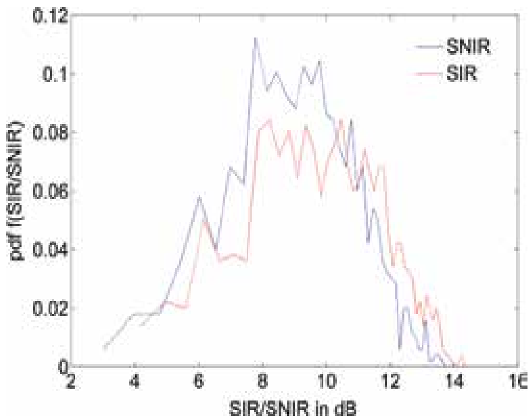

Figure 13. Beam average SIR/SNIR.

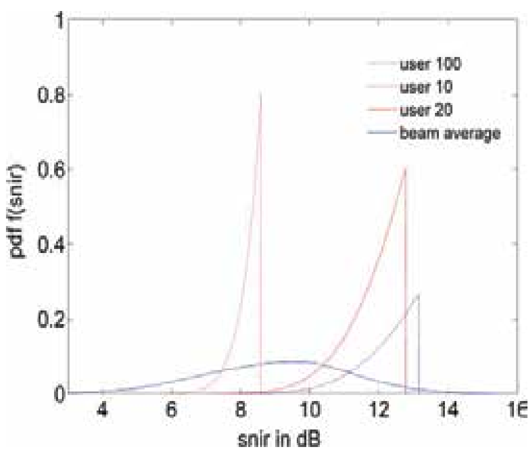

Figure 14. Individual user SNIRs.

Among the four distributions in **Figure 12**, the user location Uniform distributed SNIR pdfs show a flat top from 4 dB to approximately 9 dB whereas the user location Gaussian distributed SNIR pdfs are pointed with peaks at approximately 8 and 9 dB. This results because the Gaussian user location distribution produces higher user density at the beam center but much lower density of users at the edge of the beam coverage area compared to Uniform user location distribution in which users are evenly distributed in the beam coverage area. The former has a larger distance separation between CCI users and signal users which results in less CCIs as shown in **Figure 8**. Consequently, the SNIR performance improves. **Figure 12** also shows the ESBg3 system performs 2 dBs better on the high end due to the overall less first 2 tier CCIs. The high pdf value at low CCI power results because the many more low CCI powers produced by the many far away CCI beams exit in the ESBg3 system. **Figure 13** plots the average beam performance degradation by approximately 2 dB at the high end and 1 dB average at the low end because of the inclusion of noise.

The beam average SNIR power spreads are approximately 11 dB in **Figure 12** with ESBg3 leading 2 dB higher at high end. When noise is taken as a fixed power density value, the individual user SNIR is a fixed value for short reverse channel calls. If the noise power is taken as a random variable instead of the mean power density for individual users and by treating the CCI power as a delta function as shown in Eq. (7), we find the SNIR distributions for the typical user short calls of user case two in **Figure 14**. Clearly, for short calls where the signal and the CCI are assumed fixed power levels, the SNIR starts from a fixed value and spreads to the lower end to left as noise power spreads to the higher end at right. Compared to the beam average SNIR spread of approximately 11 dB also shown in **Figure 14**, the individual call SNIRs vary only 2–3 dB.

By treating the total CCI as a fixed value and using delta function as its pdf representation as shown in Eq. (7), we reach the beam average SNIR via probability theory using Eqs. (4–5).

Figure 15 shows the good approximation of the beam average SNIR distribution of the user Gaussian distributed ESBg3 system by applying the probability models. The high end

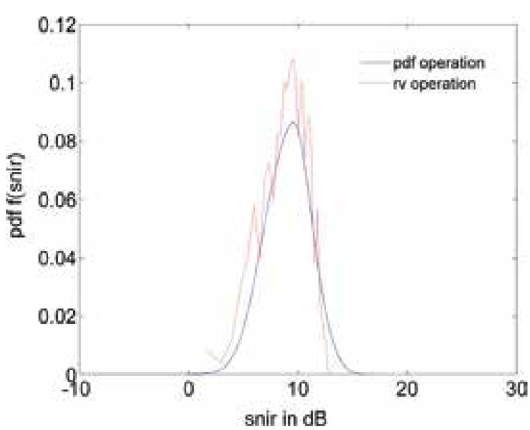

Figure 15. ESBg3 beam average SNIR distribution.

mismatch is due to the large noise power spread of the noise probability model used with pdf operation. The beam signals for the figure are approximated by a Beta distribution collectively. The parameters of the figures are given in **Table 2** in the Appendix A.

For use case three of mixed calls, we set three CCI beams holding the long call with ESB. They are the beams at Caribbean, in the middle of the continent US and at the northwest of the continental US. The rest of the CCI calls are short random calls one at a time in a beam. The ESB is located at the northeast of continental US by the edge of the service area. The total 29 short CCI calls vary a little from that of the 32 short calls in distribution. The short call CCI distributions are collectively modeled as a weighted combination of exponential and Gaussian in Eq. (16) whereas the long signal and CCI calls are each χ^2 distributed with a power scaling factor determined by the satellite antenna radiation pattern angular filtering. As shown in **Figure 16** the SNIR performance of user Gaussian distributed ESBg3 outperforms the user Uniform distributed ESBg3 SNIR by approximately 4 dB at peak and Gaussian SNIR is approximately equally probable from 6 to 12 dB indicating a consistent performance at the power spread range. Compared to the user location Uniform distributed mixed call ESBg3 SNIR, it ceases approximately 2 dB higher in power spread at lower end because of the low density CCI users at the beam edges.

Compared to the beam average SNIR performance of use case two with all short reverse channel calls shown in **Figures 12** and **15**, the SNIR of use case three mixed calls of **Figure 16** shares with them the most probable power level of about 10 dB at Gaussian user location distribution. They differ however in spreads for about 2–4 dB because, in part, that the long call χ^2 distributions and the noise exponential distribution included in **Figure 12** contribute to the larger SNIR power spread than that of the beam average SNIRs. Compared to the user case one of **Figures 10** and **11** for which the Gaussian satellite antenna radiation pattern is used, use case two and three is less performing in SNIR power level because the sharp slope of the Gaussian antenna pattern filters out much more CCI power. In those simplified random individual user location scenario, the channel SNIRs are approximately 10 dB less in spread due to the CCI power spatial simplification.

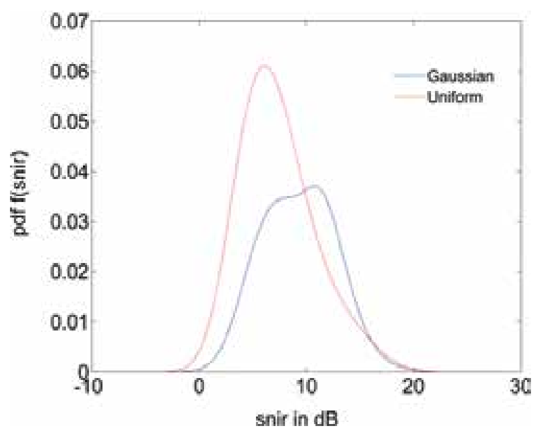

Figure 16. ESBg3 mixed call SNIR distributions.

7. Conclusions

Two approaches are commonly used in assessment of a complex system of many variables including random variables. They are simulation and analytical approaches or a combination of the two. Ka-band HTS systems are complex systems with many uncertain quantities including traffic loads, channel weather conditions, and intra-system CCIs. This study chooses the random variables of the system of statistical significance, and develops the probability models by taking an analytical approach to characterize the variations of the system or channel SNIRs, the most important QoS measure of a Ka-band HTS system. It results in a closed analytical SNIR distribution model which can be implemented by numerical evaluations, leading to significant savings in both computation time and storage. Complementary to the known studies [4–8], it captures the statistics of the random variations relevant to SNIR estimation by abstraction and fills in the blank of the study of the topic for urban and mobile user scenarios.

By treating the signal and CCI power variation caused by the channel conditions as random variables and the traffic variation and CCI power variation caused by satellite antenna radiation pattern angular filtering with collective representations, this study allows for system offered user uplink SNIR prediction with significantly reduced computational complexity in the early stage of a system design. It is statistically accurate for user uplink SNIR estimation in multi-path channel environments particularly [10].

Our discussion of the model application focuses on the collective representation of application scenarios in CS LOS channel conditions in this chapter. The sample application result of **Figure 9** confirms the beam CCI distribution reported in [4–6]. Moreover, it can be inferred by an examination of **Figures 5**, **6**, **8** that the total CCI distribution can verify the half Gaussian distribution assumption made in [8] if less mid-range CCIs more far away tier CCIs are filtered into the satellite receiver by the satellite antenna radiation pattern. The average beam CCI power distribution spread of 14 dB is comparable to the 15 dB found in [4–6]. The 1 dB difference should come from the different number of beams which is 43 in ESA's system performance simulation whereas it is 101 in this study. Though not shown here, our other sample results with ESBg3 and CSBg3 share the spatial CCI distribution characteristics with those reported in ESA.

Compared to ESA study of the same topic, the collective representation of this study can achieve the same traffic simulation effects by profiling the user traffic factor in Eq. (3b) and therefore predict the system capacity performance at offered SNIR distributions for a wide range of system architecture and operational scenarios. However, it can't achieve the interactive design of the physical layer of the system for optimal resource allocation at different system traffic loads by the total system simulation efforts reported in [4–6, 8]. It can be used in the initial stage system architecture design and performance prediction at beam and user level stand alone or incorporated into total system operation simulation packages as a SNIR estimation unit. In real system design, user population probability models and beam centers can vary in accordance with the service area and population distribution. The study is applicable to real system satellite and user antenna radiation patterns and other system parameters. The sample results should improve with them [4–6, 10].

A. Appendix A

Figure 5: use case two of ESBg3 short calls with user location Uniformly distributed
Semi-minor uniform (−164, 164); semi-major uniform (−185, 185)
a_1 = 0.6, b_1 = 0.4, α_1=1, α_2=1, β_1=17, β_2=6, max(ATT_{ccisum}) = 0.3898
Figure 6: use case three of ESBg3 mixed calls with user location Gaussian distributed
Semi-minor Gaussian (0, 53); semi-major Gaussian (0, 75)
a_2 = 0.55, b_2 = 0.45, λ = 9, μ = 5, σ = 2.5

Table 1. User total CCI collective representation parameters at $\theta_{3dB} = 0.5^{\circ}$ and $\lambda_S = -99^{\circ}$.

Figures 14 and **15**: use case two at short calls of user location Gaussian distributed
Semi-minor Gaussian (0, 53); semi-major Gaussian (0, 75);
Signal: beta(6, 1); Noise: n_0 = 0.0134; CCI: **Figure 6** parameters;
Figure 16: use case three of mixed calls
Steady call[1] CCI beam numbers: 5, 16, 25, 32; χ^2 distribution K factor: 100

[1]The four calls that last the entire SNIR evaluation time frame when the other calls are short bursts in the mixed call use case three

Table 2. ESBg3 SNIR distribution parameters at $\theta_{3dB} = 0.5^{\circ}$ and $\lambda_S = -99^{\circ}$.

Author details

Liping Ai* and Hermann J. Helgert

*Address all correspondence to: lai002@gwu.edu

George Washington University, Washington, DC, USA

References

[1] Fenech H. What pushed us into HTS systems? In: 23rd Ka and Broadband Communications Conference and 35th AIAA ICSSC; Trieste, Italy; 16 October, 2017

[2] Mclain C et al. Future Ku-band mobility satellites. In: 23rd Ka and Broadband Communications Conference and 35th AIAA ICSSC; Trieste, Italy; 17-October-2017

[3] Ippolito L. Satellite Communications Systems Engineering. 2nd ed. Wiley; 2008. ISBN-13: 9780470725276

[4] Schweikert R et al. ESA STUDY CONTRACT REPORT: Protocols and Signaling for Adaptive Fade Mitigation Techniques (FMT) in DVB-RCS Multi-Beam Systems, Final Report. European Space Agency (ESA); AUDENS ACT 2005

[5] Schweikert R et al. Protocols and Signaling for Adaptive Fade Mitigation Techniques (FMT) in DVB-RCS Multi-Beam Systems, Technical Report 1: Scenario Definition and Benchmark Assessment. ESA Reports; 2005

[6] Gallinaro G et al. Adaptive Coding Modulation Techniques for Ka/Q Band Systems – TN1 Reference Scenario Definition. European Space Agency (ESA); 2003

[7] Fenech H. High throughput satellite systems: An analytical approach. IEEE Transactions on Aerospace and Electronic Systems. 2015;**51**:192-202

[8] Rinaldo R et al. Capacity analysis and system optimization for the reverse link of multi-beam satellite broadband systems exploiting adaptive coding and modulation. International Journal of Satellite Communications and Networking (IJSCN). 2004;**22**:425-448

[9] ITU-R. Satellite antenna radiation pattern for use as a design objective in the fixed-satellite service employing geostationary satellites. ITU-R S.672-4; 1997

[10] Ai L. A Novel Study of the SNIR distributions of Ka-band HTS systems [PhD dissertation]. George Washington University; 2016

[11] Wakana H et al. Fade characteristics for K-band land-mobile satellite channels in Tokyo measured using COMETS. Electronics Letters. 1999;**35**(22):1912-1913

[12] Leon-Garcia. Probability, Statistics and Random Processes for Electrical Engineering. Upper Saddle River: Pearson Education; 2008. p. 07458

[13] ITU-R. Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems. ITU-Recommendation on Propagation.618; September 2013

Section 2

Optical Networks

Chapter 11

High-Speed Optical In-House Networks Using Polymeric Fibers

Ulrich H.P. Fischer-Hirchert, Matthias Haupt,
Mladen Joncic, Stefanie Haupt and Sebastian Höll

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72204

Abstract

Data communication over polymer optical fibers (POF) is a good alternative method for local area networks to use an optical medium to transmit data in short-range environments like cars or copper in-house networks on the basis of IEEE 802.3. Many companies offer transceivers for the area of Ethernet networks in the visible wavelength range. In the first part of the chapter, a system comparison of manufacturers with interoperability check is presented. Here, the real transfer rates within a manufacturer and between all manufacturers are measured as a cross-check. In the second part of the chapter, the limitation of bandwidth due to the use of only one wavelength channel is discussed. Wavelength Division Multiplexing (WDM) is a promising candidate to significantly increase bandwidth in POF to more than 40 Gbit/s. Here, the problems in the development and manufacture of a demultiplexer (DEMUX) for WDM over POF as well as the results of the optical separation of four wavelength channels are described. At least, the possible extension of a WDM grid of ITU G.694.2 is discussed, which seems to be a hopeful candidate to introduce a standardized WDM grid for POF in the visible range to reach data rates of 40 Gbit/s up to 50 m POF.

Keywords: polymeric fiber transmission, optical networks, WDM over POF, wavelength division multiplex, demultiplexer, injection molding

1. Introduction

Polymer optical fibers (POF) are used in various fields of applications. The core material consists of polymethylmethacrylate (PMMA), while the cover is made of fluorinated PMMA. The whole fiber has a diameter of 1 mm, which is depicted in **Figure 1**. POFs are used for optical data transmission based on the same principle as glass fiber. As a communication medium,

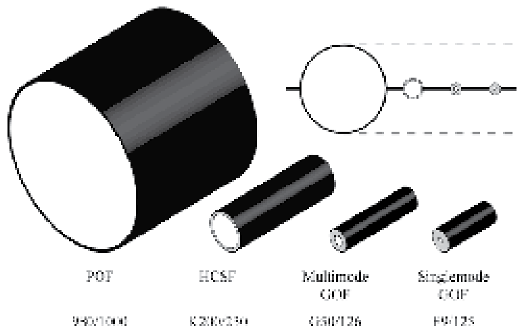

Figure 1. Comparison of optical fiber types.

they offer a couple of advantages related to other data communication systems such as copper cables, glass fibers and wireless systems and have great potential to replace them in different applications.

In comparison with glass optical fibers (GOF), POFs are easy to use in the field because of low bending losses and a large optical core of 980 μm. This makes the POF very insensitive to rough and dusty environments as well as losses on plugs in comparison with glass fibers [1]. However, one advantage of using glass fibers is their low attenuation, which is below 0.2 dB/km in the infrared range. The attenuation of polymeric fibers in the visible spectrum from 350 to 750 nm (see **Figure 2**) is much more higher with its minimum of 85 db/km at a wavelength of 570 nm. For this reason, the use of POF in communication systems is focused on short distance communication from 10 to 100 m. The larger core diameter of POFs leads to high-mode dispersion of more than 2.2 millions of optical modes. Additionally, the high attenuation at wavelengths higher than 700 nm limits the application of the POF to the visible spectrum of light (400–700 nm). Here, POFs can outperform the current standard of copper cable as a communication medium. On the one hand, they feature lower weight, low bending radius and space. On the other hand, POFs are not susceptible to electromagnetic interference [2, 3]. For these reasons, POFs are already used in various application domains, for example, in the automotive sector and for in-house communication [4–7].

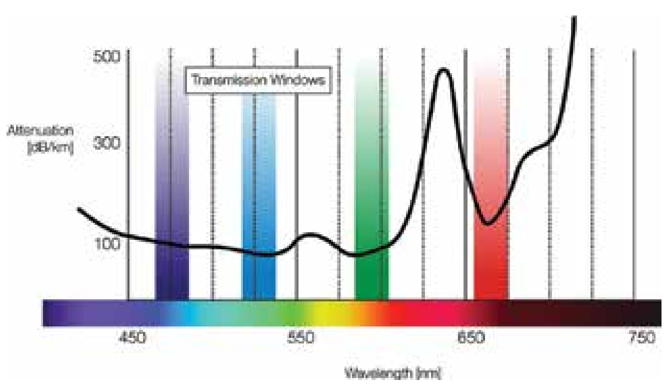

Figure 2. Attenuation of POF in the visible range [1].

http://dx.doi.org/10.5772/intechopen.72204

1.1. POF in the automotive sector

Typically, copper-wired bus systems are used in the car environment. In the past 15 years, POF has replaced the electrical wiring in many types of car (see **Figure 3**). It was first introduced by BMW in the 7er series in 2001. Since then, not only high-class cars were equipped with POF, actually more than 200 types of volume cars benefit from the advantages of POF [4, 5]. The used bus is called **M**edia **O**riented **S**ystems **T**ransport (MOST), which is a multimedia network optimized for multimedia and infotainment applications. The bus was developed by the automotive industry. It works in three data rate levels with 25, 50 and 150 Mbit/s. MOST defines basically the physical interconnection between devices by using POF as a transport medium. Additionally, it specifies and standardizes a communication protocol to develop complete systems and applications to distribute multimedia content for the car.

The replacement of the communication technique from copper wires to POF leads to lower weight. The low melting temperature of PMMA (95°C) still prevents the use of POF in the engine compartment. However, new types of fiber in the development that have higher glass transition temperature will allow the use of high-temperature POF in the engine compartment in the near future [4]. Another application in the car, where POF most likely will be used in the future, is as sensors for measuring various in-car pressures or forces. Additionally, sides emitting polymeric fibers are interesting devices for future applications for ambient lighting in the passenger cabin.

1.2. POF for in-house communication systems

Another sector where POF displaces the traditional communication medium is in-house communication [6, 7], although the possibilities of application are not confined to the inside of the house itself. In the future, POF can possibly displace copper cables for the so-called last mile between the last distribution box of the telecommunication company and the end consumer. Today, copper cables are the most significant bottleneck for high-speed Internet.

"Triple Play" is called the combination of IPTV, VoIP and the data Internet. The combination is currently introduced into the telecom market; therefore, high-speed connections are essential. It is highly expensive and bandwidth limiting to use Ethernet in-house system using copper components, thus the future will be FTTH, in combination with optical in-house wires of POF or GOF (see **Figure 4**).

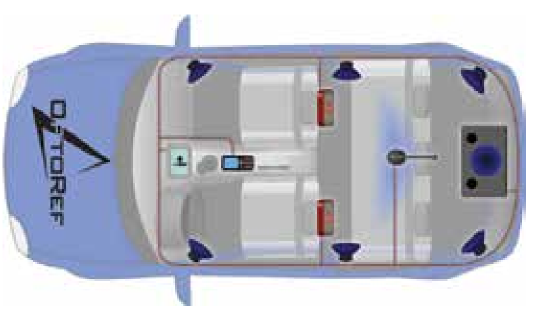

Figure 3. Multimedia bus system (MOST-bus) with POF.

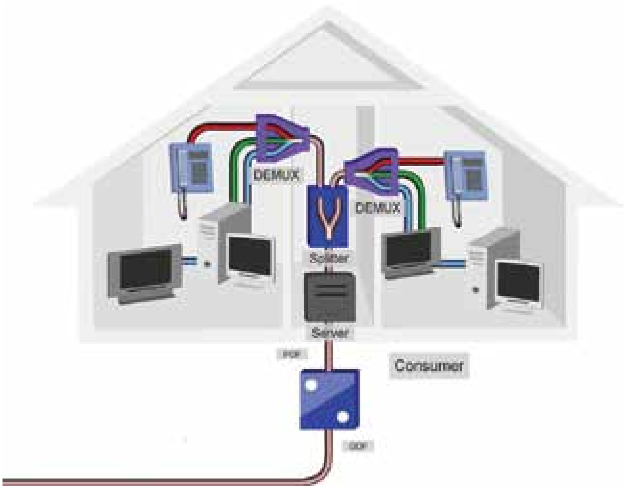

Figure 4. In-house communication with POF.

2. Studies on the interoperability of different transceivers for optical polymer fibers

In this chapter, various transceivers, which can be used for data home cabling with optical polymer fibers, are examined with respect to their interoperability. Eight different devices are tested for their effective data rate over length-varying POF transmission distances. Furthermore, the results are compared with the manufacturer's data regarding performance and the interoperability of all devices is checked.

The Photonic Communications Lab at the Harz University works closely together with manufacturers of various POF components in several projects. From this cooperation, the question of the compatibility of POF devices from different manufacturers among one another has become increasingly important. All devices tested comply with the IEEE 802.3u guidelines for Fast Ethernet. Fast Ethernet is mainly used in local networks and allows data transmission at 100 Mbit/s.

2.1. Devices under test

Table 1 shows the tested media converters or switches with specifications from the manufacturer:

As can be seen in **Table 1** and **Figure 5** devices from various companies are examined. Starting with media converters of German manufacturers (Siemens, Diemount and Rutenbeck) to transceiver-switches from Homefibre (Austria) and BSPCOM (China).

A USB media converter from the company BSPCOM could not be considered for investigations for reasons of problems with the USB driver software for Windows (**Figure 5**).

Name	Wavelength (nm)	Transmission length (m)	Data transfer rate (Mbit/s)
Speedport OptoLAN	670	30	100
Diemount CS-116	470	70	100
Rutenbeck wall socket	660	70	100
Rutenbeck socket switch	660	50	100
Media converter Rutenbeck	660	50	100
Switch OMS 126S-150 Homefibre	650	50	100
Switch CP8016 BSPCOM	650	50	250

Table 1. Used POF-transceivers.

Figure 5. Tested media converter: Speedport, Diemount, BSPCOM, Rutenbeck and Homefibre.

All the devices tested, except the CS-116 from Diemount, operate at a wavelength in the visible red range. The device from Diemount transmits data at the wavelength of 470 nm in the visible blue range (see **Figure 2**).

2.2. Experimental setup

In the investigations of the POF devices, each existing one is combined with each, connected by a POF of 50 m length and put into operation. The measurements are carried out using a certification scheme developed in the Photonic Communications Lab of Harz University in accordance with the ETSI TS 105175–1 V1.1.1 (220010-01) standard, which establishes an in-house networking of the optical polymer fibers.

The optical polymer fibers are wound up with the aid of two cylinders. These cylinders have different diameters and thus offer different bending radii (see **Figure 6**) in order to apply the typical application conditions of a typical LAN network distribution in an apartment.

Several optical polymer fibers are wound onto this structure. These differ in length and outer cable diameter. However, all have a step index profile with a core diameter of 980/1000 μm. The cable diameter varies between 1.5 and 2.2 mm. The 2.2 mm fibers are being designed for simplex transmissions only. The 1.5 mm fibers are duplex fibers. The lengths of the optical polymer fibers are: 1, 15, 30 and 50 m.

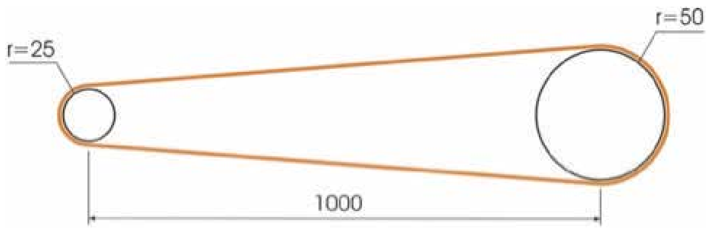

Figure 6. Setup for testing typical laying of a POF with a length of 30 m in an apartment with 15 bends of different radii (in mm).

2.3. Transmission speed measurement with Jperf 2.0.2

Iperf is a command line utility for measuring the performance of networks. Jperf is a graphical interface developed in Java for Jperf. This program is started on two PCs, one of which is the function of the server and the other is assigned to the client (see **Figure 7**). The server accepts connections on TCP port 5001. Data are transferred from the client to the server for the duration of the measurement. Thus, unidirectional data transmission always takes place.

Iperf offers different, adjustable parameters for throughput measurements. Examples of this are the selection of the transmission protocol (TCP/IP or UDP) as well as the modification of the measurement duration. In addition, the buffer size can be changed. The measurements are carried out in transmission control protocol (TCP) [8].

2.4. Transfer rates

At a transmission distance of only 1 m (back to back), all media converters and switches are working together with transmission speeds in the range of 90 Mbit/s (see **Figures 8** and **9**). However, in some combinations, the quality of the transmission rate is lower for this short distance than for a longer distance such as 30 or 50 m. Overdriving at the photodiode due to the excessive light intensity may cause this.

All the transceivers of the different manufacturers have been able to communicate with each other easily and also over the distances of 15 and 30 m, and all tests have been positive. The data rate fluctuates ±1–2 Mbit/s in the range of 92 Mbit/s. A 50 m transmission cannot be positively tested in combinations in which the Diemount CS-116 media converter is used as a client and with a blue transmit diode. This can be explained by the fact that the tolerance window of the photodiodes of the other devices is setup in the red range to this range by 650 nm. However, it should be noted that the light output up to 30 m was still intense enough to achieve a functionality of the two wavelengths without problems.

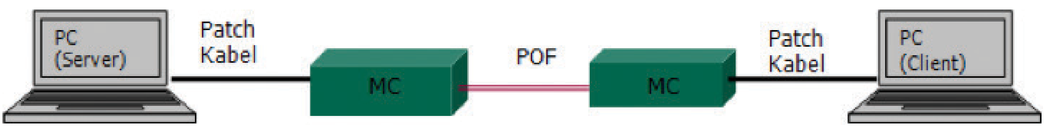

Figure 7. Measurement setup with Jperf (MC (media converter)—DUT.

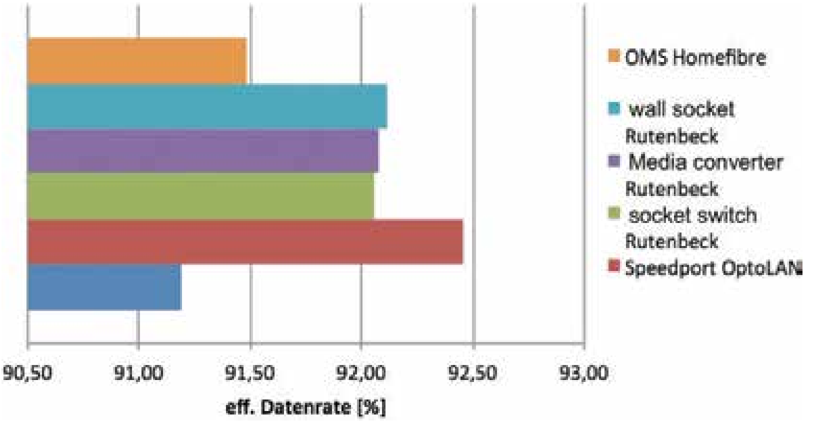

Figure 8. Effective measured data rates within one manufacturer.

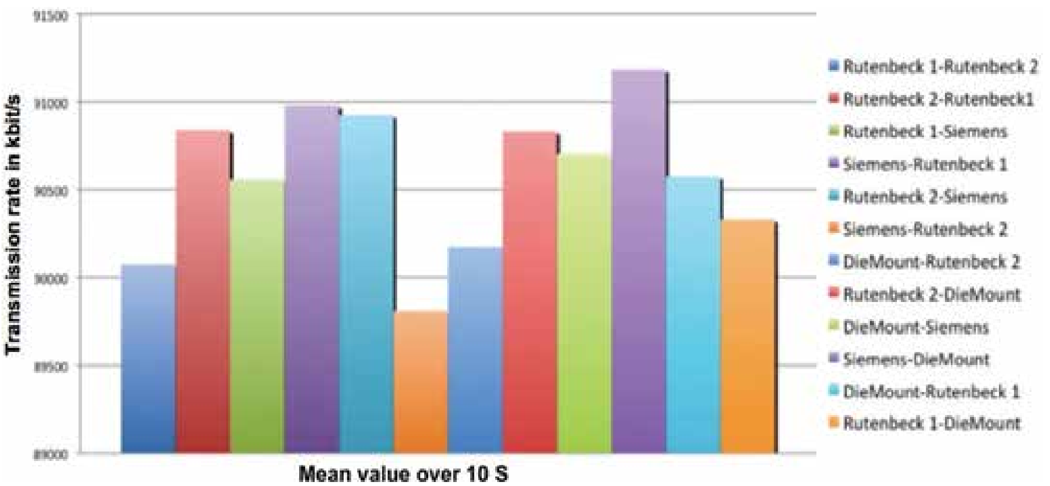

Figure 9. Measured effective data rates with 15 m POF length between different manufacturers.

In addition, the Speedport OptoLAN from Siemens is able to achieve 50 m transmission distances in almost all combinations. In general, it should be mentioned that the data rates reported by the manufacturers of 100 Mbit/s were not achieved by any system. On the other hand, all manufacturers did not provide a minimum data rate to be compared with the measurement results. The POF switch from BSPCOM from China shows the most stable transmission in all combinations and transmission lengths.

2.4.1. Measurement errors in Jperf

During the use of Jperf as a tool for recording the data rate, some points must be noted. On the one hand, higher transmission data values are always detected when the duration of the measurement is set to longer sampling values. This can be explained by the fact that the measuring interval is longer during a longer measuring period than in the case of a shorter measuring duration. Consequently, a mean value formation takes place. The reason is that the output format of Jperf of Mbit/s calculates large rounding errors. In addition, there is an error in the calculation of the average bandwidth over the whole measurement period by recalculation in Excel. The mean bandwidth was always larger than calculated externally. Therefore, the external calculated values are used in the evaluation.

3. WDM over POF

3.1. WDM over POF basics

At present, the great potential of the POF is not available as the alternative techniques offer transmission rates up to 10 Gbit/s over copper and up to 40 Gbit/s over glass fibers in the network area. The WDM technique offers an approach to achieve these high data rates also in the POF range. A sketch of the basic principle is shown in **Figure 10**.

Wavelength division multiplex systems need two basic components of a multiplexer and a demultiplexer (see **Figure 10**). To realize a working DEMUX for POF, several preconditions must be fulfilled. The basic component is a mirror, which focuses a divergent light beam coming from the input fiber. The shape of this mirror must be a toric shape to prevent spherical aberrations [9–11].

To separate the different incoming wavelength channels, a diffraction grating is used. This principle is illustrated in **Figure 11**. The light is split into different orders of diffraction. The first order is the important one to regain all information. There, the outgoing fibers with the different wavelengths channels must be arranged.

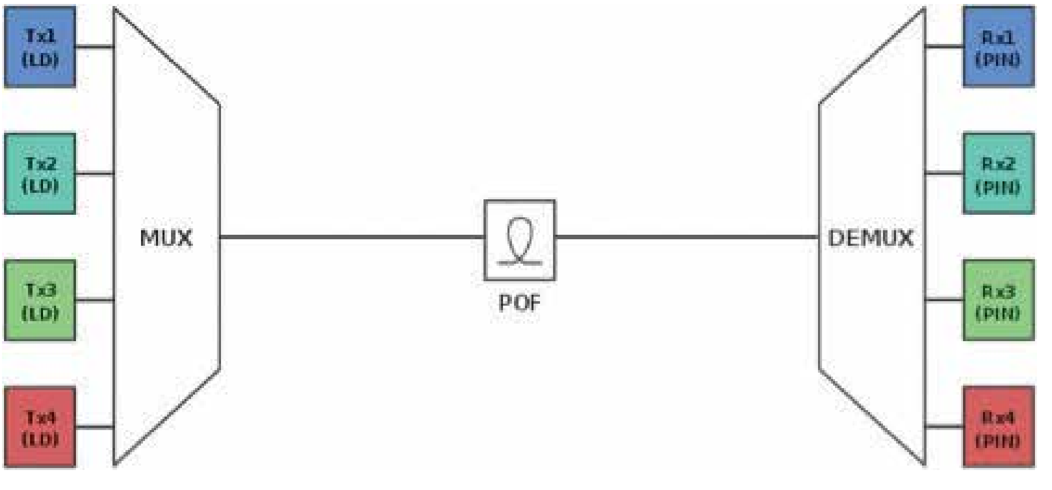

Figure 10. Schematic of the WDM over POF structure.

http://dx.doi.org/10.5772/intechopen.72204

Figure 11. Rowland setup of demultiplexer.

The development of the injection-molding process starts with the production of a master for the imprint of the entire component. This master is milled in micrometer precision by means of a diamond cutting process and created by the diamond turning process. Here, the PMMA material is processed directly. Both the moldings as well as the grid for wavelength separation can be made using this technique (see **Figure 12**). The last step is performed to validate the simulation results with the produced component.

For the injection-molding process, the production of the impression part is the most important factor. Due to the three-dimensional toric structure of the grating planar manufacturing methods like lithography, especially LIGA [a German acronym for Lithographie, Galvanoformung, Abformung (Lithography, Electroplating and Molding)] cannot be used. LIGA is used to manufacture planar spectrometers based on the glass fiber technology [12–15]. In the present approach for using a grating as a WDM element, it is necessary to manufacture the three-dimensional grating with its fine line structure and blaze precisely. In particular, the micro-structure of the grating and the exact shape of the toric surface require high precision. The

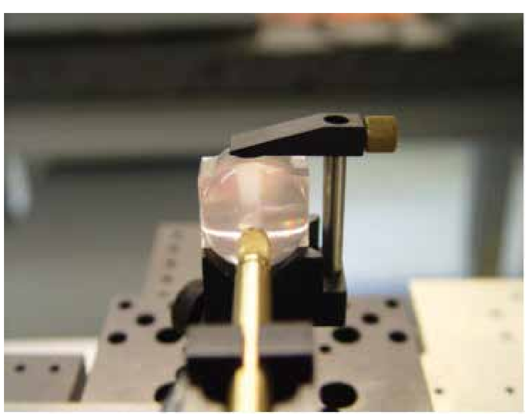

Figure 12. Integrated demultiplexer prototype.

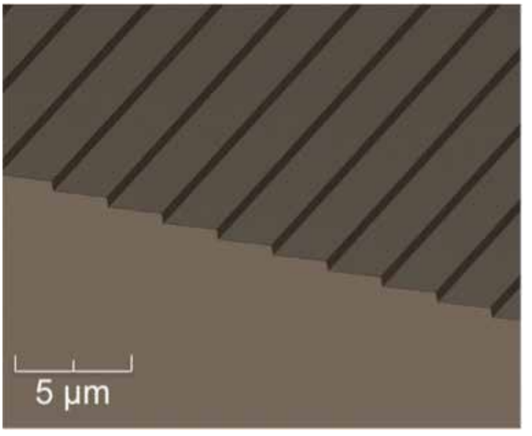

Figure 13. Grating of the demultiplexer.

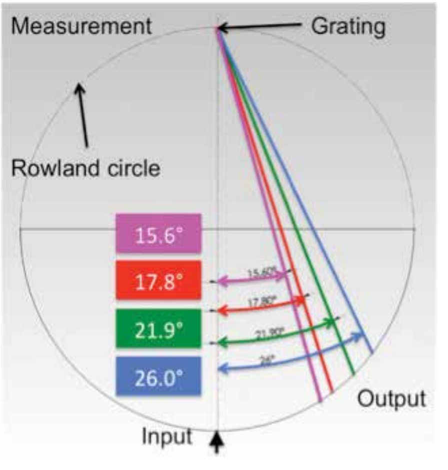

Figure 14. Measurement results of the focal points for different wavelengths (405 nm 15.6°, 450 nm 17.8°, 520 nm 21.9° and 650 nm 26.0°).

blaze with the grating lines is a microstructure in the form of a sawtooth with a distance between the teeth of 2.5 μm. **Figure 13** shows an enlarged 3D model of the grating. After analysis of other microtechnical machining processes to our knowledge, only the diamond turning meets the stringent requirements of the microstructured grating (**Figure 14**).

3.2. Design of the first demonstrator

The DEMUX elements must be manufactured in injection-molding technology. The capability of injection-molding technology for the cost-effective mass production of large volume and micrometer-accurate plastic components has made this technology the industrial standard production method for plastic parts. More and more high-quality optical components

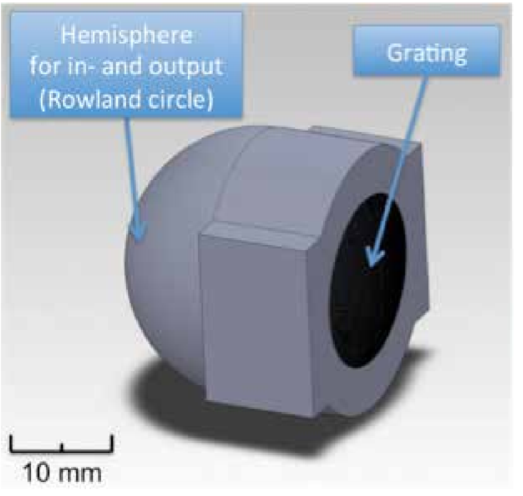

Figure 15. 3D model of the DEMUX demonstrator.

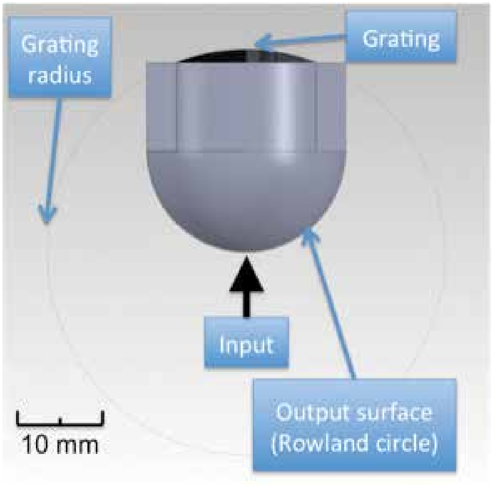

Figure 16. Cross-section view of the DEMUX demonstrator.

are produced with injection molding. With the aid of the injection molding, dimensionally stable and stress-free molded parts can be produced. In particular, the reduction of internal mechanical stress makes this technique ideal for optical mold components [12]. With this cost-effective production, the components for WDM can be made available via POF for a very wide application market. To further reduce the production costs, a self-adjustment of the individual optical components of the DEMUX such as fiber, grating, focusing mirror is necessary. This is why the various functions are combined in a molded part. However, this makes DEMUX technologically more difficult to implement, and therefore the individual process steps are discussed in detail.

In the first step, a demonstrator was produced. In order to verify the concept of the demultiplexer and to compare the simulation results with the real setup, it is necessary to proceed step by step. For this reason, a special optomechanical design was chosen. **Figure 15** shows the new design where a hemisphere at the output of the DEMUX represents the radius of the Rowland circle. This is shown in the cross-sectional view in **Figure 16**. The light reflected from the grating and emanating from the circle is focused on this radius. Therefore, the light is coupled into the center of the hemisphere, and the separated wavelengths can be detected on the surface of the hemispheres. This is illustrated in **Figure 14**. For detection, scanning of the surface is performed to determine the positions of the outgoing, separated light for each wavelength.

4. Materials and methods

Prior to the production of the DEMUX, some preliminary investigations have taken place to find the best suitable material for the demultiplexer. Therefore, both the processability of the material and the optical parameters had to be considered in detail. The injection-molding process was tested with a thick-walled mold. This test tool had the same shape as the final DEMUX, except for the grid. The test runs were carried out with an injection-molding machine from Babyplast 6-10P. This device was able to inject precisely small parts. **Table 1** lists all the materials used for the study. Further, parameters such as the respective melt volume rate (MVR) and light transmittance (according to the manufacturer's specification) are depicted. The test was additionally used to find the optimized injection-molding process parameters for the material.

In addition, the optical quality of the polymer materials must be investigated. Therefore, a mold for injection-molding test plates was designed. The test plates had a thickness of 2 mm. The mold is used to make samples from each material listed in **Table 2**. The DIN EN ISO 13468–2 standard describes the measurement of the optical transmission of polymer materials. Therefore, the test plates are designed to meet this standard.

Transmission measurements were carried out with all test plates. The results are shown for 405 nm in **Figure 17**. It can be seen that both ZEONEX types and PMMA POQ62 show the highest value for the light transmission. PMMA POQ62 is a polymer grade with high purity of polymer granulates. The measurement is made at a wavelength of 405 nm because it is one of the wavelengths used for the WDM system.

4.1. Manufacturing of the demonstrator

By using the injection-molding process, the manufacturing of the mold insert is the most important factor. Due to the three-dimensional toric structure of the grating planar manufacturing methods like lithography, especially LIGA cannot be used [15].

In our case, however, the three-dimensional grating requires a different processing method. The microstructure of the grating and the exact shape of the toric surface require a particularly

Name	Type	MVR [cm³/10 min]	Transmission [%]
Plexiglas 6 N	PMMA	12	92
Plexiglas POQ62	PMMA	21	92
Topas 5013 L-10	COC	48	91.4
Topas 6013 M-07	COC	14	91
ZEONEX F52R	COP	22	92
ZEONEX 350R	COP	26	92
Makrolon LED2245	PC	35	90

Table 2. Injection-molding materials for MUX/DEMUX-element.

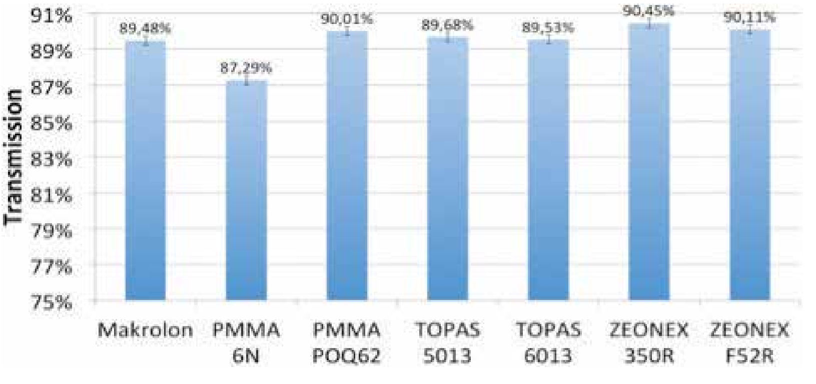

Figure 17. Transmission of different material at 405 nm.

high precision of manufacture. The microstructure has the shape of a sawtooth with a distance between the teeth of 2.5 μm. **Figure 13** shows an enlarged 3D model of the grating. An in-depth investigation of various processing methods has shown that only the diamond turning fulfills the high requirements of the production of the microstructural lattice. The diamond twisting technique is a special machining method using a single crystal diamond cutting tool. It is also possible to produce a surface with an optical quality at the edge of the optical component. It offers several advantages:

- True three-dimensional contour generation.
- Accuracy of one part in 106 with absolute accuracy of 1 part in 108 on a single axis for ideal conditions.
- Surface finish of 5 nm Rz for a range of materials and as good as 1 nm Ra.
- Ability to generate surfaces with variable aspect ratios and
- Feature sizes that exceed the limits of optical microscopy [14, 15].

A metallization process was used to analyze the surface of the lattice. The surface was sputtered with a thin aluminum layer depicted in **Figure 18**. It is now possible to measure the shape of the surface with a white light interferometer and to examine the lattice structure under the scanning electron microscope (SEM). The metallized surface of the grating is shown in **Figure 18**. It can be seen that the structure on the left side has a dull and mat surface instead of the glossy residue of the surface. This is a first indication that the surface roughness in this part is higher and does not meet the requirements for the component precision. The first visual impression was then confirmed by the analysis under the SEM.

The cause of the degradation of the grating quality is the change in the strain as the milling tool passes the highest point in the center of the surface. It changes the way the force is exerted by a pushing movement on the surface. This results in a coarse structure on the other half of the surface. From the measurement of the structure size in **Figure 19**, a width of 2.55 μm can be determined, which is within the tolerances of the reference of 2.5 μm.

In addition to the structural quality, the dimensions of the surface are also important for the functionality of the DEMUX and must be considered in detail. The shape of the radius of the

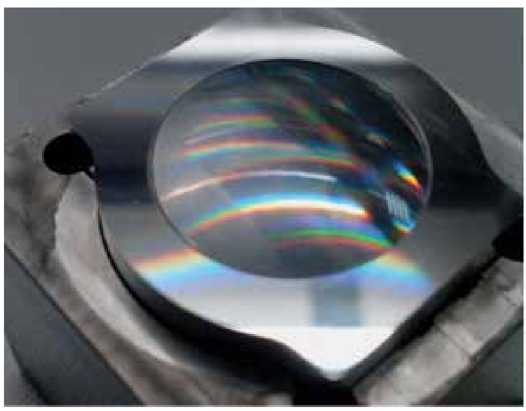

Figure 18. High-quality structures of the grating.

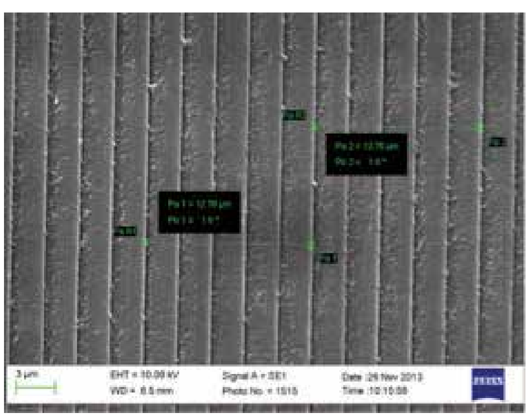

Figure 19. Metallized grating surface of the DEMUX.

toric surface was designed to focus the colored light beams on the Rowland circle. Therefore, it was analyzed using a white light interferometer (FRT MicroProf). The cross-section of the toric surface was measured. The measurement shows that the dimensions of the surface correspond to the tolerances of the DEMUX. An exception can be seen in the diameter in the x-axis, which is somewhat out of tolerance (**Table 3**).

Several parameters have to be optimized in order to correct the manufacturing errors. This is performed in several iterations in close cooperation with the manufacturer. For example, the adjustment of the force applied to the surface was varied and optimized by the diamond tip. The next part with optimized parameters is now in production and is then analyzed in the same way to check the adjustments of the parameters.

4.2. Optical measurements

In order to measure the position of the focal points of the different separated wavelengths on the Rowland circuit, a special measurement setup was chosen. It uses a parallel kinematic precision alignment system to align a POF on the surface of the hemispheres. An input fiber firmly bonded to index matching is used to couple white light into the DEMUX, as shown in **Figure 16**. In this figure, it can be seen that the separated wavelengths are focused on a ring on the hemisphere. This ring is scanned by the fiber on the alignment system. The light from the scanning fiber is analyzed by using a spectrometer.

From the spectra along the Rowland ring, the location of the maxima of the wavelengths is determined. For the first component, the entire measurement was performed and compared with the simulation results. Four different wavelengths that were used to analyze the wavelength separation are as follows: 405, 450, 520 and 650 nm.

The positions of the wavelengths measured by the setup are also depicted in **Figure 14**. In comparison to the simulation, a shift of the positions of 2,3° are found. Nevertheless, the separation of the wavelengths was measured and confirmed the functionality of the demultiplexer.

The derivations to the simulation could be caused through the following reasons:

- Derivation of the blaze angle of the sawtooth grating
- Inhomogeneous structure of the grating
- Manufacturing tolerances

These depend strongly on the precision of the manufacturing process. As mentioned in the previous section, the production of such complex structures on a toric surface is a major

Dimension	Measurement (mm)	Reference (mm)
Diameter x-axis	15.869	>16.000 ± 0.1
Diameter y-axis	15.887	>15.170 ± 0.1
Height of grating	1.862	1.872 ± 0.05

Table 3. Measurement results of the DEMUX dimensions.

challenge. Therefore, the process parameters must be improved and optimized to fully meet the optical requirements of the demultiplexer component.

5. Spectral grids in the visible spectrum for POF WDM applications

Besides developing low-IL cost-effective POF WDM components and fast POF WDM transmission systems, it is also important to allocate a unique set of WDM transmission channels in the visible spectrum to support WDM applications over SI-POF. To evaluate the applicability of a spectral grid to support visible spectrum WDM applications over SI-POF, the appropriate criteria were first established. Those criteria refer to:

- Channel distribution with respect to the spectral attenuation of SI-POF;
- Performances of different demultiplexing techniques;
- Availability of laser diodes in the visible spectrum.

5.1. Extension of ITU-T G.694.2 CWDM grid into the visible spectrum

If ITU-T G.694.2 CWDM wavelength grid would be extended into the visible spectrum, 15 equidistant channels between 400 and 700 nm would be obtained, as shown in **Figure 14**. The parameters of the grid including the nominal central wavelengths are depicted with arrows in **Figure 20**.

The channel spacing of 20 nm makes good utilization of the available spectral range. In the red window, the extension has a channel at 651 nm, which is very close to the attenuation minimum at 650 nm. The channel distribution also corresponds well to three other attenuation windows. The channels experiencing the highest attenuation are those at 611, 631, 671

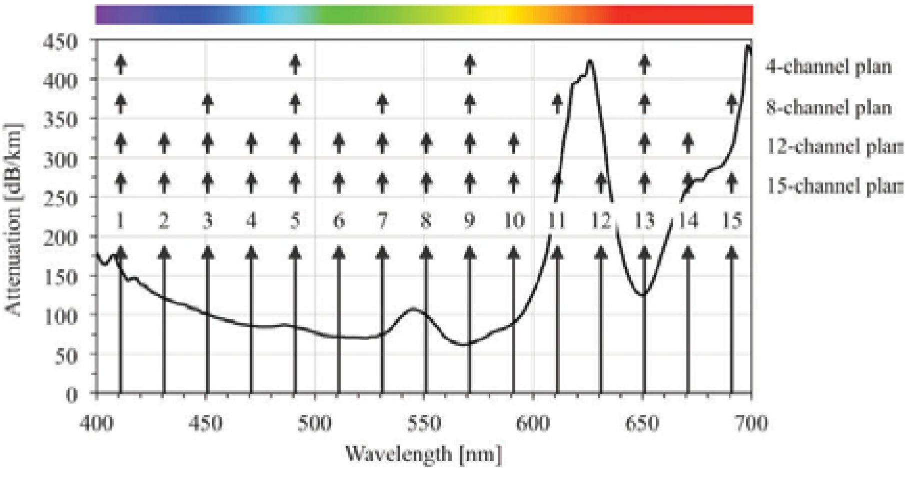

Figure 20. Extension of CWDM wavelength grid into the visible spectrum and channel plans for 4-, 8-, 12- and 15-channel applications.

and 691 nm. Those channels could be used for distances up to 20 m since they would experience approximately the same attenuation as 651 nm channel over 50 m, but lower intermodal dispersion. Good channel allocation, sufficient channel spacing, high channel count and good availability of the transmitters make the extension of CWDM grid very suitable to support WDM applications over SI-POF.

6. Conclusions

Currently, commercially available POF transmission systems are able to fulfill the needs of IEEE 802.3 requirements with a data rate of 100 Mbit/s. The interoperability between the devices of the different manufacturers is also in a good condition.

To realize higher bitrates WDM over POF will be an interesting application. We produced a DEMUX by injection molding.

The realization of this DEMUX element for POF presents several challenges, in particular the microstructure of the grating on the three-dimensional surface. It is shown that it is possible to realize the structure size and the exact radius for the DEMUX with the current optimized production process. The high challenge of producing the blazed grating leads to some errors in the milling process, which still needs to be improved. This will be done in the future by optimizing the process parameters. The next parts will be produced and analyzed with the optimized parameters.

This hopeful result shows that WDM applications over SI-POF with high Gbit/s transmission are a realistic aim for the next future. The technique will be able to extend the bandwidth in POF systems strongly. It seems to be possible to transmit 40 Gbit/s via 15 channels and a channel rate of 2,7 Gbit/s data rate with WDM over POF. This opens the range of POF applications to existing cloud centers and future in-house networks with to link length up to 100 m.

Acknowledgements

We gratefully acknowledge the funding by the German Ministry of Education and Research (BMBF) under grant number 16 V0009 (HS Harz) /16 V0010 (TU BS). All injection molded parts are done with the support of the Institute of Micro and Sensor Systems at the Otto-von-Guericke University Magdeburg and Prof. Bertram Schmidt.

Author details

Ulrich H.P. Fischer-Hirchert*, Matthias Haupt, Mladen Joncic, Stefanie Haupt and Sebastian Höll

*Address all correspondence to: ufischerhirchert@hs-harz.de

Harz University of Applied Sciences, Wernigerode, Germany

References

[1] Daum W, Krauser J, Zamzow PE, Ziemann O. POF Handbook: Optical Short Range Transmission Systems. Berlin, Heidelberg, New York: Springer-Verlag; 2008

[2] Nalwa HS, editor. Polymer Optical Fibres. California: American Scientific Publishers; 2004

[3] Club des Fibres Optiques Plastiques (CFOP) France. In: Marcou J, editor. Plastic Optical Fibres—Practical Applications. Masson: John Wiley & Sons; 1997

[4] Brandrup J, Immergut EH, Grulke EA. Polymer Handbook. 4th ed. Wiley-Interscience; 1999

[5] Chen RT, Lipscomb GF, editors. WDM and photonic switching devices for network applications. Proceedings of SPIE, vol. 3949, 2000

[6] Colachino J. Mux/DeMux optical specifications and measurements. In: White Paper. Lightreading, The Woodland Texas: Lightchip Inc.; 2001

[7] Gnauck AH, Chraplyvy AR, Tkach RW, Zyskind JL, Sulhoff JW, Lucero AJ, et al. One terabit/s transmission experiment. Proceedings OFC'96; 1996

[8] www.nwlab.net/know-how/JPerf/, Website 09-01-2017

[9] Fischer-Hirchert UHP. Photonic Packaging Sourcebook: Fiber-Chip Coupling for Optical Components, Basic Calculations, Modules; 2015

[10] Fischer UHP, Haupt M. WDM over POF: The inexpensive way to breakthrough the limitation of bandwidth of standard POF communication. SPIE Symposium on Integrated Optoelectronic Devices, Photonics; Bellingham WA: West San Jose; 2007

[11] Fischer UHP, Haupt M. Integrated WDM System for POF Communication with Low Cost Injection Moulded Key Components. Access Networks and In-house Communications; 2010

[12] Stricker M, Pillwein G, Giessauf J. Focus on precision—Injection molding optical components. Kunststoffe International. 2009;**4**:15-19

[13] Ferguson JP, Schoenfelder S. Micromoulded spectrometers produced by the Liga process. Searching for Information: Artificial Intelligence and Information Retrieval Approaches, IEE Two-day Seminar (Ref. No. 1999/199); 1999. pp. 11/1-11/4

[14] Davies MA, Evans CJ, Vohra RR, Bergner BC, Patterson SR. Application of precision diamond machining to the manufacture of microphotonics components. Proc. SPIE 5183, Lithographic and Micromachining Techniques for Optical Component Fabrication II, 94; November 2003

[15] Dornfeld D, Min S, Takeuchi Y. Recent advances in mechanical micromachining. CIRP Annals: Manufacturing Technology. 2006;**55**(2):7

Chapter 12

Content Defined Optical Network

Hui Yang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72432

Abstract

Optical interconnection has become one of the key technologies to adapt the needs of large-scale data center networking with the advantages of large capacity, high bandwidth, and high efficiency. Data center optical interconnection has the characteristics of resource and technology heterogeneity. Its networking and control face enormous challenges for the increasing number of users with a high level quality of service requirements. Around different scenarios, there are a series of key networking and control problems in data center optical interconnection, such as multiple layers and stratums resources optimization in inter-data center, and time-aware resource scheduling in intra-data center. To solve these problems and challenges, this chapter mainly researches on content defined optical networking and integrated control for data center. For networking of vertical "multi-layer-carried" and horizontal "heterogeneous-cross-stratum", the chapter launches research work around application scenarios about inter-data center optical interconnection with optical network, and intra-data center. The model architecture, implementation mechanism and control strategy are analyzed and demonstrated on the experiment and simulation platform of data center optical interconnection. This chapter will provide important references for future diverse applications of data center optical interconnection and software defined networking and control in practice.

Keywords: software defined optical network, content, data center, optical interconnect, OpenFlow

1. Introduction

With the rapid development of cloud computing and high rate services, data center services have attracted a great deal of attention from network service providers. With the variety and massiveness of applications, the high-performance network-based datacenter applications have the features of high burstiness and wide-bandwidth, particularly for the super-wavelength services [1]. Flexi-grid optical networking with big capacity, low power density and distance adaption, provides a promising solution for the new datacenter network bottlenecks. A novel optical orthogonal frequency division multiplexing (OFDM)-based architecture with high spectral efficiency and high energy efficiency is presented for data center networks [2]. Networking architecture, algorithm and control plane for inter-datacenter network is also addressed in flexi-grid optical networks [3]. Lower power, improved scalability and port density which is the advantages of software defined optical networking for highly virtualized datacenters are studied [4].

Compared with optical interconnect networks between datacenters, optical interconnect networks in a datacenter is a more imperative requirement and respective case to serve the services in a flexible and high-efficient way [5]. Besides, miscellaneous datacenter services have a lower delay and higher availability requirements whose quality of service (QoS) can be guaranteed in corresponding levels [6]. Many studies focus on the architecture and equipment in datacenter interconnection [7–9]. For instance in Ref. [7], arrayed waveguide grating router (AWGR)-based interconnect architecture is proposed. A distributed all-optical control plane is designed with low latency and high-throughput at high traffic load in the case of sufficient packet transmission time. The authors in [8] achieve the hitless adaptation between Ethernet and time shared optical network (TSON) by designing a novel network on-and-off chip approach for highly efficient and transparent intra-datacenter communications. The work [9] make the datacenter offloads heavy inter-pod traffic onto an optical multi-ring burst network by proposing efficient scheme to all-optically inter-networking the pods. However, the time feature of application to guarantee the services delivery with various QoS in intra-datacenter networks from the view of service is relatively unexplored. Recently, as a centralized software control architecture, the software defined networking (SDN) enabled by OpenFlow protocol has become a focus of study by making the network functions and protocols programmable [10–13], where maximum flexibility is provided for the network operators and the integrated optimization of services can be achieved in a centralized control over multi-dimensional resource [14–17]. Hence, to introduce SDN method to centrally control network and application resources in optical interconnect of intra-datacenter has a great significance.

In our previous study, the network architecture with cross stratum optimization (CSO) based on SDN including muti-stratums resources in inter-datacenter networks has been designed to partially satisfy the QoS requirement [18–21]. Even in the edge of the network, the similar architectures with CSO based on SDN have been studied for improving the performance of cloud-based radio access network, which is similar as the inter-datacenter networks [22–24]. Based on the previous work, this paper proposes a Content Defined Optical Network (CDON) architecture in OpenFlow-based datacenter optical networks for service migration, in which a time-aware service scheduling (TaSS) strategy is introduced. CDON considers the time factor, in which the applications with required QoS can be arranged and accommodated to enhance the responsiveness to quickly provide for datacenter demand. By the experimental implementation on our testbed with OpenFlow-based intra-datacenter and inter-datacenter optical networks and the statistics collection of blocking probability and resource occupation rate, the overall feasibility and efficiency of the proposed architecture are verified. Intra-datacenter and inter-datacenter networks are considered in this paper. Based on the unified and flexible control advantages, SDN is deployed in both two networks.

The rest of this paper is organized as follows. In Section II, we propose the novel datacenter-network-based CDON and builds functional models. Then intra-datacenter optical interconnection architecture and inter-datacenter optical network architecture are designed. Section III describe the TaSS strategy. Finally, we describe the testbed and present the experimental results and analysis in section IV and section V conclude the paper.

2. Datacenter-network-CDON

In order to promote the control efficiency of datacenter networks, control architecture based on CDON is described as shown in **Figure 1**. Different OpenFlow controllers for different resources including intra-datacenter computing recourse and inter-datacenter communication resources have been developed. The latter is mainly flexi-grid optical network resource in this work. All resources are software-defined with OpenFlow and support datacenter application. Then OpenFlow-enabled network controller and OpenFlow-enabled application controller can work together. By using user and application interface (UAI), application plane which is served through application and controller interface (ANI) can provide users with various services. The architecture of intra-datacenter and inter-datacenter networks is discussed in detail as follows.

2.1. Intra-datacenter optical interconnection architecture

The CDON architecture for OpenFlow-based intra-datacenter optical interconnect is shown in **Figure 2(a)**. Top-of-rack (ToR), aggregation and core optical switches three kinds of optical switches are used to interconnect datacenter servers with the deployment of application stratum resources (e.g., CPU and storage). Application controller (AC) and network controller (NC) respectively centrally control each stratum resources which are software defined with OpenFlow. To control intra-datacenter networks for service migration with

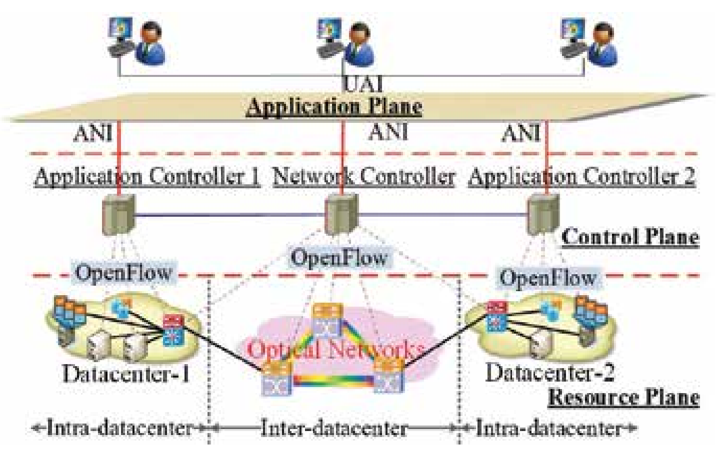

Figure 1. Datacentre-network-based CDON.

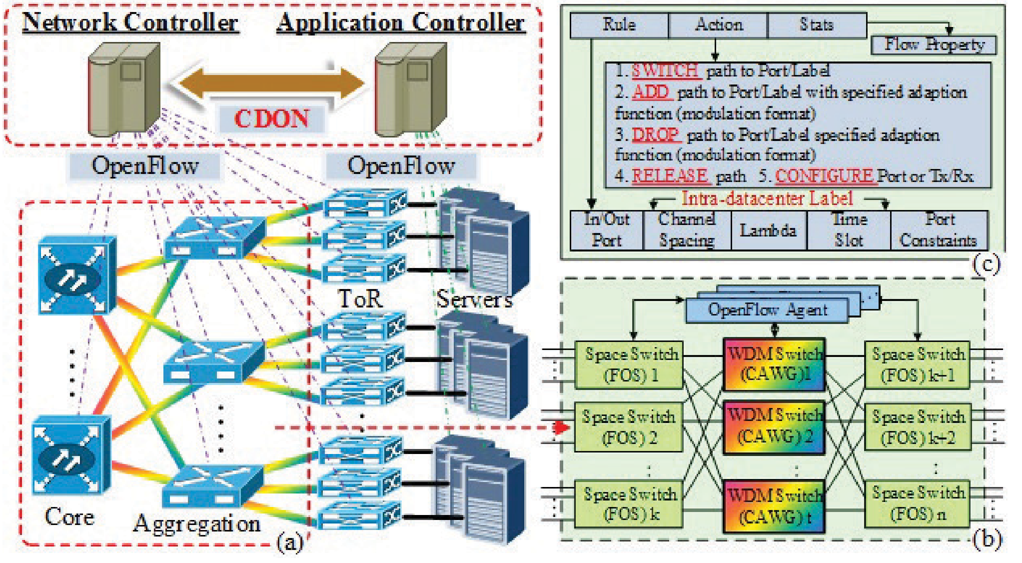

Figure 2. (a) The SDN-based intra-datacenter network architecture of CDON. (b) The SDN-based switching fabric of intra-datacenter. (c) The extension of protocol for CDON.

extended OpenFlow protocol (OFP), OpenFlow-enabled optical switches with the OFP agent software (OF-OS) is used. The CDON architecture in intra-datacenter networks have twofold motivations. Firstly, CDON can highlight the cooperation between AC and NC for supporting TaSS strategy to schedule datacenter applications based on different time sensitivity requirements reasonably and optimize application and network stratum resources efficiency. Secondly, considering the burstiness, burst applications fast provisioning with unified CDON control and process can be supported.

Figure 2(b) describes the OpenFlow-switching structure of intra-datacenter. The functions and interaction descriptions of relevant functional modules are described below. The AC is responsible for monitoring and maintaining the resources of application stratum for CDON, NC supports abstraction of the network information from physical stratum and lightpath provisioning of optical networks in intra-datacenter. When request arrives, AC processes it by using TaSS strategy considering the requirements of delay sensitivity and achieving CSO of the computing and storage in an internal database and sent the decision to NC. The control of virtual and physical network are concerned in NC. The former controls the virtual network and send abstracted network information to AC, the responsibilities of latter is to monitor and control the programmable physical modules. With the extended OFP, the lightpath can be built according to the request from AC. It is noteworthy that burst traffic can be correspondingly serviced by the high level optical switch by fast tunable laser (FTL) and burst mode receiver (BMR) in ToR switch. The aggregation and core switching fabrics are built based on space and wavelength circuit-switching technologies, fast optical switch (FOS) and cyclic arrayed waveguide grating (CAWG) respectively. The OFP agent software embedded in optical module has four functions which are maintaining flow table, modeling the information of node with programmability, mapping the content to configure and controlling hardware.

In connection to networks control of intra-datacenter, **Figure 2(c)** shows extended flow entry of OFP. The rule is added with the main characteristics of intra-datacenter including the in/out port, intra-datacenter label (e.g., channel space, lambda and time slot) and port constraints The action is extended as five types: add, switch, drop and configure to set up a path, and release a path. Using combinations of rule and action, the control of optical node is realized. The responsibility of stats function is monitoring the flow property to provide service provisioning for CDON.

2.2. Optical network architecture of inter-datacenter

Flexi-grid optical networks are the promising technology for inter-datacenter networks as these networks can satisfy the requirement of burstiness. The CDON architecture is built and shown in **Figure 3(a)**. The distributed datacenters are interconnected through the flexi-grid optical networks. The network architecture mainly consists of the optical resources stratum and application resources stratum. In a unified manner, network controller and application controller can control each resource stratum which is software defined with OpenFlow. Software defined OTN (SD-OTN) is necessary to control the flexi-grid optical networks for inter-datacenter network with extended OFP. SD-OTNs are essentially OpenFlow-enabled elastic optical device nodes with OFP agent software. It has twofold motivations to design the CDON architecture over inter-datacenter optical network. Firstly, the CDON can realize the global interworking of cross stratum resources that the physical

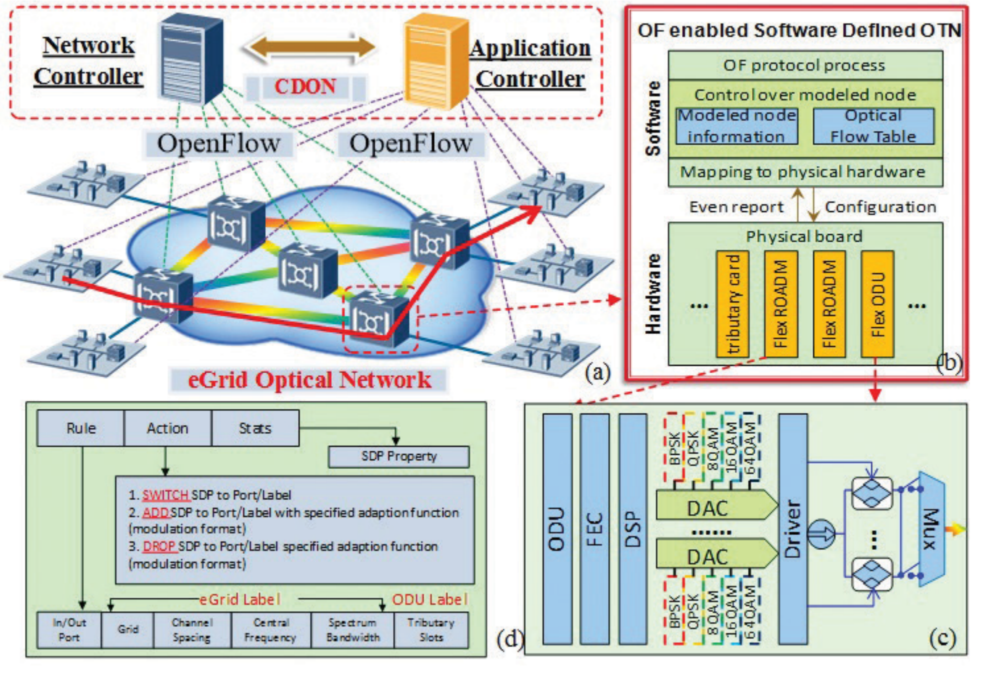

Figure 3. (a) The architecture of CDON in OpenFlow-based inter-datacenter networks. (b) OF-enabled SD-OTN functional models. (c) Flex ODU. (d) The extension of protocol for CDON in inter-datacenter.

layer parameter (e.g., bandwidth and modulation format) can be adjusted. So the cooperation between AC and NC is emphasized to realize software defined path (SDP) with application and spectrum elasticity. Secondly, the different time sensitivity requirements of services can be considered reasonably through scheduling data center services with time elasticity to optimize the application and network resources utilization further. Based on functional architecture described above, TaSS scheme is proposed in the AC and it can arranging the start time, transport time and corresponding transport bandwidth for services for realizing the application and network stratums resources optimization.

The functional modules of AC and NC and the coupling relationship between different modules are shown as follow. NC is responsible for the control in physical and virtual network. The former includes controlling spectrum resource and modulation format. The latter is responsible for managing the virtual network and sending virtual resource information to AC. While a request arrives, AC runs TaSS strategy based on muti-stratum resource information and sent the decision to NC through application-transport interface (ATI). The SDP can be found out and built up based on extended OFP according to the request from AC. It is noteworthy that, the length of SDP decides the modulation format of service (e.g., QPSK and 16QAM). In the case of short distance, spectrum bandwidth which is more precious than other resource can be economized by using high-level modulation format. In SD-OTN, OpenFlow-enabled agent software is embedded to realize the communication between NC and optical node. The SD-OTN maintains optical flow table and modeled node information as software. The physical hardware which contains flexible ROADM and ODU boards shown in **Figure 3(b)** and **(c)** respectively is configured and controlled through the content mapping. In the side of the control of flexi-grid optical networks in inter-datacenter, flow entry of OFP is extended and shown in **Figure 3(d)**. In this architecture, the rule is extended as the main characteristics of flexi-grid optical networks which including the in/out port, flexi-grid label (e.g., central frequency and spectrum bandwidth) and ODU label (e.g., tributary slots). The action of optical node mainly includes three types: add, switch and drop. Through combining rule and action, the control of flexi-grid node can be realized. The responsibility of stats function is monitoring the flow property to provide SDP provisioning.

3. Service scheduling strategy

In the side of the service accommodation of datacenter optical networks, the traditional strategy can.

allocate the optimal datacenter server application and corresponding lightpath network resources when a request arrives. That may satisfy the following conditions, which are shown in **Figures 4** and **5**. We assume that two datacenter servers are deployed in the candidate destination node. The storage utilization of server #1 and server #2 are 80% and 85% respectively when service arrives. After a relatively short time, the services which are provided by the two servers have changed, i.e., some services are released when they are complete and new ones arrive. It leads that the storage utilization of server #1 increases to 95%, while that in server #2 is down to 25%. The server #1 (80% at arriving time) would be chosen as the destination

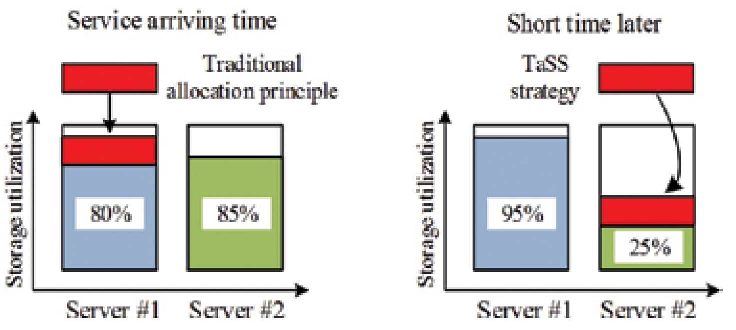

Figure 4. Illustration of time-aware datacenter application resource allocation.

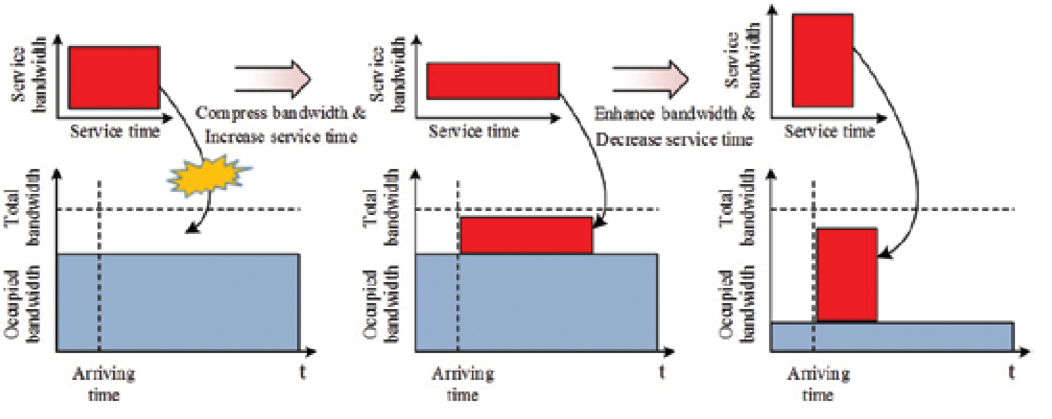

Figure 5. Illustration of time-aware network resource allocation.

node When the traditional allocation principle is used. However, if the resource is allocated after a short time (within delay tolerance of user), the better optimization and more effective resource utilization may be realized. The other instance is about the service bandwidth. The bandwidth which most of datacenter services use is specified and the time to transmit service is fixed for achieving the overall data volume of service. If the bandwidth is fixed with traditional methods, the service cannot be provided with enough resource in case of less available bandwidth. With compressing the bandwidth to adapt the available bandwidth and increasing the service time (overall data volume of service is constant), the allocation scheme can be feasible. Similarly, by enhancing the provide bandwidth, the service can be completed as soon as possible with relatively abundant resources. The time factor of datacenter service is considered in the two issues. Therefore, we propose the time-aware service scheduling strategy.

3.1. Network modeling

G (***V***, ***L***, ***F***, ***A***) denotes the OpenFlow-based datacenter interconnect with optical networks, where $\boldsymbol{V} = \{v_1, v_2, \ldots, v_n\}$ is represented as the set of SDN-based optical switching nodes, $\boldsymbol{L} = \{l_1, l_2, \ldots, l_n\}$ denotes the set of optical fiber links connecting nodes in ***V***. $\boldsymbol{F} = \{\omega_1, \omega_2, \ldots, \omega_F\}$

indicate the set of wavelengths of optical fiber and ***A*** is the set of servers in datacenter. While source node *s sends* a service request, the request contains total data volume of service *D* and storage space *S*. The services are classified with latency-sensitive and delay-tolerant service. The former requires immediate service process and its *i*th request is denoted as $SR_i(s, D, S)$, and the latter incluces the arriving time t_c and tolerant delay *T*, and we denote its *i*th service request as $SR_i(s, D, S, t_c, T)$. The request SR_{i+1} will be the next request in time order when the connection demand SR_i arrives. In addition, **Table 1** shows some requisite notations and their definitions.

3.2. Time-aware service schedule strategy

Developing on the functional architecture, we present a novel time-aware service scheduling (TaSS) strategy which is implemented in application controller to schedule datacenter service with time sensitivity requirement. For the requirements of the arriving service, we classify them as burst latency-sensitive and delay-tolerant service including flow volume and tolerant latency, based on the delay sensitivity of each service. For the delay-tolerant service, we search servers and lightpaths of datacenter to judge whether the volume of S_r, B_r and *t is enough* to satisfy the incoming service. If the $S_r \geq S$ and $B_r * t \geq D$, by comparing CSO factor [16], the minimum one would be chosen to provision. If the resources for accommodating is not enough, the service would wait until maximum tolerant delay arrives, i.e., $t = T - D/B_a - t_G$, which considers B_a and t_G. New lightpath will be allocated with CSO to accommodate the service. In the side of delay-sensitive requests, by searching the existing candidate with the same method, the available resources are decided. The path which has minimum propagation delay obtained from candidate paths can be prepared to provision. When an available built path is found, the TaSS.

Strategy builds new lightpath for the service immediately. In order to realize the quick response for service provisioning, service time is used to reckon the release lightpath procedure immediately. The flowchart of TaSS strategy is shown in **Figure 6**.

Symbol	Definitions
$\cdot S$	The total capacity of storage space in target data center
$\cdot S_r$	The residual storage space in target data center
$\cdot D$	The total data volume of service
$\cdot B_r$	The product of available bandwidth of lightpath
$\cdot T$	The tolerant delay of delay sensitive service
$\cdot t$	The lightpath duration
$\cdot B_a$	The network average transmission bandwidth
$\cdot t_G$	The guard time
N	The number of network nodes
L	The number of links
F	The number of wavelengths
A	The number of datacenter nodes

Table 1. Symbols and definitions.

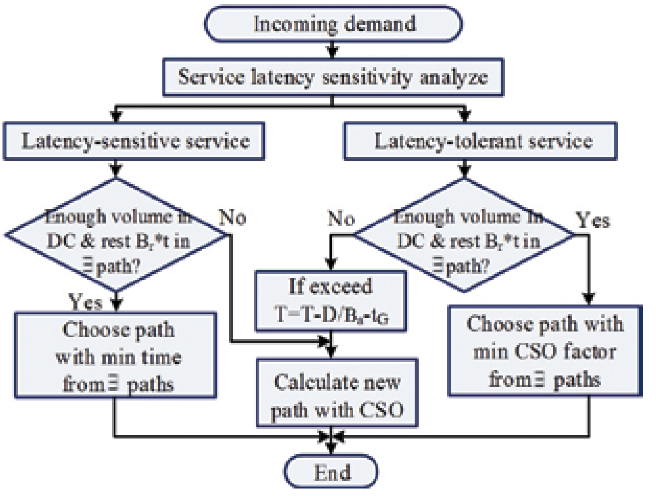

Figure 6. Flowchart of TaSS strategy.

4. Experimental demonstration and results discussion

A testbed of datacenter networks consisting of intra-datacenter and inter-datacenter is built. SDN is deployed on the two parts together. The following is a detailed description of the testbed demonstrations.

For experimentally evaluating CDON architecture, we deploy it into optical intra-datacenter networks and realize the service migration in this architecture on our testbed as shown in **Figure 7(a)**. In data plane, it contains 4 optical switches with FOS, Two CAWG cards in the core side, BMR embed into burst mode transceiver (BMT) card with FTL and the software OFP agent. The switching time of optical switches is 25 ns and the insert loss is lower than 4.5 dB. The frequency deviation of CAWG cards is 12.5GHz and the insert loss is 10.5 dB. The receiving power sensitivity of BMR is -25 dBm. The switching time of FTL has 98 ns and the frequency deviation is 2.5GHz according to ITU-T standard. The software OFP agent use PI to control the hardware through OFP. We use VMware software to build groups of virtual machines to realize Datacenters. Each virtual machine model a real node with the independent operation system, CPU and storage resource. In control plane, NC is realized with optical module control function, PCE computation function and resource abstraction function corresponding to three servers, database server is deployed to maintain transmission resources and the database of traffic engineering. AC server is used to carry TaSS strategy and monitoring the computing resources in servers. User plane is built in a server for running the required service.

We have designed experiment to verify the lightpath provisioning in CDON architecture for datacenter service migration. AC runs TaSS strategy to determine the path and schedule for service migration based on various application utilizations among datacenters and current network resource, then setup the path from source to destination node chosen by CSO. **Figure 7(b)-(d)** show the eye diagram and tuning waveform of FTL and spectrum of CAWG port reflected on the filter profile. The experimental results are further

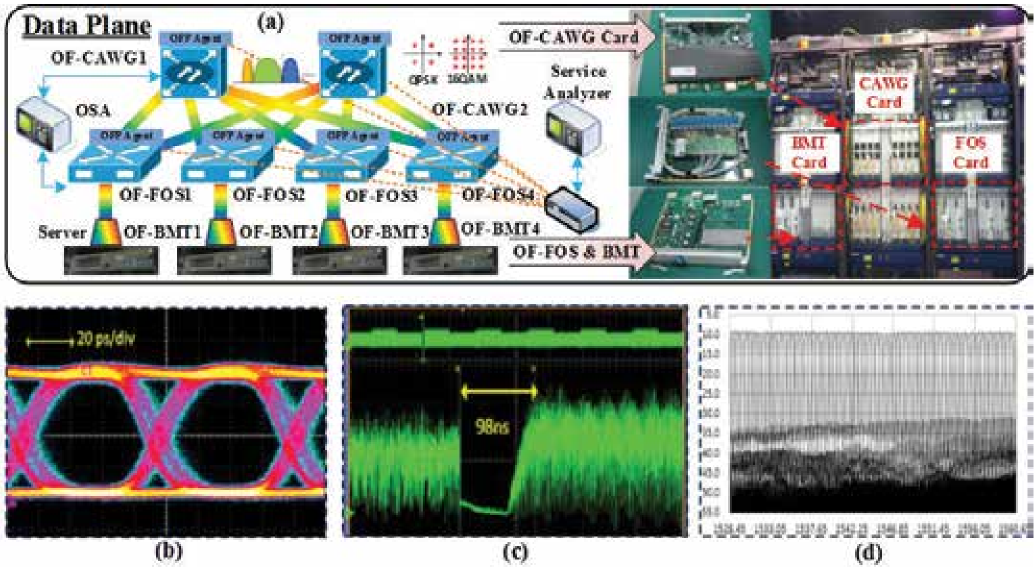

Figure 7. (a) Experimental testbed and demonstrator setup. (b, c) eye diagram and tuning waveform of FTL. (d) Spectrum of CAWG port.

shown in **Figure 8(a)** and **(b)**. **Figure 8(a)** demonstrates the OpenFlow message exchange sequence for CDON by capturing in NC with Wireshark. In **Figure 8**, 10.108.50.74 denotes AC and 10.108.65.249 denotes NC, OF-OS nodes are 10.108.50.21 and 10.108.51.22. AC performs TaSS strategy and NC control all OF-OS via flow mod message through the allocated lighpath to provide the path. Then the path releasing information is sent via flow mod message, and keep synchronization by updating the application usage with UDP to. **Figure 8(b)** verifies OPF extensions for CDON in intra-datacenter networks with a snapshot of the extended flow table modification message for lightpath provisioning. We also evaluate the performance of CDON in intra-datacenter networks under the condition of heavy traffic load through simulation by comparing TaSS with traditional CSO (TCSO) strategy [16]. The migration data volumes from datacenter node following a Poisson process are randomly from 50Gbit to 400Gbit. **Figure 9(a)** and **(b)** compare the performances of two strategies. Comparing to TCSO, TaSS can reduce blocking probability effectively, especially when the network is under heavy load. We can also find that TaSS outperforms TCSO in the resource occupation rate significantly. The reason is TaSS realizes global optimization considering the time schedule with various delay sensitivity requirements and adjusts the service bandwidth according to the distribution of network resources.

in order to evaluate the proposed architecture in experiments, we build CDON including both control and data planes based on testbed, as shown in **Figure 10(a)**. In data plane, SDN-based flexible optical nodes are deployed, each of which including flex ROADM and ODU boards.

Datacenters and the other nodes are also implemented on virtual machines created by VMware software on servers. As each virtual machine has independent operation system, IP address, CPU and memory resource, it can be regarded as a real node. The virtual OS technology can set up experiment topology including 14 nodes and 21 links which is similar to the backbone of US. For OpenFlow-based CDON control plane, the NC is set up to support the proposed architecture and deployed in three servers corresponding to elastic spectrum control, physical layer parameter adjustment, PCE computation and resource abstraction, while the database server work for maintaining traffic engineering database, management information base, connection status and the configuration of the database and transport resources. The AC server support CSO agent and monitoring the application resources from datacenter networks with TaSS strategy. User plane is set up in a server for running the required application.

We have designed experiment to verify SDPs provisioning in CDON over flexi-grid optical networks for datacenter service migration whose results are shown in **Figure 10(b)-(d)**. AC runs TaSS strategy to determine the destination datacenter based on various application utilizations among datacenters and current network resource, and then set up SDP for the service migration along the lighpath from source to destination node. Moreover, for different SDP distances, the spectrum bandwidth and corresponding modulation format is tunable. **Figure 10(b)** and **(c)** show the spectrum of SDPs on the flexible link between two SD-OTNs which is reflected on the filter profile. The setup/release time of end-to-end SDP is evaluated and shown in **Figure 10(d)** by collecting the statistics of the strategy processing time, OFP transmission delay and device handle times of software and hardware.

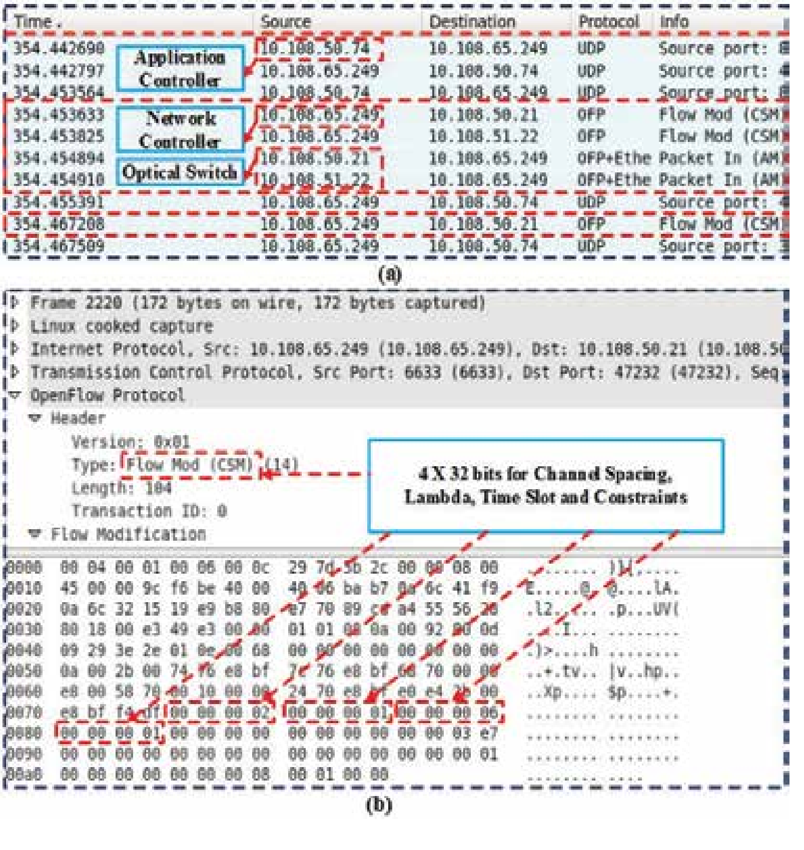

Figure 8. (a)The capture of the OpenFlow message sequence and (b) extended flow mod message for CDON.

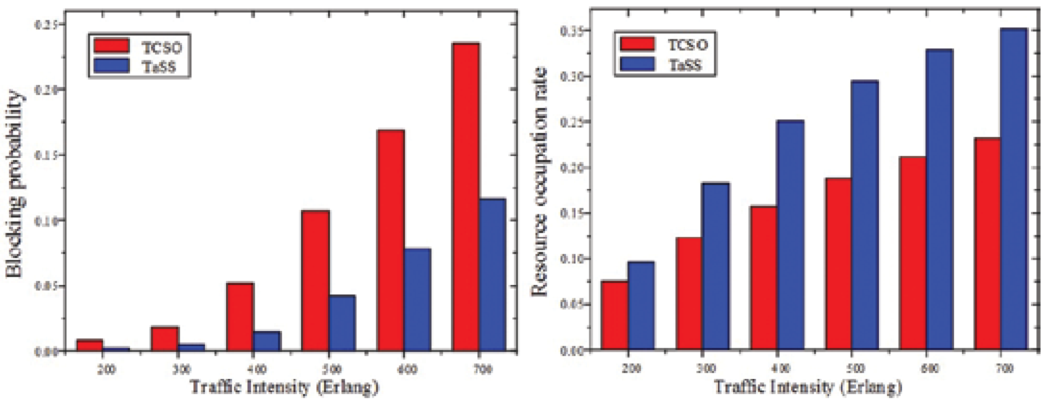

Figure 9. Comparisons on (a) blocking probability and (b) resource occupation rate between two strategies in heavy traffic load scenario.

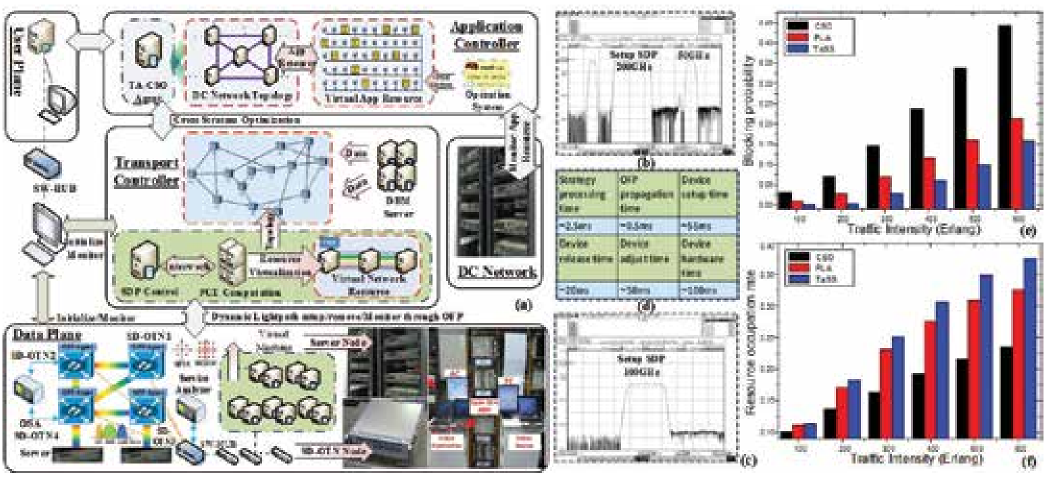

Figure 10. (a) Experimental testbed. (b, c) optical Spectrum of SDPs. (d) SDP setup/release delay. (e) Blocking probability (f) resource occupation rate under heavy load.

We test the performance of CDON by collecting the performance statistics in TaSS with CSO and physical layer adjustment strategies (PLA) under the condition of heavy load. The spectrum bandwidth of requests are randomly from 50GHz to 400GHz and the adjustable minimal frequency slot is 12.5GHz. All the destination nodes are datacentre in this network. These requests follow a Poisson process and statistics have been collected through the generation of 1×10^5 requests per execution. **Figure 10(e)** and **(f)** compare blocking probability and resource occupation rate of three strategies. In the figure, TaSS has a lower blocking probability than CSO and PLA, especially when the network is under the condition of heavy load. We can also find that TaSS can provide a better resource occupation rate than the other strategies. That is because both application and network resources can be considered in TaSS integrally and global optimization can be realized. Furthermore on the basis of it, by choosing high-level modulation format the spectrum resource is economized again. These results are further emphasized in **Figure 11(a)-(c)**.

http://dx.doi.org/10.5772/intechopen.72432

Figure 11. (a)Application graphical user interface (GUI) of testbed. (b) The capture of the OF messages for network discovery. (c) The capture of extended flow table message for SDP setup.

Figure 11(a) shows the application interface of CDON with status and destination node choice of datacenter service migration and various bandwidths of SDP. **Figure 11(b)** illustrates the OpenFlow message exchange for CDON with the capture in TC. **Figure 11(c)** shows a capture of the extended flow table modification message for SDP setup to verify the OPF extensions for CDON over flexi-grid optical networks.

5. Conclusion

Datacenter server resources will become more and more important in the future information society. It will be the main decision factor for reducing cost to allocate the resources efficiently. Considering the high capacity and low energy consumption requirements, Optical networking is regarded as promising technology for both intra-datacenter and inter-datacenter networks. The SDN technology can improve the resource utilization of application and network efficiently. In this chapter, we provide a CDON architecture in SDN-based datacenter optical network for service migration. Besides, we design the TaSS strategy for CDON and the extended OFP further. Based on our intra-datacenter and inter-datacenter networks testbed, we verify the feasibility and efficiency of CDON. The blocking probability and resource occupation rate of our approach under the condition of heavy traffic load are collected and compared with TaSS strategy. The experimental results indicate that the service with time sensitivity can be scheduled effectively and cross stratum resources utilization efficiency is improved in the CDON with TaSS strategy.

Our future works for CDON are to improve TaSS performance with dynamic parameters and consider the scalability in optical network. Then we will realize the network virtualization in datacenter optical network on our OpenFlow-based testbed.

Author details

Hui Yang

Address all correspondence to: yanghui@bupt.edu.cn

State Key Laboratory of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications (BUPT), Beijing, China

References

[1] Kachris C, Tomkos I. A survey on optical interconnects for data centers. IEEE Communications Surveys and Tutorials. 2012;**14**:1021-1036 http://dx.doi.org/10.1109/SURV.2011.122111.00069

[2] Kachris C, Tomkos I. Energy-Efficient Bandwidth Allocation in Optical OFDM-Based Data Center Networks. Los Angeles, USA: OFC 2012; Mar 2012

[3] Yoo SJB, Yin Y, Wen K. Intra and Inter Datacenter Networking: The Role of Optical Packet Switching and Flexible Bandwidth Optical Networking. Colchester, England: ONDM 2012; April 2012 http://dx.doi.org/10.1109/ONDM.2012.6210261

[4] Casimer D. Optical Networking in Smarter Data Centers: 2015 and beyond. Los Angeles, USA: OFC 2012; Mar 2012

[5] Yang H, Zhang J, Zhao Y, Ji Y, Han J, Lin Y, Qiu S, Lee Y. Experimental Demonstration of Time-Aware Software Defined Networking for OpenFlow-Based Optical Interconnect in Intra-Datacenter Networks. GLOBECOM 2013, WS-CCSNA; Dec. 2013. pp. 1-6 http://dx.doi.org/10.1109/GLOCOMW.2013.6825027

[6] Yang H, Zhao Y, Zhang J, Wang S, Gu W, Ji Y, Han J, Lin Y, Lee Y. Multi-stratum resources integration for OpenFlow based data center interconnect [invited]. Journal of Communications and Networks. Oct. 2013;**5**(10):A240-A248 http://dx.doi.org/10.1364/JOCN.5.00A240

[7] Proietti R, Nitta CJ, Yin Y, Akella V, Yoo SJB. Scalability and Performance of a Distributed AWGR-based All-Optical Token Interconnect Architecture. In: Proc. of OFC/NFOEC 2013, OW3H.1. pp. 1-3

[8] Rofoee BR, Zervas G, Yan Y, Amaya N, Qin Y, Simeonidou D. Network-on-and-off-Chip architecture on demand for flexible optical intra-datacenter networks. In: Proc. ECOC, 2012. Th.2.B.2. http://dx.doi.org/10.1364/ECEOC.2012.Th.2.B.2

[9] Deng N, Xue Q, Li M, Gong G, Qiao C. An Optical Multi-Ring Burst Network for a Data Center. In: Proc. of OFC/NFOEC 2013, OTh1A.5. pp. 1-3

[10] Liu L, Peng W, Casellas R, Tsuritani T, Morita I, Martínez R, Muñoz R, Yoo SJB. Design and performance evaluation of an OpenFlow-based control plane for software defined elastic optical networks with directdetection optical OFDM (DDOOFDM) transmission. Optics Express. 2014;**22**(1):30-40. http://dx.doi.org/10.1364/OE.22.000030

[11] Yang H, Zhang J, Zhao Y, Han J, Lin Y, Lee Y. SUDOI: Software defined networking for ubiquitous data center optical interconnection. IEEE Communications Magazine. February 2016;**54**(2):86-95. http://dx.doi.org/10.1109/MCOM.2016.7402266

[12] Yang H, Zhang J, Ji Y, Tian R, Han J, Lee Y. Performance evaluation of multistratum resources integration based on network function virtualization in software defined elastic

data center optical interconnect. Optics Express. November 2015;**23**(24):31192-31205. https://doi.org/10.1364/OE.23.031192

[13] Yang H, Zhang J, Ji Y, Tan Y, Lin Y, Han J, Lee Y. Performance evaluation of data center service localization based on virtual resource migration in software defined elastic optical network. Optics Express, September. 2015;**23**(18):23059-23071. https://doi.org/10.1364/OE.23.023059

[14] Channegowda M, Nejabati R, Rashidifard M, Peng S, et al. Experimental demonstration of an OpenFlow based software-defined optical network employing packet, fixed and flexible DWDM grid technologies on an international multi-domain testbed. Optics Express. Mar. 2013;**21**:5487-5498. http://dx.doi.org/10.1364/OE.21.005487

[15] Yang H, Zhang J, Zhao Y, Ji Y, Han J, Lin Y, Lee Y. CSO: Cross Stratum Optimization for Optical as a Service. IEEE Communications Magazine. August 2015;**53**(8):130-139. https://doi.org/10.1109/MCOM.2015.7180520

[16] Yang H, Zhang J, Zhao Y, Ji Y, Wu J, Lin Y, Han J, Lee Y. Performance evaluation of multi-stratum resources integrated resilience for software defined inter-data center interconnect. Optics Express, May 2015:**23**(10):13384-13398. https://doi.org/10.1364/OE.23.013384

[17] Yang H, Zhang J, Zhao Y, Ji Y, Li H, Lin Y, Li G, Han J, Lee Y, Ma T. Performance evaluation of time-aware enhanced software defined networking (TeSDN) for elastic data center optical interconnection. Optics Express, July 2014;**22**(15):17630-17643. https://doi.org/10.1364/OE.22.017630

[18] Yang H, Zhao Y, Zhang J, Wang S, Gu W, Lin Y, Lee Y. Cross Stratum Optimization of Application and Network Resource based on Global Load Balancing Strategy in Dynamic Optical Networks. In: Proc. OFC/NFOEC. 2012. JTh2A.38

[19] Yang H, Zhao Y, Zhang J, Wang S, Gu W, Han J, Lin Y, Lee Y. Multi-Stratum Resources Integration for Data Center Application Based on Multiple OpenFlow Controllers Cooperation. In Proc. OFC/NFOEC, 2013. NTu3F.7

[20] Zhang J, Yang H, Zhao Y, Ji Y, Li H, Lin Y, Li G, Han J, Lee Y, Ma T. Experimental demonstration of elastic optical networks based on enhanced software defined networking (eSDN) for data center application. Optics Express. 2013;**21**(22):26990-27002. http://dx.doi.org/10.1364/OE.21.026990

[21] Yang H, Zhang J, Zhao Y, Li H, Huang S, Ji Y, Han J, Lin Y, Lee Y. Cross stratum resilience for OpenFlow-enabled data center interconnection with flexi-grid optical networks. Optical Switching and Networking. January 2014;**11**(Part A):72-82. http://dx.doi.org/10.1016/j.osn.2013.10.001

[22] Yang H, Zhang J, Ji Y, He Y, Lee Y. Experimental demonstration of multidimensional resources integration for service provisioning in cloud radio over fiber network. Scientific Reports, July 2016;**6**:30678. http://dx.doi.org/10.1038/srep30678

[23] Yang H, Zhang J, Ji Y, Lee Y. C-RoFN: Multi-stratum resources optimization for cloud-based radio over optical fiber networks. IEEE Communications Magazine, August 2016;**54**(8):118-125. https://doi.org/10.1364/OE.22.017630

[24] Yang H, He Y, Zhang J, Ji Y, Bai W, Lee Y. Performance evaluation of multistratum resources optimization with network functions virtualization for cloud-based radio over optical fiber networks. Optics Express. May 2016;**24**(8):8666-8678. https://doi.org/10.1364/OE.22.017630

Chapter 13

An Integrated SDN-Based Architecture for Passive Optical Networks

Hamzeh Khalili, David Rincón, Sebastià Sallent and José Ramón Piney

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72036

Abstract

Passive optical network (PON) are often managed by non-flexible, proprietary network management systems. Software defined networking (SDN) opens the way for a more efficient operation and management of networks. We describe a new SDN-based architecture for Ethernet passive optical networks (EPON), in which some functions of the optical line terminal (OLT) are virtualized and located in an external controller, while keeping the rest of the passive optical network (PON) functionality around an OpenFlow switch. This opens the way for an improved management of the resource usage, bandwidth allocation, quality-of-service (QoS) monitoring and enforcement, or power consumption management, among other possibilities. In order to maintain the time-sensitive nature of the EPON operations, synchronous ports are added to the switch. OpenFlow messages are extended in order to cope with the PON-related parameters. Results based on simulations demonstrate that our proposal performs similarly or better than legacy architectures, in terms of delay and throughput.

Keywords: software defined networking (SDN), OpenFlow protocol, passive optical network (PON), Ethernet passive optical network (EPON), service interoperability for EPON (SIEPON)

1. Introduction

1.1. Ethernet passive optical networks

Passive optical networks technologies (among them, Ethernet passive optical network (EPON) [1] and GPON [2]) are currently used in access networks. As shown in **Figure 1**, passive optical network (PON) is composed of a central office (optical line terminal, OLT), passive optical splitters, and one terminal (optical network unit, ONU) at each one of the customer premises, in a tree topology. First-generation PONs use time division multiplexing (TDM) for sharing the medium in a point-to-multipoint scenario. In the downstream, the OLT broadcasts the data frames towards the ONUs, while in the upstream an arbitration mechanism is needed in order to avoid collisions when ONUs send frames to the OLT. The EPON standard, which will be the focus of this work, uses the Ethernet format for the data frames.

A key piece in EPON is the concept of dynamic bandwidth allocation (DBA) algorithms, which allow the OLT to orchestrate the access to the shared medium and to dynamically adapt to the changing requests from the ONUs. In the EPON architecture, an entity called multi point control protocol (MPCP) manages the upstream channel and harmonizes the transmission of data from ONUs to OLTs. In addition, it manages operations such as terminal discovery, registration, and bandwidth allocation. MPCP uses two standard messages for DBA operations, namely GATE and REPORT. REPORTs are used by ONUs to request transmission opportunities, and GATEs are sent by the OLT to grant a transmission slot (time and length) to a specific ONU.

1.2. EPON service interoperability

In order to ensure service interoperability in a multivendor scenario, IEEE developed the 1904.1 Service Interoperability for Ethernet passive optical networks standard (SIEPON) [3–5]. One of the key concepts in SIEPON is the definition of a unified data path architecture to ensure that data services can be managed, provisioned and monitored across the EPON. SIEPON covers the physical (PHY), medium access control (MAC), and upper layers to manage tasks related to the data path, such as multicast delivery, tunnels, VLANs, and quality-of-service (QoS) enforcement.

EPON service pathways (ESPs) are defined as the path that a data frame follows through the media access control (MAC) client functional blocks, specifying both the connectivity and QoS of the frame. Connectivity is defined by classifying the frame according to header fields (VLAN tag, MAC address, IP address, to name a few possibilities), while QoS parameters can

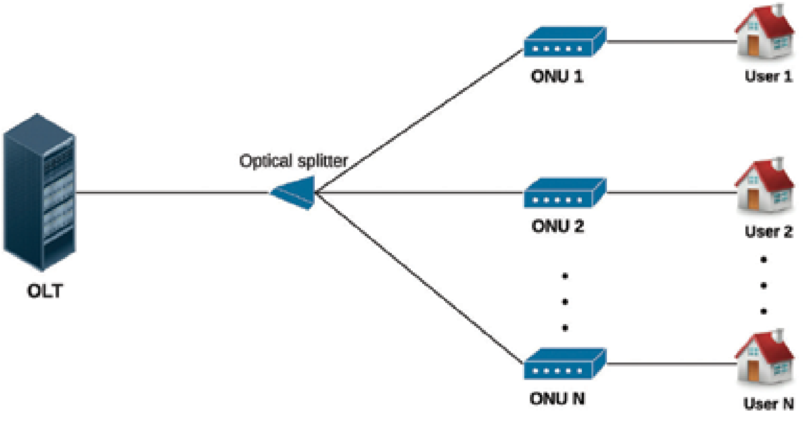

Figure 1. PON topology.

be enforced by queues with priorities and scheduling. **Figure 2** shows the SIEPON architecture in both the OLT and the ONUs, and specifically the ESP-related blocks: the Input block [I] is the ingress port that receives frames from user-network interface (UNI) or network-network interface (NNI); the classifier block [C] classifies incoming frames; the modifier block [M] is able to modify frame fields; the Policer/Shaper block [PS] enforces conformance to the service contract; the Cross-connect block [X] is used to direct a frame to the proper queue; the Queue block [Q] holds frames until they are selected by the scheduler for transmission; the Scheduler block [S] transmits the frames from the queue block to the output block using a predefined scheduler algorithm; and finally Output block [O] is the egress port that receives frames from a scheduler and forwards them to the UNI or NNI. Each one of the ESP blocks can involve several independent instances operating on various flows of traffic.

1.3. Software defined networking

Software defined networking (SDN) [6] is a new paradigm for network management that introduces network programmability and separation of the data and control planes, migrating the complex, "intelligent" functions from the network devices to a centralized controller. These characteristics allow the SDN controller to obtain a centralized and accurate view of the state of the network, thus opening the way for the optimization in several ways (bandwidth assignment, QoS guarantees, minimization of power consumption, or resilience, to name a few possibilities).

The SDN architecture defines two interfaces in the controller. The northbound interface defines an API that allows network management systems (NMS) or ad-hoc applications to communicate with the controller. The southbound interface communicates the control

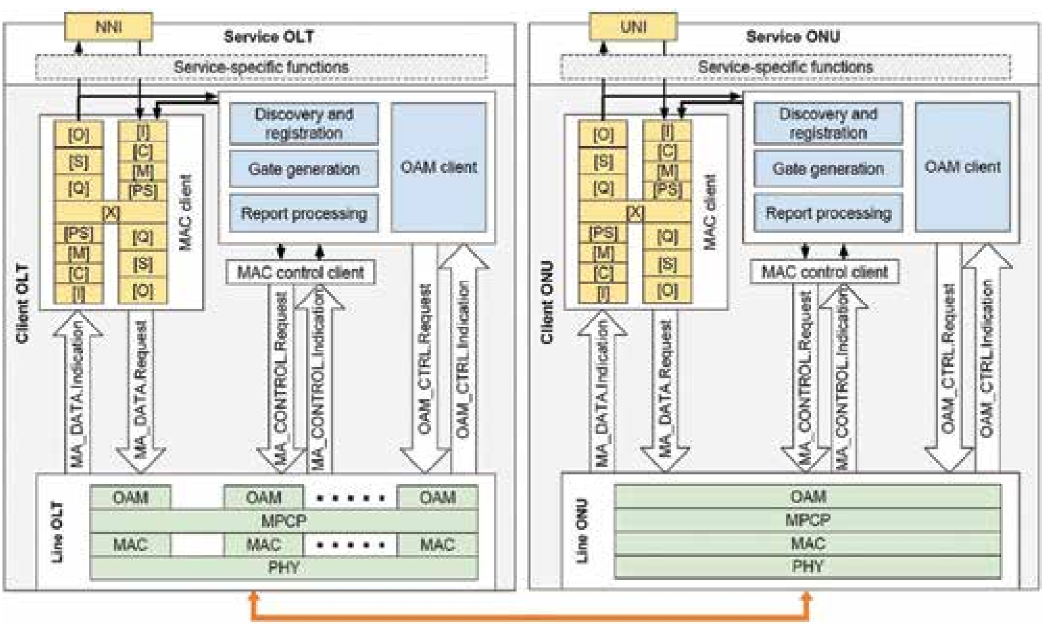

Figure 2. SIEPON architecture, OLT (left) and ONU (right).

plane that runs in the controller with the data plane that resides in the network equipment (switch). The original southbound API, and probably the most popular, is OpenFlow (OF). The OpenFlow protocol [7] defines a standardized instruction set that allows the controller to manage any OpenFlow-enabled device and specify the path to be followed by traffic flows through the network of switches, by means of the definition of matching rules (based on packet header fields) and operations to be performed (e.g., forwarding to a specified port, modifying the headers, or dropping the packet).

1.4. SDN-based PON networks

The management of access and backbone optical networks is becoming more and more important in the research literature [8], and how to apply SDN is being now intensively investigated. In the specific case of PON networks, given its centralized nature (the OLT manages almost all the operations), applying the centralized control envisioned by SDN makes even more sense. However, the aspects related to service interoperability, or a detailed SDN architecture (including extensions of OpenFlow messages and actions), are relatively unexplored. The authors of [9] provide a useful and up-to-date survey on the contributions on SDN-based optical networks. Here, we discuss the most relevant works in the field of SDN and optical access.

In [10, 11], authors envision a scenario where each optical switching node is virtualized, in order to obtain a unified control plane. Each physical interface is mapped to a virtual interface. In order to ensure a smooth transition from legacy equipment, an extra layer would translate the messages between the controller and the switches [12], at the cost of adding latency by using message proxies. The authors of [13] developed an architecture for PON networks based on SDN including an OpenFlow extension for traffic mapping and forwarding capabilities, with no effect on data link layer latency. In [14] an architecture control plane for converged metro-access networks under SDN is described. Reference [15] presents a novel software-defined optical access network (SDOAN) architecture. The purpose of developing a Service-Aware Flow Scheduling (SA-FS) strategy is to assign network bandwidth resources in an efficient and flexible way. In order to operate PONs coordinated with a core SDN-based network, Ref. [16] describes an architecture for a Software-Defined Edge Network (SDEN) able to control end-to-end the traffic flows. In [17], authors propose a new SDN- and NFV- (Network Function Virtualization) based architecture for EPON networks; in this scheme, OLTs and ONUs are partially virtualized and moved to a central controller. Reference [18] describes, at a high level, an EPON architecture based on SDN that replaces the hardware-based DBA with a software-based one, residing in the controller. The authors of [19] describe a GPON architecture where an OpenFlow agent is located in the OLT to communicate with the SDN controller, claiming that the approach can connect several sites in different locations in a cost effective manner.

Central office re-architected as a datacenter (CORD) is a novel architecture developed in [20] aimed at substituting the telephone exchange hardware with software-based equipment, converting central offices into datacenters, with the purpose of speeding the deployment and increasing the efficiency of services. CORD decouples the control and data planes, using the open network operating system (ONOS) [21] SDN controller, and virtual machines running on top of OpenStack. CORD is currently supported by service providers such as AT&T and NTT Communications.

To the best of our knowledge, Refs. [22, 23] are the only works that tackle the topic of how to adapt the SIEPON standard to the SDN architecture. Reference [22] describes methods for enabling nodes to exchange information between a management protocol and other protocols, including control plane and data plane interfaces to support SDN. Reference [23] expands the work presented in [17], converting the OLT into an OpenFlow switch, and describes the extensions to make it compatible with the SIEPON architecture, and thus reach the goal of having a virtualized, simplified and centralized management system migrated to the SDN controller.

The remainder of this chapter is organized as follows: In Section 2, we describe the SDN-based EPON architecture and the OLT functions. Section 3 expands the description by defining the new elements and modifications, together with details of its operation and implementation. Section 4 presents the validation of the proposal, together with results obtained with a simulator and a reference implementation. Section 5 concludes the chapter.

2. SDN-SIEPON architecture

This work builds on [17] and expands [23] with more details and results. We present a novel SDN-SIEPON architecture based able to decouple, virtualize, and simplify the management and operation of the OLT and OAM functions, and opens the way for multi-tenant and multi-provider optical access networks, where various service providers use the same basic infrastructure.

Our architecture is based on the principle that the OLT is built around an OpenFlow switch that takes care of the forwarding plane and carries out the functions related to the EPON service path, while some of the control plane-related functions are migrated to an OpenFlow controller. The OpenFlow switch performs the following functions: (1) classifies incoming packets following the matching rules; (2) modifies packet header fields if required; (3) schedules and ensures QoS for each flow; and (4) forwards packets towards their destination ports. In order to do so, the OpenFlow switch emulates some of the EPON functionalities defined in the OLT, such as control multiplexer (CM), control parser (CP), and multipoint transmission control (MPTC). Many of these functionalities work in real time, and therefore we need to define synchronous ports and make it compatible with the SDN architecture by defining a set of registers. The operations that must be executed in real-time are kept in the switch, while those that work at longer time scales can be migrated to the SDN controller.

As shown in **Figure 3**, the SDN-SIEPON architecture includes three main elements: an SDN-based OLT; an SDN controller; and the EPON network, including the passive splitters, the fibers and the ONUs. We focus here on the redesign of the OLT, and although the partial virtualization of the ONUs might be of interest (for example, in the context of energy saving or QoS management), it is out of the scope of this work and left for future research.

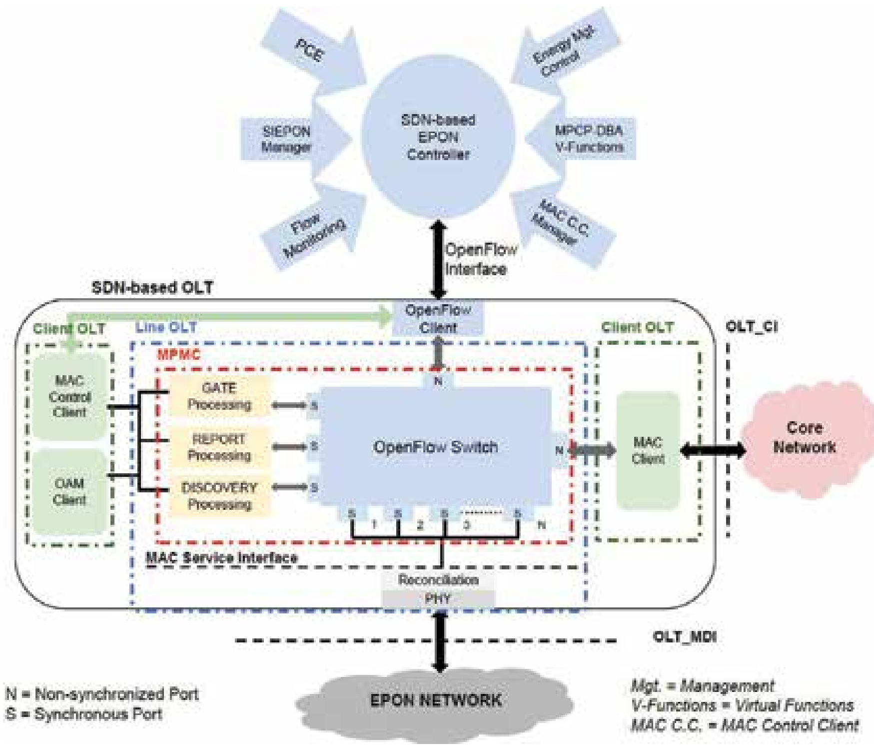

Figure 3. Elements of the SDN-SIEPON architecture.

2.1. SDN-OLT

The SDN-based OLT is built around an OpenFlow switch, and its tasks include the forwarding functions of the MPMC sub-layer and the operations of the Control Multiplexer, Control Parser, and Multipoint Transmission Control elements of the EPON architecture.

A non-synchronous port is in charge of the connection of the SDN-OLT to the SDN controller, while a set of synchronous logical ports link the ONUs and the operations of the SDN-OLT (among them the processing of the GATE and REPORT messages, the discovery process, and the DBA management). There is a synchronous port per ONU (with a maximum of 64 ONUs per OLT), attached to a specific MPMC instance running in the SDN-OLT, plus one extra logical port for broadcasting messages to all ONUs. The 64 port-instance pairs are activated on demand, depending on the number of active ONUs. Whenever an ONU is activated, the SDN controller creates a MPMC instance and links its three logical ports to the DISCOVERY, GATE, and REPORT processing modules. In parallel, the controller creates a unicast MAC service logical port and connects it with the recently started up instance. While legacy OLTs require several complex entities with multiple coordinating entities, our design simplifies the architecture by setting the OpenFlow switch as the single coordinator of all the entities, thus arbitrating the access to the downstream and upstream channels, as illustrated in **Figure 4**.

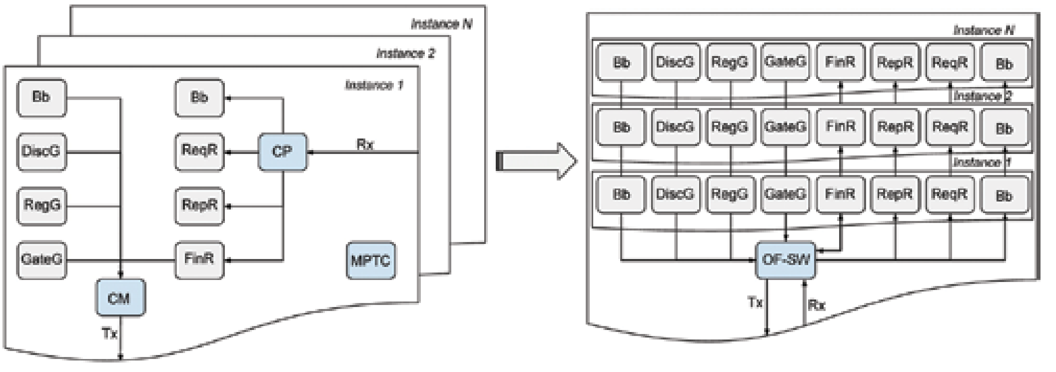

Figure 4. Left: architecture of a legacy OLT with several complex MPMC instances. Right: the simplified SDN-OLT architecture. Acronyms of the processing entities: OF-SW (OpenFlow switch); Bb (backbone), DiscG (discovery gate generation); RegG (register generation); GateG (gate generation); FinR (final registration); RepR (report reception); and ReqR (request reception).

2.2. SDN controller

The SDN controller is a piece of software that manages one or several SDN-OLTs, and runs the complex, non-real time functions that compose the control plane of the EPON network. It is usually run on virtual machines in a datacenter, and has a centralized view of the network. One key aspect of our design is the decoupling of the virtualizable functions that work in longer timescales and can be migrated to the controller from the functions that work at shorter timescales and have to be kept at the switch. To name a few cases, the controller manages the parameters of the MAC control client; the policies of the MAC client; or the translation of the linking the high-level SIEPON QoS class definition to the low-level MAC QoS parameters.

An important SIEPON module that is also migrated to the controller is the ESP manager of the MAC client, responsible of aspects such as provisioning, QoS policies, flow management, bandwidth assignment policy, time-related decisions, addition of unregistered ONUs to the network, and data forwarding services through the SDN-OLT. It is also in charge of handling and configuring the initial parameters of the MAC control client in the OpenFlow switch, and the processing of GATE, REPORT, and DISCOVERY messages as well. It also provides a database for the OpenFlow switch to keep information regarding logical link identifiers (LLIDs), status of logical port, and ONU MAC address so as to configure flow tables for the switch.

The MPCP-DBA virtual module handles the overall network policy function and allows the controller to communicate with SDN-based OLT to alter parameters and policies of DBA, ONU priority in DBA, or shift between several available DBA algorithms.

3. Extensions and modifications of OpenFlow

This section describes the extensions and additions to the OpenFlow architecture that are needed for our proposal. OpenFlow messages and actions will have to be extended to support

EPON operations, and synchronous ports will be needed in the switch in order to keep the synchronous nature of EPON. Another aspect to tackle is how to implement the ESP functional block of SIEPON. Finally, a new and simpler MPMC sub-layer will be described.

3.1. Extension of OpenFlow messages

The original OpenFlow messages lack specific headers and parameters needed to execute some of the SIEPON-related operations. Basically, we need:

1. A way to determine whether the OF switch is a SDN-OLT type or not. This is solved with an extension of the OFPT_FEATURES_REPLY message sent in response to the controller query.
2. Maintain correspondence of the logical link identifier (LLID) of the port and the MAC address of the ONU. This is done by extending the OFPT_PORT_STATUS message.
3. Set up the initial parameters of the MAC control client and MAC client (for example, setting the LLIDs) during the initial phase of the dialog between the SDN-OLT and the SDN controller. This can be done by extending OFPT_SET_CONFIG message.
4. Transport the set of synchronous rules, matching fields and actions. This information is carried in an extension of the OFPT_FLOW_MOD message.

OpenFlow match fields must also be extended in order to support the new SIEPON-related functionalities. A PDU_type field is created to maintain the type of received packet (a MAC control frame or a data frame). An opcode type distinguishes the control frames as a DISCOVERY GATE, REGISTER_REQ, REGISTER, REGISTER_ACK, GATE or REPORT message. Flag is used to classify the type of GATE packet (DISCOVERY GATE or normal GATE). LLID contains the port's LLID number, which is given by the controller and allocated by the MAC control client. Finally, a Grants_number stores how many grants have produced as result of the processing of the GATE message. It can range from 0 to 4, where 0 signals a periodic GATE packet and 1 to 4 are related to the bandwidth allocation by a normal GATE.

3.2. Synchronous ports

The synchronous nature of PON networks requires the SDN-OLT to be able to handle real-time operations in synchronous ports, through which packets are sent or received in specific times specified by the synchronous flow entries. We developed a MPMC sub-layer extension that retains synchronicity and allows us to replace control parser, control multiplexer, and multipoint transmission control components of a legacy OLT with the OpenFlow switch.

The synchronous operations related with a newly added ONU include:

- Setting an instance_ID in the OpenFlow switch for the respective MPMC instance.
- Allocating three synchronized logical ports to each gate, report, and discovery processing instance.

- Allocating an input/output synchronized logical port with an associated LLID so the port is linked with the MAC address of the newly connected ONU.
- Timestamping of packets delivered to/received from the ONUs. The timestamps are obtained from the local clock of the OpenFlow switch.
- Computing and monitoring the round trip time (RTT) of the packets received from the ONUs.
- Assigning bandwidth and transmission time slots in the upstream direction (packet forwarding strategy), in cooperation with the SIEPON manager (based on its QoS policies).
- Creating and destroying flow entries in the OpenFlow switch.

Several registers are used to emulate the synchronous operation:

- P (transmitPending), I (transmitInProgress), and E (transmitEnable) are employed to synchronize the MPMC instances with the DBA (thus allowing the packets from the instances to share the channel in an arbitrated way). The OpenFlow switch controls the transmission of packets towards the Reconciliation Sub-layer (RS) by activating the transmitEnable signal.
- Two additional registers are used for round-trip-time (RTT) calculation and packet timestamping. RTT calculation is used for synchronization between OLT and ONU.

Figure 5 illustrates the extension of the MPMC sub-layer of the SDN-OLT.

3.3. ESP functional block

The aforementioned extension of the Flow-Mod message allows us to introduce the functionality of the SIEPON ESP functional blocks in the SDN-OLT. **Figure 6** describes an example of an OpenFlow flow entry extended with the new matching fields and actions. The actions related with the ESP functional blocks are:

1. Set-Field: it is consistent with the Modifier block [M] and implements all the set-field actions in the packet. It modifies the value of the packet header fields (e.g., Ethernet/IPV4/IPV6 source and destination addresses, VLAN ID and VLAN priority, for example).
2. Apply meter: it is consistent with the Policer/Shaper block [PS] and allows packets' rate measurement and control (setting rate limit), and implement DiffServ-like policy. In addition, meters can assist per-port queues in scheduling packet on an output port with respect to their priority, thus ensuring QoS.
3. Set-Queue: it is consistent with the Queues block [Q], and sets the queue_id for a packet when it is delivered to the port. It is used for packet scheduling and forwarding.
4. The Scheduler block [S] controls the actions related with the transmitPending, transmitInProgress and transmitEnable actions by altering their associated registers P, I, E.

5. Output port: it is consistent with the Output block [O]. Packets are allocated to a physical port (i.e., controller) or a logical port (either unicast or broadcast, depending on the ONUs targeted).
6. Operations performed by the Classifier [C] and Cross-connect [X] blocks are executed via rule matching and actions in the flow table. Incoming packets are compared and classified based on their matching fields with the existing rules, and packets are later sent to the destination based on the associated action.

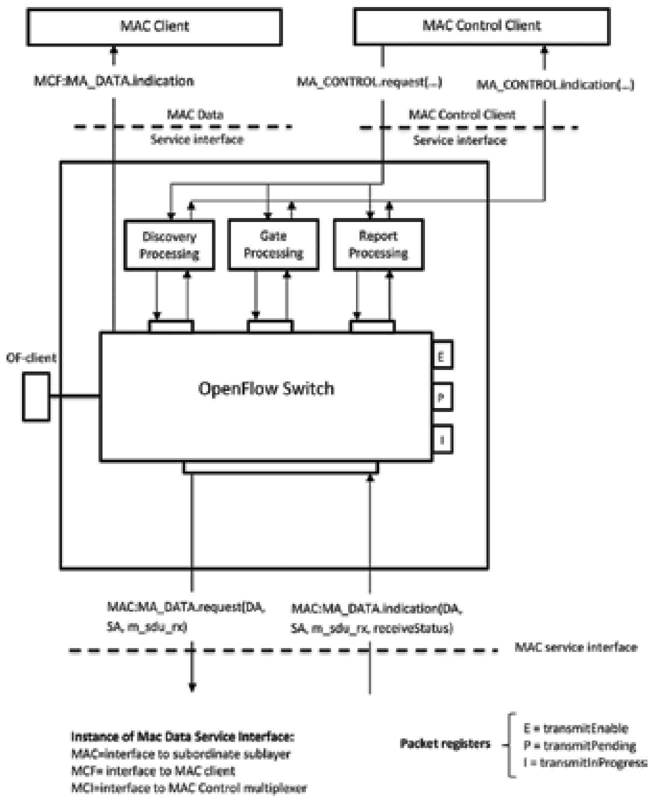

Figure 5. MultiPoint MAC Control functional blocks, from [23].

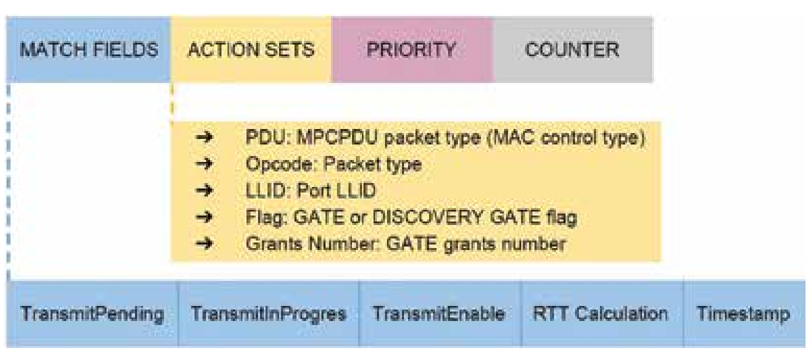

Figure 6. Example of extended match fields and actions.

3.4. MPMC sub-layer

In the legacy OLT the MPMC sub-layer is quite complex: a MPTC module synchronizes multiple instances of MPMC (one for each ONU, with a control multiplexer and a control parser entity). Our proposal simplifies this by introducing three single entities of MPTC, CM, and CP, implemented as rules and flow entries in the OpenFlow switch. Therefore, the switch's task is to coordinate with different MAC instances of the OLT and coordinate their requests for sending or receiving packets.

Figure 7 illustrates the main blocks that compose an OpenFlow switch: OF agent; control path; and data path. The OpenFlow agent interacts with the SDN controller in order to manage both the control and data paths. The control path is responsible of emulating the OLT functionality through the set of rules and actions in the flow entries. The data path task is to match the received packets with respect to the matching rules, and execute the attributed actions (e.g., frame forwarding and scheduling for transmission).

The MPMC sub-layer carries out the following functions: after receiving a frame from the underlying MAC, forwards it to the OpenFlow switch flow table in which the frame is analyzed based on its opcode. The REPORT and data frames coming from ONUs are analyzed by CP functions. In **Figure 7** each data path flow entry related to the control path element is shown in the same color. The MPMC also coordinates the processing of DISCOVERY GATE, GATE, and data frames created by the MAC instances and forwards them to the RS layer (shown in green in **Figure 7**), and manages the transmission of data frames using the three registries (i.e., P, I, and E) defined in the OpenFlow switch control path.

3.5. An example of operations

We now provide an example to illustrate the procedure for generating a new flow entry in the SDN-OLT for processing a DISCOVERY GATE, GATE and REPORT messages. Let us recall that a GATE message tells a specific ONU when to send data (start time) and how

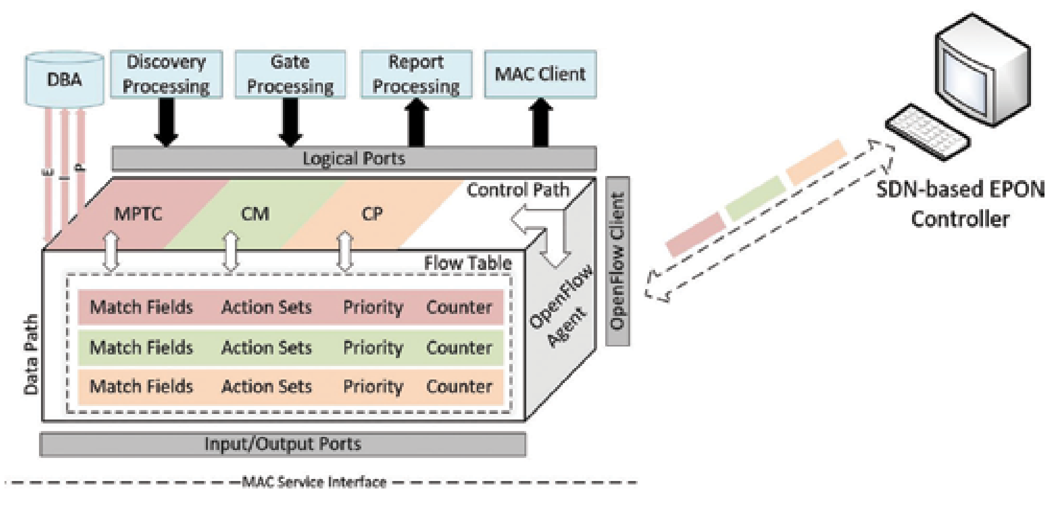

Figure 7. MPMC sub-layer design.

much data to send, in response to a previous REPORT message from the ONU requesting a transmission opportunity. Regarding the discovery process, DISCOVERY GATE messages are broadcasted periodically to all ONUs, and are replied by the newly connected ONU with a REGISTER REQUEST (randomly delayed in order to minimize collision with other ONUs). The OLT confirms by sending a REGISTER message immediately followed by a GATE that grants and schedules an ONU response in the form of a REGISTER-ACK message, finishing the discovery and register process. **Figure 8** illustrates the state diagram for the actions related to processing the aforementioned messages.

When the SDN-OLT establishes the communication with the SDN controller, the latter sends a series of FlowMod messages with pre-defined rules to be inserted in the flow table of the OpenFlow switch. These rules define how to process the DISCOVERY GATE, REGISTER_REQ, REGISTER, REGISTER_ACK, GATE, and REPORT messages. The MAC control client sends a request to a discovery processing entity of an instance to creates a DISCOVERY GATE message with the required fields (discovery information and policy to use by the ONU) and sends it through the appropriate synchronous logical OpenFlow switch port. After receiving a packet, a flow table lookup is performed by comparing the packet header fields with the matching fields of the installed rules by following a priority order. When the packet is matched, the associated counter for the chosen flow entry is increased and the actions linked with the flow entry are performed.

When, later on, the controller receives a request (through Packet-In) from the SDN-OLT, for processing a GATE message, it gathers and analyses the received packet data (source and destination MAC addresses, PDU type, opcode, flag, grant number, LLID and VLAN IDs) to set a rule in the OF switch for determining the grants' number and establish the QoS policy, and

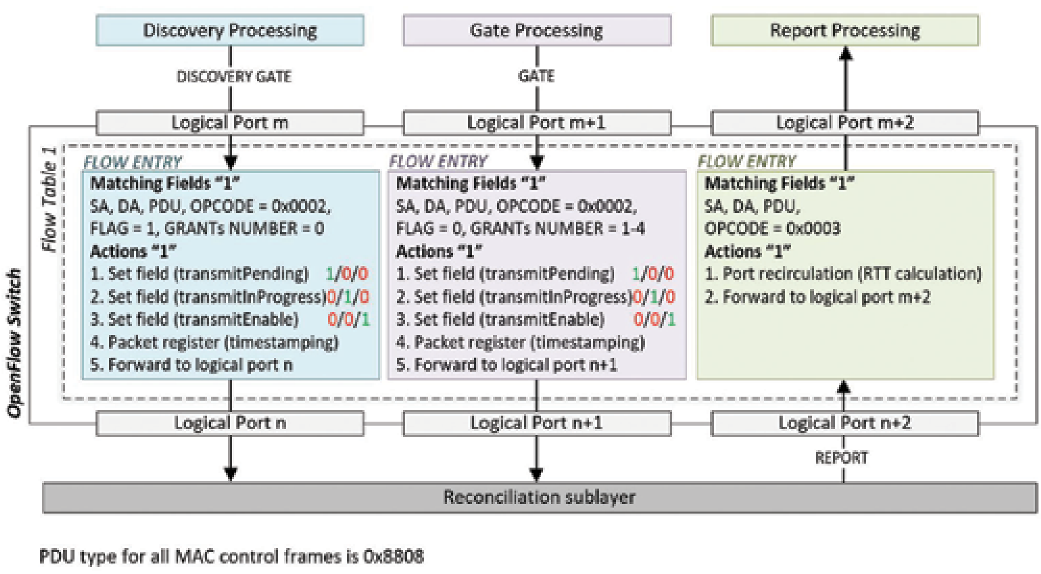

Figure 8. Operations related to the processing of DISCOVERY GATE, GATE and REPORT messages.

prioritize packet delivery towards an output port. The rule is transmitted through a FlowMod and a Packet-Out messages to the OpenFlow switch, and the received GATE message is processed accordingly.

For both the DISCOVERY GATE (right side of **Figure 8**) and GATE (center of **Figure 8**) messages, when the packet is prepared for broadcasting, the packet is scheduled for transmission by setting register P (transmitPending). The packet, placed in the queue, might wait for higher priority packets to be processed by the DBA. The transmitInProgress action is performed by setting register I so as to alter transmission state of the packet to transmitInprogress (by setting I to 1 and transmitPending to 0). Then, transmitEnable action is performed when no more packets are available to be transmitted by setting register E, and the transmission state of the packet is altered to enable (E is set to 1 and transmitInProgress to 0). The next action is to send the packet towards a specified logical output port. A timestamp is then added to the frame and the packet is sent to the RS layer. In the case of REPORT message (at the left side of **Figure 8**), as soon as the REPORT message is received from an ONU and matched against a rule in the flow table, the action of calculating the RTT is executed on the packet (via port re-circulation and interacting with the packet registers) and the value is communicated to the DBA. Once the RTT is calculated, the packet is forwarded to the REPORT entity for further processing.

4. Evaluation

This section describes the simulation model and the performance evaluation of our proposal. We built models of both the legacy EPON architecture of the SDN-based EPON architecture using the OPNET Modeler package, and compared them in terms of delay, and throughput. The results expand our initial validation described in [23] with an evaluation of the influence of the traffic pattern, the number of active ONUs, and end users in the performance of the system.

4.1. Description of the scenario and delay analysis

A scenario with a SDN controller, a SDN-OLT, and a tree topology with 16 ONUs was built. The distance between the SDN controller and the OLT is initially set to 1 km, and the ONUs are at distances ranging from 16 to 18 km from the OLT. The link rate in both the downstream and upstream directions is 1 Gbit/s, the guard time is 1 μs, and the maximum cycle time is set to 1 ms. IPACT [24] is used as DBA algorithm, and the message processing delay in the SDN-OLT is set to 0.0164 ms [25].

Two traffic generators are used: constant bitrate and self-similar traffic [26, 27]. For CBR traffic, packets have a constant length of 791 bytes, while in the second case packet size is uniformly distributed between 64 and 1518 bytes (with a mean of 791 bytes, to compare with the CBR case), and the Hurst parameter (a measure of the long-range dependence of self-similar traffic) was set to 0.7, 0.8, and 0.9. In order to understand the delay analysis that follows,

Figure 9 shows the scenario and the delay measurement points, where the dashed path is used for calculating the round trip time (RTT) delay, and the DE path is used for evaluating the queuing delay and throughput of the system.

Table 1 shows the time it takes for the controller to process each type of packet, and the time for transmitting the packet through the logical output port in the SDN-OLT, for the cases of the six initial FlowMod messages (DISCOVERY GATE, REGISTER_REQ, REGISTER, REGISTER_ACK, GATE, and REPORT,) where the OpenFlow installs the flow rules at the

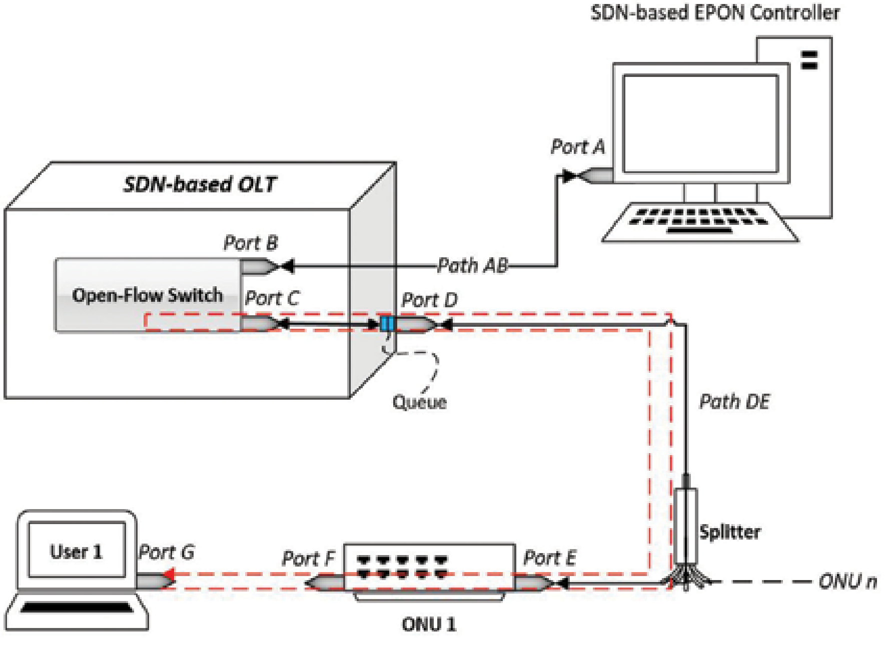

Figure 9. Measuring points in the SDN-based EPON scenario.

FlowMod	Tx delay in the controller (port A) (μs)	Rx delay in the switch (port B), d = 1 km (μs)	Rx delay in the switch (port B), d = 10 km (μs)	Tx delay through the path AB, d = 1 km (μs)	Tx delay through the path AB, d = 10 km (μs)
DISCOVERY GATE	0.52	5.76	51.66	5.243	51.14
REGISTER REQ	1.04	6.28	52.18	5.243	51.14
REGISTER	1.56	6.80	52.70	5.243	51.14
REGISTER_ACK	2.08	7.32	53.22	5.243	51.14
GATE	2.60	7.84	53.70	5.243	51.14
REPORT	3.12	8.36	54.26	5.243	51.14

Table 1. Average values of packet processing delays.

switch during the connection establishment. The average packet delay is measured for two different cases of distance between the controller and the SDN-OLT, 1 Km and 10 Km. Even for the longer distance, in which a single controller located in a datacenter could manage simultaneously several OLTs in a diameter of 20 Km (i.e., a medium-sized city), response times are almost negligible.

A metric of interest is the influence of the presence or absence of the flow processing rules in the performance of the SDN-OLT. In the first case we assume that the SDN controller has previously installed in the switch the rules for processing the control frames (DISCOVERY GATE, REGISTER_REQ, REGISTER, REGISTER_ACK GATE, and REPORT). The delay is evaluated from the moment a packet (such as control frame or data frame) enters the OpenFlow switch via a synchronous port, including the look-up for a rule in the flow table to match and perform a group of actions (such as read and write a register, and forward the packet to the output port). The average processing time in presence of the rules in the flow table is $T_{presence\ of\ rule} = 0.488\mu s$.

For the case when the rules are not yet installed in the switch, whenever a new packet arrives, the OpenFlow switch sends a Packet-In message to the controller, who in turn creates a FlowMod and a Packet-Out message to insert a rule in the flow table, followed by the look-up time and action execution. In this case, the average processing time can be described as the first look-up time ($T_{1st\ look-up} = 0.297\mu s$), plus forwarding time of the packet to the logical port B ($T_{forwarding-to-port-B} = 0.096\mu s$). In addition, the packet forwarding time from logical port B to a physical port A is, and the controller processing time for analyzing the incoming the Packet-In message, installing the rule, and issue a FlowMod and Packet-Out messages is $T_{controller\ processing} = 1.986\mu s$. The second look-up and actions execution times are, as in the first case, $T_{2nd\ look-up+actions} = 0.488\mu s$. Therefore, the average processing time in absence of the rules is $T_{absence\ of\ rule} = 13.353\mu s$. Since the look-up time process will vary depending on the length of the flow table and the number of flow entries, we carried some experiments. For both cases (in presence and absence of the rules) the average look-up time when 10 flow entries are present in the table is 0.294 μs, increasing to 0.398 μs, when 100 flow entries are present. Therefore, the influence of the number of flow rule entries does not play an important role in the delay performance of the system.

4.2. QoS performance evaluation

We now focus on the evaluation of the QoS in the path between the OpenFlow switch and the end users, shown as a dashed line in **Figure 9**. This path includes the look-up and rule matching times, registers read and write operations, and waiting time in the port queue (point D in the MAC service interface, queue in **Figure 9**). The results are obtained by averaging 50 values for each parameter.

Figure 10 shows the round-trip-time (RTT) delay against the offered load for both the SDN-based and legacy architectures. The RTT delay is evaluated for the data frames, which are generated and forwarded to the logical port C of the OpenFlow switch by the end user through the physical port G, and returned back. In a real situation, packets would actually ingress the core network, but since we are interested in the PON segment, we do not include this term. The results are presented using box plots, showing the average delay value (center of the box), the first and third quartiles (bottom and top of the box), lowermost and

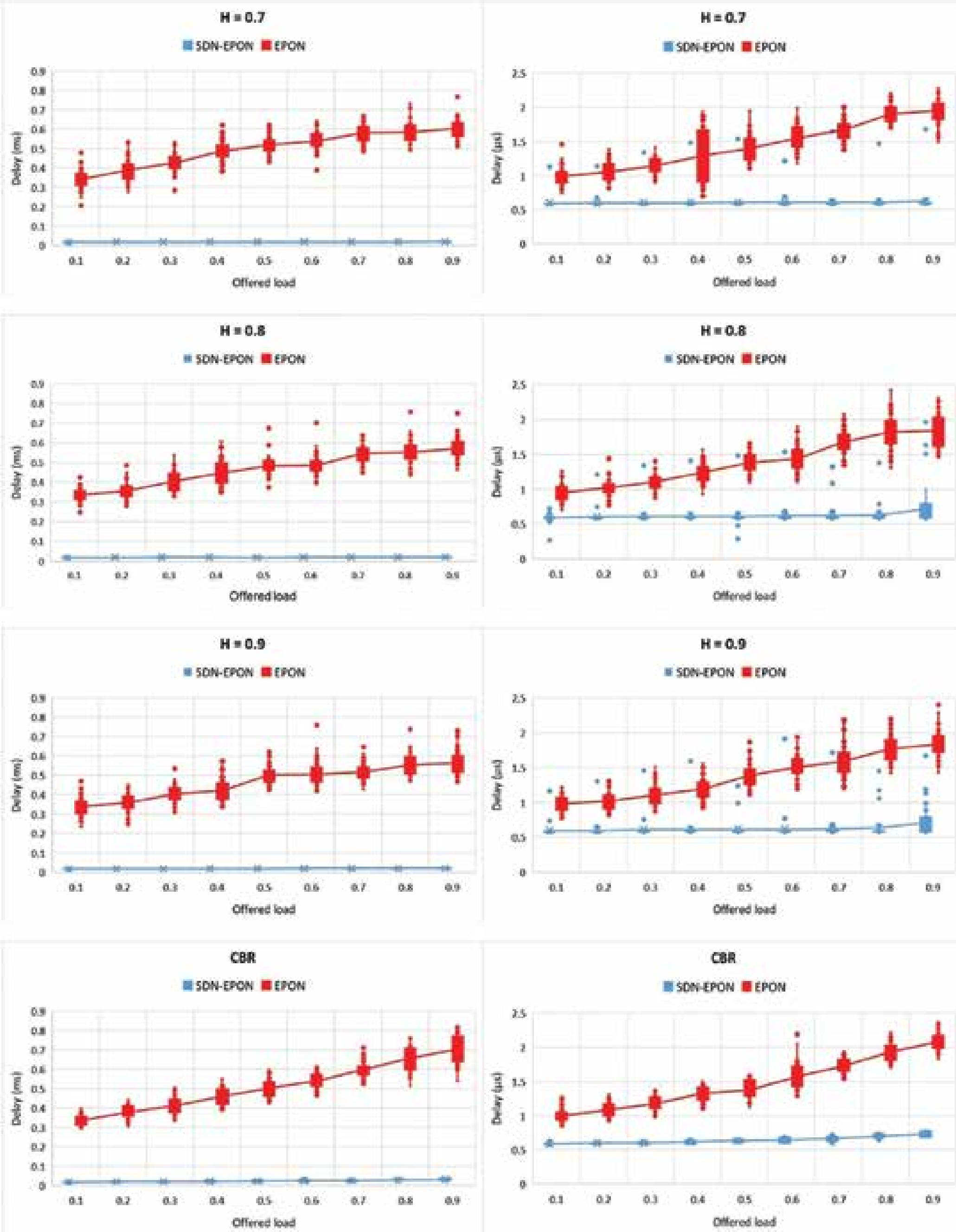

Figure 10. Comparison of RTT delay (left column) and downstream queuing delay (right column) in SDN-based EPON and legacy scenarios under different offered load, for CBR and self-similar traffic patterns and different Hurst parameter.

uppermost values (whiskers below and above the box) and the low min and high max values (red dots scattered). Please note that, although they are related to the same x-axis value, red dots not coincide with the blue dots at the same x-axis position in order to enhance visualization and avoid overlapping. Both SDN-based and legacy EPON architectures are evaluated in the same conditions (global offered load) and each blue or red dots in the figures relates to the average value obtained for all the ONUs. The offered load is expressed as 10% fractions of the maximum capacity of the system.

The main conclusion from the left column of **Figure 10** is that the delay performance is much better in the SDN-based architecture, particularly when the network load is high. The almost constant RTT in the SDN-OLT scenario is explained by to the presence of pre-defined rules in the flow table, thus ensuring almost constant processing time. In the SDN-based architecture the DBA procedures are highly simplified, while in the legacy EPON architecture, several Control Multiplexer and Control Parser entities (one per instance) are involved. This also explains the higher variability (red dots dispersion) in the legacy case, compared with the more deterministic behavior of the new architecture. The traffic pattern does not have a significant affect in the results. CBR and self-similar traffic offer the same average results, with less variance in the case of constant traffic. Long-range dependence characteristics of the traffic pattern do not seem to affect significantly the performance, with only a slight increase in the dispersion of the measurements when H is increased (probably related to the presence of more time-correlated traffic bursts when the Hurst parameter is increased).

The right column of **Figure 10** shows the queuing delay in the downstream direction against the offered load for both architectures and for different traffic sources (self-similar with different Hurst parameters and CBR). The queuing delay is evaluated for both control and data frames at the OLT physical port (port D of the MAC service interface), and correspond to measurements of the packet waiting times in the downstream direction of the queue. The delay is measured from the moment when a packet is forwarded to the transmitter channel queue and until the time the last bit of the packet is transmitted. Our goal here is to evaluate the possible impairments in the downstream transmission in the SDN-based architecture. As discussed earlier, the OpenFlow switch affects the downstream transmission by controlling the synchronous logical ports that communicate with the DBA entity. The results show a behavior similar to that of the RTT. In the legacy case, the queuing delay is very dependent on the load, while it almost does not vary in the case of the SDN-based architecture—and in this case the delay is much smaller. Again, this is caused by the simplified architecture of our proposal.

In the next set of experiments we vary the number of active ONUs (16 and 8), for different offered loads, and Hurst parameter 0.9. The case of 16 active ONUs is the base case already analyzed previously, while in the second scenario only 8 ONUs send and receive the total traffic of the network (we double the rate of each ONU, in order to maintain the global load equivalent) and the other eight ONUs do not generate any user-related traffic and only communicate with the OLT through the control frame messages. As **Figure 11** illustrates, in both architectures, the RTT delay increases with the input traffic, but as in the previous experiments, the delay for the legacy EPON case is much higher. The number of active ONUs has a limited influence, shown as a small increase in the delays experienced in both architectures. The increase is caused by the fact that we are averaging the delays experienced by the packets generated by ONUs that generate a rate that has been doubled, and therefore their packets experience increased waiting times.

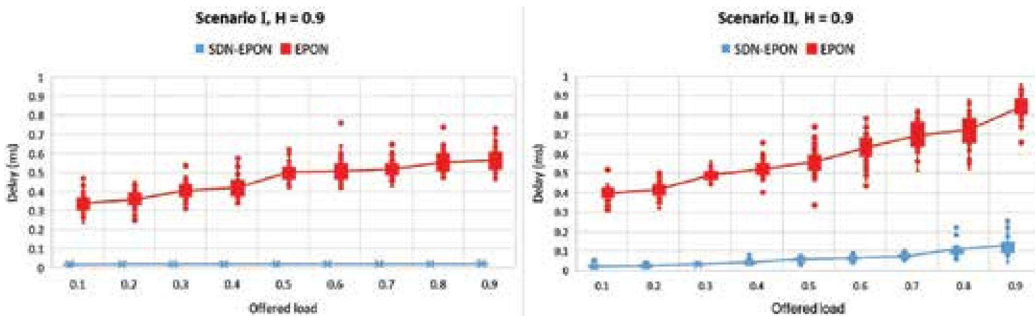

Figure 11. RTT delay in SDN-EPON and EPON architectures under different number of end users (16 active ONUs on the left, 8 active ONUs on the right), for different offered traffic load.

The same phenomenon is seen in the queuing delay analysis, whose results are shown in **Figure 12**. As the number of packets per second increases at each ONU, the queuing delay increases too, but much faster for the legacy EPON scenario, and both scenarios suffer a slightly higher delay when the load is concentrated in a few the ONUs.

Figure 13 shows the throughput of both the downstream (left column) and upstream (right column) channels against the offered load for total ONUs, for CBR and self-similar traffic sources, where self-similarity is evaluated under 16 and 8 active ONUs (scenarios I and II, respectively). As shown in **Figure 9**, the throughput is evaluated for the control and data frames traversed through path DE. As expected, the total throughput is very similar, although there is a small advantage for the SDN-based EPON architecture (3.7% better in the downstream, 2.9% better in the upstream), caused by a faster processing of packet is performed in existence of predefined rules, thus lowering the queuing delay. Results show the throughput for both architectures and scenarios (I and II) is quite similar no matter what configuration and Hurst parameters are used. As the number of packets is increased, the average transmission delay and the QoS will be degraded. Again, the traffic pattern does not have any relevant influence in the results.

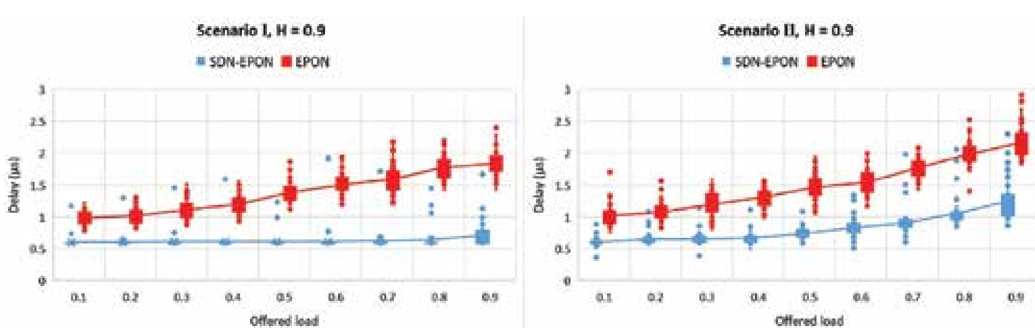

Figure 12. Queuing delay in SDN-EPON and EPON architectures under different number of end users (16 active ONUs on the left, 8 active ONUs on the right), for different offered traffic load.

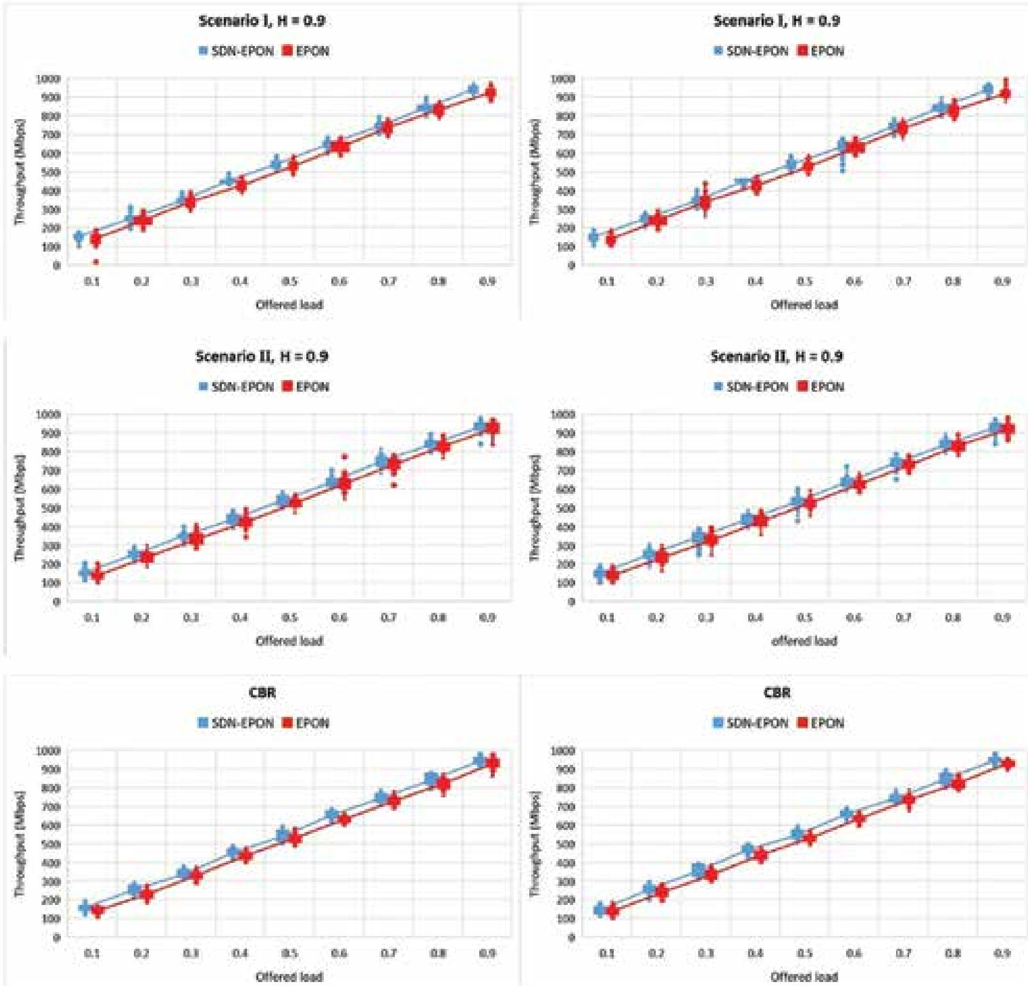

Figure 13. Downstream and upstream throughput in SDN-EPON and EPON architectures under different traffic sources and number of end users (16 active ONUs on the left, 8 on the right), for different traffic patterns (CBR and self-similar) and different offered traffic load.

5. Conclusions

We have described a novel architectural design for SDN-controlled EPON networks, together with system implementation details. Our architecture minimizes the management and operational complexity of the EPON, while optimizing the flexibility and controllability of the network. An extension of MultiPoint MAC Control sub-layer is defined in the OLT to decouple some of its functionalities and distribute them between the SDN controller and the SDN OLT. The management and control functionalities of the MAC client and MAC control client are migrated to the SDN controller, (an example of the operations that take place at longer time scales), and the main functional blocks kept in the SDN-based OLT to be integrated with the OpenFlow switch (operations that work at short time scales are kept at the OLT). Synchronous operations are introduced in the OpenFlow architecture by defining a group of synchronous and non-synchronous ports (via operation of registers), and extending the OpenFlow's rules and messages. Several simulations have demonstrated a significant improvement in data packet delay and downstream and upstream throughput when compared with the legacy EPON architecture.

Among the advantages of our approach we want to emphasize that it opens a lot of possibilities related to the optimization in the resource usage. For example, we envision end-to-end QoS by the coordination of the SDN controllers in charge of the access, metro and core networks. Another important feature of our architecture is the possibility of sharing the same EPON infrastructure among different service providers, by providing separate slices of resources to each provider, under the control and coordination of the SDN controller. This opens the way for multitenant, virtualized EPON networks, with a substantial reduction in OPEX and CAPEX, and the creation of new business models in the field of optical access networks.

Our work is currently focused in the development of power-saving algorithms in the OLT and the ONUs, and its management by the SDN control plane. Power saving can be obtained by setting the laser transmitters and the receivers of both the OLT and ONUs to a sleep state when there is no traffic [28]. The SDN controller could modify the behavior of the DBA and use the GATE/REPORT messages to switch the state of the ONU between sleep, doze, and active mode. The decision could be taken with the information transported in REPORT messages about the queue status of the ONU for the upstream channel, and the OLT queue information for the downstream channel. The action of setting off transmitters affects the traffic patterns by grouping the data frames in burst, thus introducing delays and jitter, and affecting the QoS. That is why a global approach in which the SDN controller has both the control over the QoS and the energy saving, and can reach an appropriate trade-off.

Acknowledgements

This work has been supported by the Ministerio de Economía y Competitividad of the Spanish Government under project TEC2016-76795-C6-1-R and AEI/FEDER.

Author details

Hamzeh Khalili*, David Rincón, Sebastià Sallent and José Ramón Piney

*Address all correspondence to: khalili.hamzeh@entel.upc.edu

Department of Network Engineering, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

References

[1] Kramer G, Pesavento G. Ethernet passive optical network (EPON): Building a next-generation optical access network. IEEE Communications Magazine. 2002;**40**(2):66-73. DOI: 10.1109/35.983910

[2] Cale I, Salihovic A, Ivekovic M. Gigabit passive optical network—GPON. In: 29th International Conference on Information Technology Interfaces, 2007 (ITI 2007); 25-28 June, 2007; Cavtat, Croatia. IEEE; 2007. pp. 679-684. DOI: 10.1109/ITI.2007.4283853

[3] Kramer G, Khermosh L, Daido F, Brwon A, Yoon H, Suzuki K-I, Bo W. The IEEE 1904.1 standard: SIEPON architecture and model. IEEE Communications Magazine. 2012; **50**(9):98-108. DOI: 10.1109/MCOM.2012.6295719

[4] IEEE Standards Association. 1904.1-2013—IEEE Standard for Service Interoperability in Ethernet Passive Optical Networks (SIEPON). Piscataway, NJ, USA: IEEE; 2013. DOI: 10.1109/IEEESTD.2013.6605490

[5] Suzuki K-I. G.epon and Current Status of Related Standardization [Internet]. 2014. Available from: https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201401gls_s.html

[6] Kreutz D, Ramos FMV, Verissimo PE, Rothenberg CE, Azodolmolky S, Uhlig S. Software-defined networking: A comprehensive survey. Proceedings of the IEEE. 2015;**103**(1):14-76. DOI: 10.1109/JPROC.2014.2371999

[7] McKeown N, Anderson T, Hari Balakrishnan MIT, Parulkar G, Peterson L, Rexford J, Shenker S, Turner J. OpenFlow: Enabling innovation in campus networks. ACM SIGCOMM Computer Communication Review. 2008;**38**(2):69-74. DOI: 10.1145/1355734.1355746

[8] Simmons JM. Optical Network Design and Planning. Cham: Springer International Publishing; 2014. p. 529. DOI: 10.1007/978-3-319-05227-4

[9] Thyagaturu AS, Mercian A, McGarry MP, Reisslein M, Kellerer W. Software defined optical networks (SDONs): A comprehensive survey. IEEE Communication Surveys and Tutorials. 2016;**18**(4):2738-2786. DOI: 10.1109/COMST.2016.2586999

[10] Liu L, Tsuritani T, Morita I, Guo H, Wu J. OpenFlow-based wavelength path control in transparent optical networks: A proof-of-concept demonstration. In: 37th European Conference and Exhibition on Optical Communication (ECOC), 2011; 18-22 September 2011; Geneva, Switzerland. Optical Society of America: IEEE; 2011. p. Tu-5

[11] Liu L, Zhang D, Tsuritani T, Vilalta R, Casellas R, Hong L, Morita I, Guo H, Wu J, Martinez R, Muñoz R. First field trial of an OpenFlow-based unified control plane for multi-layer multi-granularity optical networks. In: Optical Fiber Communication Conference and Exposition (OFC/NFOEC), 2012 and the National Fiber Optic Engineers Conference; 4-8 March 2012; Los Angeles, CA, USA. Optical Society of America: IEEE; 2012. p. PDP5D-2

[12] Farias FNN, Salvatti JJ, Cerqueira EC, Abelém AJG. A proposal management of the legacy network environment using OpenFlow control plane. In: IEEE Network Operations and Management Symposium (NOMS), 2012; 16-20 April 2012; Maui, HI, USA. IEEE; 2012. pp. 1143-1150. DOI: 10.1109/NOMS.2012.6212041

[13] Parol P, Pawlowski M. Towards networks of the future: SDN paradigm introduction to PON networking for business applications. In: 2013 Federated Conference on Computer Science and Information Systems (FedCSIS); 8-11 September 2013; Krakow, Poland. IEEE; 2013. pp. 829-836

[14] Ruffin M, Slyne F, Bluemm C, Kitsuwan N, McGettrick S. Software defined networking for next generation converged metro-access networks. Optical Fiber Technology. 2015;**26**:31-41. DOI: 10.1016/j.yofte.2015.08.008

[15] Yang H, Zhang J, Zhao Y, Jialin W, Ji Y, Yi L, Han J, Lee Y. Experimental demonstration of remote unified control for OpenFlow-based software-defined optical access networks. Photonic Network Communications. 2016;**31**(3):568-577. DOI: 10.1007/s11107-015-0547-6

[16] Amokrane A, Hwang J, Xiao J, Anerousis N. Software defined enterprise passive optical network. In: 2014 10th International Conference on Network and Service Management (CNSM); 17-21 November 2014; Rio de Janeiro, Brazil. IEEE; 2014. pp. 206-411. DOI: 10.1109/CNSM.2014.7014203

[17] Khalili H, Rincón D, Sallent S. Towards an integrated SDN-NFV architecture for EPON networks. In: Meeting of the European Network of Universities and Companies in Information and Communication Engineering; 5-8 September 2014; Rennes, France. Advances in Communication Networking, EUNICE: Springer; 2014. pp. 74-84. DOI: 10.1007/978-3-319-13488-8_8

[18] Li C, Guo W, Wang W, Weisheng H, Xia M. Programmable bandwidth management in software-defined EPON architecture. Optics Communications. 2016;**370**:43-48. DOI: 10.1016/j.optcom.2016.02.070

[19] Lee SSW, Li K-Y, Ming-Shu W. Design and implementation of a GPON-based virtual OpenFlow-enabled SDN switch. Journal of Lightwave Technology. 2016;**34**(10):2552-2561. DOI: 10.1109/JLT.2016.2540244

[20] Peterson L, Al-Shabibi A, Anshutz T, Baker S, Bavier A, Das S, Hart J, Palukar G, Snow W. Central office re-architected as a data center. IEEE Communications Magazine. 2016; **54**(10):96-101. DOI: 10.1109/MCOM.2016.7588276

[21] Open Networking Foundation. Open Network Operating System (ONOS) [Internet]. Available from: http://onosproject.org/

[22] Kusano T. Architecture for an Access Network System Management Protocol Control under Heterogeneous Network Management [Internet]. Available from: http://www.patentsencyclopedia.com/app/20140314087 [Accessed: October 23, 2014]

[23] Khalili H, Sallent S, Piney JR, Rincón D. A proposal for an SDN-based SIEPON architecture. Optics Communications. 2017;**403**:9-21. DOI: 10.1016/j.optcom.2017.07.009

[24] Kramer G, Mukherjee B, Pesavento G. Interleaved polling with adaptive cycle time (IPACT): A dynamic bandwidth distribution scheme in an optical access network. Photonic Network Communications. 2002;**4**(1):89-107. DOI: 10.1023/A:1012959023043

[25] Garfias P. Resource management research in ethernet passive optical networks [PhD thesis]. Barcelona: Universitat Politecnica de Catalunya UPC-BarcelonaTech; 2013

[26] Rincón D. Contributions to the wavelet-based characterization of network traffic [PhD thesis]. Barcelona: Universitat Politecnica de Catalunya—Barcelona; 2007

[27] Rincón D, Sallent S. Segmentation of fractal network traffic with wavelets and log-likelihood statistics. In: IEEE International Conference on Communications, ICC 2005; 16-20 May 2005; Seoul, South Korea: IEEE; 2005. pp. 11-15. DOI: 0.1109/ICC.2005.1494312

[28] Argüello JI. Power saving in passive optical networks with distributed bandwidth allocation [MSc thesis]. Barcelona: Universitat Politecnica de Catalunya—Barcelona; 2017

FTTx Access Networks: Technical Developments and Standardization

Krzysztof Borzycki

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71785

Abstract

This chapter provides a review of factors driving technical development of broadband access networks, mostly toward higher bit rates and symmetrical services, together with a review of "fiber to the x" (FTTx) technologies for fixed access networks, including development and performance limitations of digital subscriber line (DSL) systems using twisted-pair copper cables, as well as fiber to the home systems. Characteristics and standardization of these systems are presented, together with a review of the two main competing broadband technologies: Data over cable service interface specification (DOCSIS) in coaxial cable TV networks and the 4G and 5G wireless networks. Additionally, a short list of recent developments in passive technologies (fibers, cables, and connectors) is included. Finally, the issues related to dismantling of the traditional copper telephone network and ensuring continuity of voice services in emergency situations are analyzed.

Keywords: broadband access network, FTTx, FTTH, PON, DOCSIS, standardization, backup power

1. Introduction

Developments of fixed (wired) access networks are driven by services and related demands on network infrastructure, including:

a. Bandwidth adequate for all services, stable performance, and reliability

b. Low cost

c. Continuity of "lifeline" services in emergency situations

Requirement (a) is most important, with bit rate being the differentiator and marketing tool, complemented by consistent performance, including bit rate, error rate, and latency. As the importance of digital services grows, customers and regulators no longer accept "best effort" and "up to" services, but the cost sensitivity of access market (b) forces operators to make compromises between quality and expenditures and reuse existing facilities wherever possible. Currently, available access technologies are capable of providing gigabit services, while streaming of 8 K video needs less than 100 Mb/s. However, with the shift of data storage and processing to "cloud" services, customer requirements will evolve toward even faster and symmetrical access.

Access to basic "lifeline" services in emergency situations (c) is overlooked by operators and customers alike; it comes to light only in times of major disaster like the Sandy hurricane [1].

Finally, there is a possibility of 5G wireless networks largely replacing the wired ones after 5G standards are finalized and equipment becomes available after 2020.

2. Bandwidth demands

The majority of traffic today is generated by video services, characterized by steady increase of resolution, which during the last 15 years rose from 320 × 240 clips and 640 × 480 standard definition (SD) broadcasts to 4 K (3840 × 2160) now and 8 K (7680 × 4320) in the near future. 8 K broadcasts with H.265 coding require a 75–100 Mb/s bit rate, exceeding the requirement of voice service (64 kb/s) by a factor of 1500. Video is predicted to constitute 81% of all consumer Internet traffic in 2021, with the latter growing at an annual rate of 26% [2].

2.1. Nielsen's law and residential Internet access

According to a Nielsen's Law of Internet Bandwidth, the fastest speeds offered to residential customers rise 40–50% each year and even faster recently (**Figure 1**).

Available bandwidth, in fact, rises faster than demands imposed by streaming services. While SD video in 2005 required 2–6 Mb/s, which digital subscriber line (DSL) network barely delivered, full HD needs 10–15 Mb/s today, while 40–100 Mb/s access is available in FTTH, FTTC, and data over cable service interface specification (DOCSIS) 3.0 networks (4, 5, 6). Even the 100 Mb/s required for an 8 K video after 2020 is well below top rates (200 Mb/s–10 Gb/s) in current FTTH and DOCSIS 3.1 networks.

Proliferation of cloud services and social networks after 2010 has also changed the directions of traffic. While watching TV or web browsing required mostly downloading of data, residential users now upload a lot of content they created (photos, videos, scans, etc.); the same applies to professionals and business. Consequently, customers want symmetrical access with equal bit rates in both directions. In this respect, the differences between specific broadband access technologies like XG-PON and XGS-PON (5.2.2) are substantial.

2.2. Is there a limit?

Is there a fundamental limit to demand for faster Internet access?

http://dx.doi.org/10.5772/intechopen.71785

In principle, there should be, because the performance of the human brain and senses (primarily vision) is limited. In particular, the resolution of 4 K and 8 K video already matches or exceeds the capacity of human vision, while adding a 3D and surround capacity will probably increase the streaming data rate by a factor of 2–4x, to some 300 Mb/s. The best audio (192 kHz sampling, 24-bit coding, six channels, uncompressed) needs only approx. 30 Mb/s. Still, there is the need for rapid transfer of massive files with movies, games, or photos. Or, a complete backup of data stored on a PC to a cloud storage. Transfer of 20 TB from a machine with the two largest hard disk drives available now (10 TB) in 5 hours requires 10 Gb/s, exactly the fastest fiber access available in 2017 (see **Figure 1**).

Will subscribers need even more? Probably not, if cloud services replace local storage and streaming replaces file downloading, the mainstream demand can then saturate at 1 Gbit/s.

A significant increase of bit rate above 10 Gb/s will be difficult even in FTTH networks due to high chromatic dispersion of standard single-mode fiber at 1550 nm [4], and the need to use low-cost transceivers with non-return-to-zero (NRZ) modulation and direct detection. With this technology, transmission distance is inversely proportional to the square of bit rate: approx. 80 km at 10 Gb/s, 20 km at 20 Gb/s, and only 5 km at 40 Gb/s. This is a "hard" limit, costly to break (by employing coherent detection and digital dispersion compensation), resembling the situation in commercial aviation, where cost and noise issues associated with breaking the sound barrier keep speeds below 950 km/h [5].

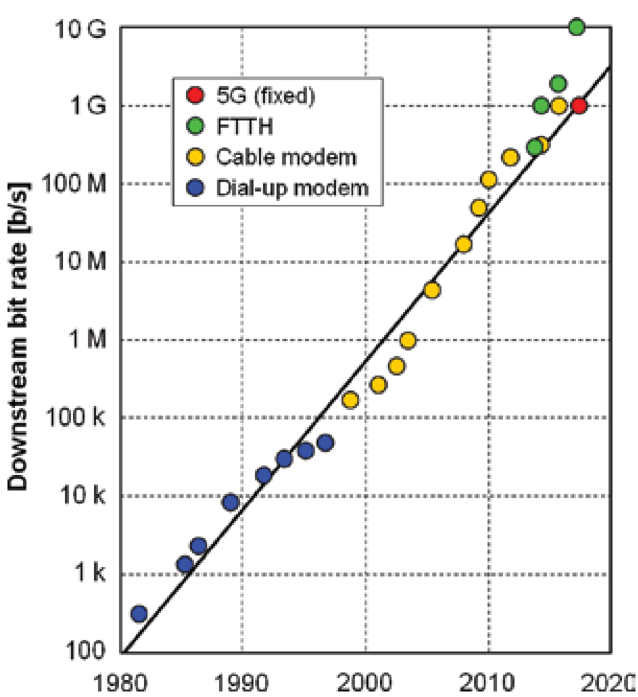

Figure 1. Top download rates offered to residential customers in the USA. (Pre-2012 data from [3]).

3. FTTx access networks

3.1. Reference architecture

The reference architecture of fiber access networks defined by ITU-T Recommendation G.984.1 allows to use both fiber and copper cables as transmission medium, with a possible transition between them in the middle of access loop, as shown in **Figure 2**. Because optical fiber can reach various locations with respect to customer premises, marked as "x," this architecture is called fiber to the x (FTTx).

The termination on customer's side is called either NT/ONT or ONU, e.g., by IEEE [7]. ONU performs termination of optical fiber link to transmit data further, using a medium other than fiber, extending toward the customer; its functionality may include multiplexing of several data streams to and from multiple customers. ONT is located at customer's premises and has interfaces to his devices like PC, TV set, telephone set, etc. In practice, this distinction is blurred, and ONT can have fiber interfaces to a home PC or router or include the router, performing multiplexing of data to/from all devices at home.

3.2. Distances to subscribers

OLT equipment is preferably located in existing central offices (CO), having space, backup power, cable ducts, etc., and is expected to serve all subscribers in associated area. **Figure 3** shows lengths of telephone loops in selected countries before introduction of remote units.

3.3. Transmission media

Optical fibers in access networks are exclusively of non-dispersion-shifted single-mode type, standardized in ITU-T Recommendations G.652 and G.657 [4, 8]. To reduce cabling costs, FTTH networks transmit signals in both directions over a single fiber. LAN cables in FTTB networks have four twisted pairs of 0.5 mm or 0.6 mm copper wires, of which either two or four are utilized for data transmission. Telephone cables have twisted pairs with wire sizes ranging from 0.4 to 0.8 mm, with 0.5 mm most common; a standard telephone loop consists of a single pair.

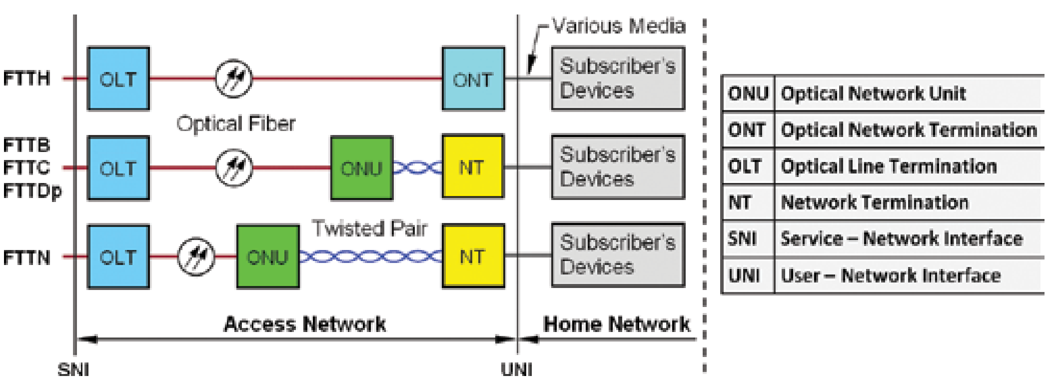

Figure 2. The reference architecture of fiber access network defined by ITU-T [6].

http://dx.doi.org/10.5772/intechopen.71785

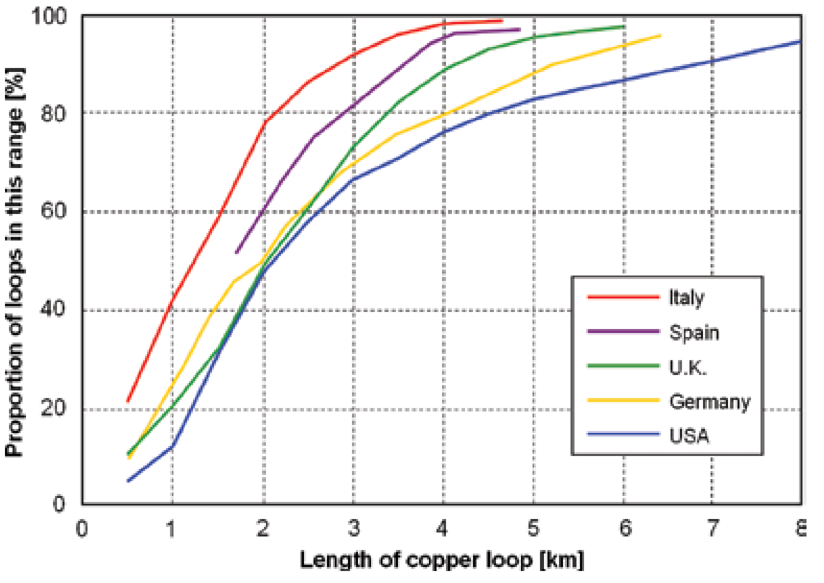

Figure 3. Lengths of subscriber loops in selected countries. Note: There are large differences between data published, probably due to variations of network characteristics with time.

3.4. Classification of FTTx networks

The most common implementations of FTTx networks, listed below and shown in **Figure 4**, differ by the length of remaining copper loop (if any) and splitting of optical fibers (PON) or use of point-to-point fibers (P2P), while the ONU performing media conversion from fiber to

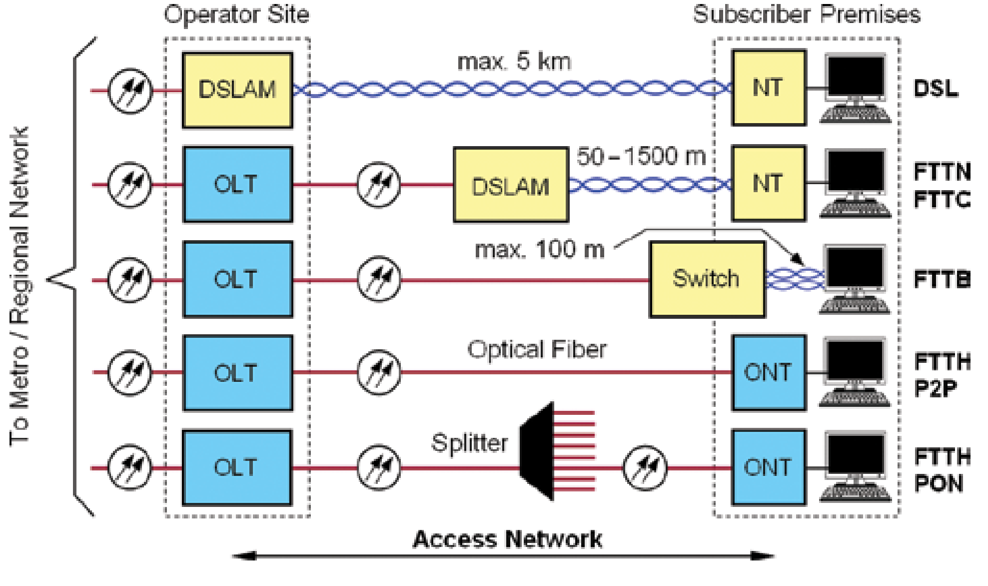

Figure 4. Basic variants of FTTx broadband access networks.

twisted pair is called a digital subscriber line access multiplexer (DSLAM) or Ethernet switch in case of LAN in apartment block.

– FTTB (fiber to the building)

The apartment block is wired with unshielded twisted-pair (UTP) data cables (newly installed), forming a LAN serving all inhabitants, with its Ethernet switch performing termination of fiber link. Data rates up to 1 Gb/s, most often 100 Mb/s.

– Fiber to the curb/cabinet (FTTC)

Optical fibers extend to a remote unit with a DSLAM. From there, short (50–400 m) twisted-pair loops extend to NTs at customer's premises. Data rates up to 100 Mb/s.

– Fiber to the distribution point (FTTDp)

Access network based on G.fast technology, with very short (10–250 m) twisted-pair or coax loops terminated at distribution point unit (DPU) located close to subscriber's premises: at the corridor, on pole, in manhole, etc. Data rates up to 1 Gb/s.

– Fiber to the home (FTTH)

Network with optical fibers extending all the way to ONT at customer's premises, also known as **fiber to the premises (FTTP)**. No active devices in the middle.

– FTTH-PON

FTTH network where a feeder fiber extending from OLT is split into 8–128 distribution fibers reaching customer's premises, forming a passive optical network (PON). OLT bandwidth is shared among all users in a PON with time division multiplexing (TDM).

– FTTH-P2P (point to point)

FTTH network with a separate optical fiber to each customer, no fiber splitting. Each customer is connected to a separate port at OLT. No sharing, data rates up to 1 Gb/s.

– Fiber to the node (FTTN)

Network similar to FTTC, but with longer copper loops, up to approx. 1500 m; fever remote units; and lower data rates up to 20–40 Mb/s.

4. Hybrid networks: FTTN, FTTC, FTTDp, and FTTB

4.1. DSL technology and its limits

Digital subscriber line (DSL) technology was developed to reuse existing copper telephone cables for broadband services, with the minimum use of optical fibers. This approach reduced costs, but limitations imposed by the use of twisted pairs as transmission medium are severe:

a. Attenuation and cross talk rise with frequency, reducing bit rate and reach (**Figure 5**).

http://dx.doi.org/10.5772/intechopen.71785

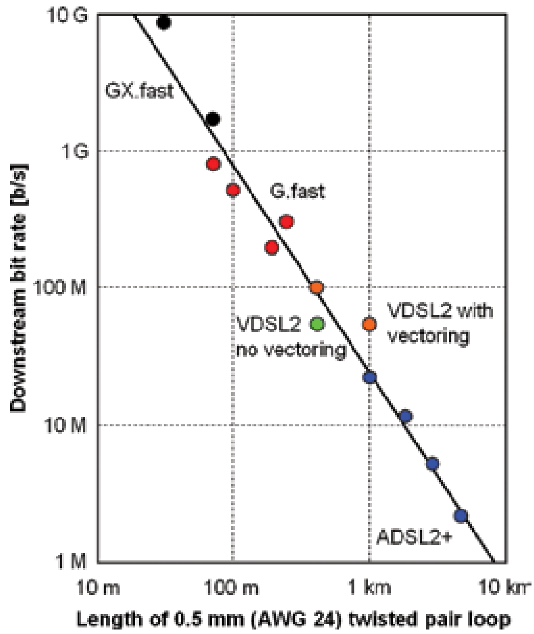

Figure 5. Typical reach vs. bit rate in DSL systems using twisted pairs in telephone cables.

b. Cross talk between adjacent pairs prevents NTs from simultaneously operating at the same frequencies, making the cable a medium with shared bandwidth.

c. Old telephone cables are frequently in poor condition (deteriorated insulation, entry of moisture, deformations), and failures are frequent.

d. Copper cables are often stolen for metal scrap, increasing maintenance costs.

Problem (b) can be eliminated with digital cancelation of cross talk, known as vectoring [9].

Twisted pairs exhibit rise of attenuation and cross talk with frequency, resulting in interdependence between bit rate B and reach L of DSL link, in accordance with a formula:

$$\mathbf{B} = \mathbf{k}\mathbf{L}^{-1.5} \tag{1}$$

where k is a factor dependent on wire diameter, design, and condition of cable, as well as signal processing employed in active equipment. This dependence is presented in **Figure 5**. Higher bit rate requires shorter copper loops and larger number of remote units, which in turn need construction permits and electric power. The acute need for remote units becomes clear after comparing reach values presented in **Figure 5** and distances from the CO to 90 or 95% of customers shown in **Figure 3**.

Development of DSL systems is focused on increasing bit rates, combined with attempts to deliver a good fraction of full capacity at longer distances. Equipment must dynamically adopt to a frequency-dependent, variable transmission characteristics of copper loops, including attenuation, cross talk, and external interference. This is achieved by employing a discrete multitone (DMT) modulation and division of occupied bandwidth into a large number of equally spaced narrow sub-bands, e.g., 2048 in the 2.2–106.0 MHz band with 51.75 kHz spacing in the G.fast system [10]. Only the sub-bands with sufficient signal-to-noise ratio are used, resulting in variable bit rate and "up to" service specifications.

Bandwidth and bit rate data in **Table 1** are maximum values set in standards. Reach achieved in service depends on cable design and its technical condition; data in **Figure 5** and **Table 1** are only indicative.

Performance of DSL link can be improved by pair bonding: the use of two or more copper pairs for parallel transmission, most effective when combined with vectoring.

Comparison of **Figures 1** and **5** suggests that further upgrades of legacy copper networks are not feasible, as system reach becomes too short for most purposes, barring exceptional situations like historical buildings in Europe, where installation of new cables is strongly opposed. This has been already experienced with gigabit version of G.fast system. Despite issuance of standards in 2014 [10, 11], large-scale deployments began only in 2017, and operators prefer 250 Mb/s products with reach of approx. 200 m. However, development continues—a prototype XG-FAST system transmitting 10 Gb/s at a distance of 30 m over one or two pairs, using 500 MHz bandwidth, was demonstrated in 2014 [12, 13].

4.2. FTTB networks

The main advantage of FTTB network (**Figure 4**) over DSL one (4.1) is that the transmission medium between the Ethernet switch and subscriber's premises (a flat in apartment block) is a new Category 5, 6, or 7 unshielded twisted-pair (UTP) data cable, installed in accordance with local area network (LAN) standards. The use of dedicated LAN cables with lengths restricted to 100 m and rated performance up to 100 MHz (Cat. 5), or even 600 MHz (Cat. 7), removes limitations of DSL systems presented in Section 3.2, therefore 100 Mb/s and 1 Gb/s symmetrical services are possible. There is also no need for separate NT devices, as this functionality, at least at 100 Mb/s bit rate, is included in all PCs, laptops, routers, TV sets, etc.

System	Standard	Max. bandwidth occupied (MHz)	Max. bit rate [Mb/s]	Typical reach with a single 0.5 mm pair	Vectoring
ADSL2+	ITU-T G.992.5 (2009)	2.208	24	1600 m @ 20 Mb/s 3000 m @ 10 Mb/s	No
VDSL	ITU-T G.993.1 (2001)	12	52	1200 m @ 40 Mb/s	No
VDSL2	ITU-T G. 993.2 (2006)	30	200	250 m @ 200 Mb/s 800 m @ 50 Mb/s 3500 m @ 4 Mb/s	Optional
VDSL2+	ITU-T G. 993.2 Amd. 1 (2015)	35	300	200 m @ 300 Mb/s	Optional
G.fast	ITU-T G.9700 (2014) ITU-T G.9701 (2014)	106 212 (in the future)	1000	25–70 m @ 1000 Mb/s 100 m @ 500 Mb/s 250 m @ 250 Mb/s	Yes

Table 1. Comparison of DSL systems in use or being introduced. (dates of standards refer to publication of the first edition).

http://dx.doi.org/10.5772/intechopen.71785

How big is the improvement provided by LAN cables? Cat. 7A cable has useful bandwidth of 1200 MHz at distances up to 100 m, supporting a 10 Gb/s bit rate, while the newest Cat. 8 provides 2 GHz bandwidth and is expected to support 40 Gb/s applications, but with cable length reduced to 30 m. The G.fast system with a similar reach (**Table 1**) uses 106 MHz of bandwidth and reaches 1 Gb/s, so the difference in terms of achievable bit rate is 40:1.

Assuming a 100 m reach, allowing to wire all apartments in a 15-floor block to one technical room, FTTB technology can provide 10 Gb/s services at low cost. This is competitive with all FTTH technologies and shall satisfy customer demands for another decade, if the demand trend shown in **Figure 1** has to continue.

5. Fiber networks

5.1. FTTH point-to-point (P2P) networks

Architecture of FTTH-P2P network is simple: the whole path between OLT and ONT (**Figure 1**) is made of one, continuous single-mode fiber, without splitting or intermediate active equipment, resembling the traditional telephone network. There is also no sharing of OLT capacity between multiple ONT units belonging to different subscribers. Full-duplex, symmetrical transmission, with equal bit rates in both directions, is possible due to wavelength division multiplexing (WDM), using the following wavelengths:

– Downstream (to subscriber): 1480–1500 nm (nominal 1490 nm)

– Upstream (from subscriber): 1260–1360 nm (nominal 1310 nm)

The necessary WDM coupler (duplexer) is part of a "single-fiber" optical transceiver module, usually of SFP type. Choice of the 1490 nm wavelength (**Figure 6**) resulted from allocation of 1550 nm in all FTTH networks to analog distribution of TV programs. This was initially necessary because the copyright law in several jurisdictions like Japan and the USA made digital distribution of video content subject to new, burdensome licensing, while a retransmission in original analog format (PAL or NTSC) was covered by existing licenses.

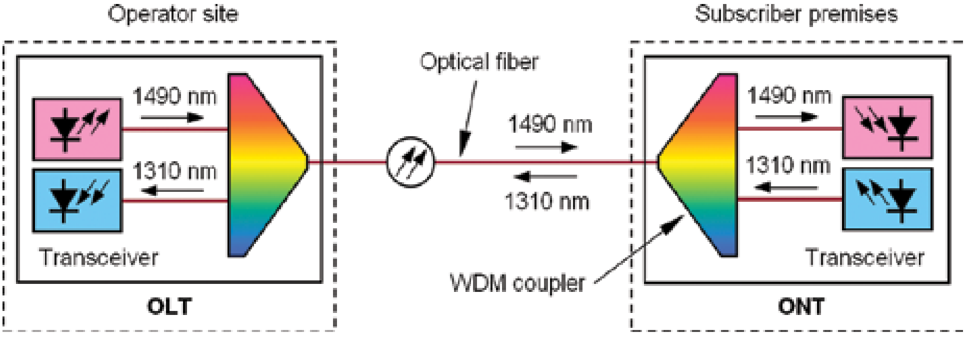

Figure 6. Optical connections and wavelengths in FTTH-P2P network. Optical transceivers are manufactured in separate versions for use in OLT and ONU.

Active equipment for FTTH-P2P networks is based on Ethernet standards [7], and this type of network is known as "active Ethernet." The dominant physical interface is 1 Gb/s 1000BASE-BX10, with a power budget of 5.5/6.0 dB at 1310/1490 nm and 10 km maximum fiber length. This reach corresponds to maximum length of subscriber loops in Europe (**Figure 3**), and conversion of access network from copper to fiber is possible without installing remote units.

More expensive 1 Gb/s transceivers with extended reach of 20, 40, or 80 km are available for use in sparsely populated areas. The 1 Gb/s FTTH-P2P networks are covered also by ITU-T Recommendation G.986 [14], where the power budget of OLT-ONT path in the lowest category S is extended to 15 dB but still combined with a 10 km reach.

10 Gb/s SFP Ethernet transceivers are available but still too costly for use in ONTs. However, prices began to fall after the start of mass production of 10G PON equipment in 2016.

5.2. FTTH-PON networks

5.2.1. Technology

The idea of passive optical network (PON) was first proposed at British Telecom in 1987 [15] to reduce demand for fibers and splicing, both expensive items at the time. As the cost of passive infrastructure of typical FTTH network makes 35–45% of the total [16, 17], this rationale remains valid despite falling prices of optical fiber cables. In a PON there are no active devices between site housing OLT equipment and subscriber premises, as can be seen in **Figure 7**, hence the term "passive." The single feeder fiber extending from OLT optical port is passively split into multiple (8–64) drop fibers extending to each ONT, using a passive, spectrally nonselective, and compact fiber splitter. The whole passive fiber network between OLT and group of ONTs is called an optical distribution network (ODN).

The benefit of splitting is a great reduction in the number of feeder fibers, splicing costs, and duct space for feeder cables. In most cities, duct space is scarce, while construction of new cable ducts

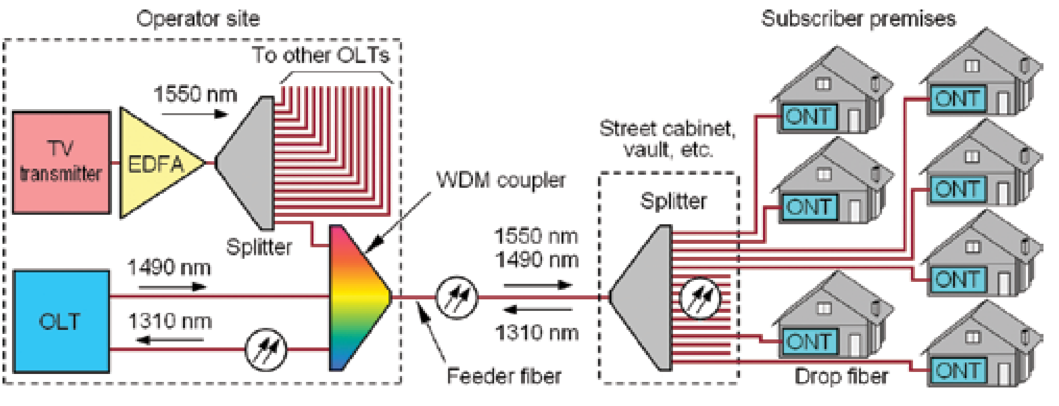

Figure 7. Connections and wavelengths in FTTH-PON access network (GPON). Wavelengths depend on a specific PON system (**Table 4**). Analog distribution of TV signals is optional.

http://dx.doi.org/10.5772/intechopen.71785

is slow and expensive. Savings grow with distance between CO and served residential area, so the FTTH-PON technology is of particular interest to large operators who want to consolidate their access infrastructure into a smaller number of facilities.

Splitting may be executed in two or more stages. For example, the first 1:4 splitter may feed four apartment blocks, while the final splitting into subscriber drops is done by a second splitter in each building, as presented in **Figure 8**. Such arrangement is flexible, and savings on cables are made, but the combined loss of two splitters is larger than of equivalent single device (see representative splitter specifications in **Table 2**).

In general, passive optical networks have three issues absent in P2P networks:

a. High insertion loss of splitter, rising with the split ratio, as shown in **Table 2**.

b. Sharing of OLT bandwidth between all ONTs in a PON.

c. Different transmission delays to each ONT.

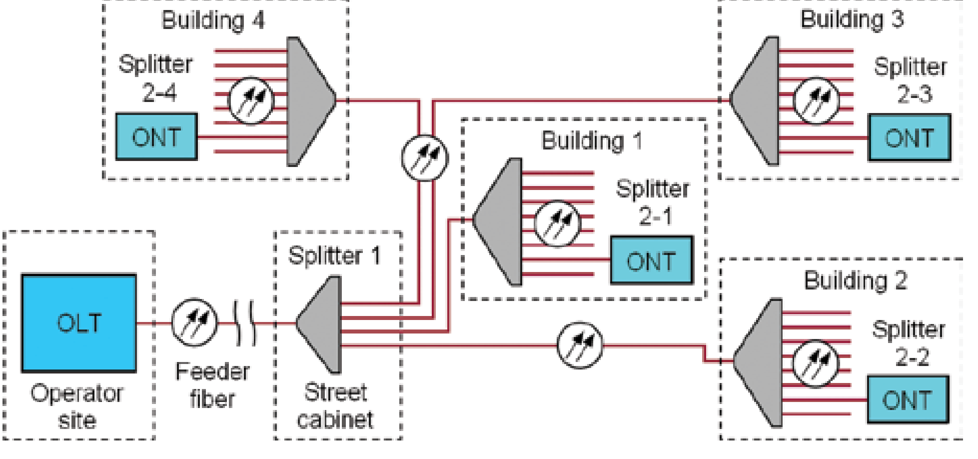

Figure 8. PON covering four apartment blocks with fiber splitting arranged in two stages.

Split ratio	Power division loss (dB)	Actual insertion loss (dB)
1:2	3.01	≤ 4.0
1:4	6.02	≤ 7.3
1:8	9.03	≤ 10.7
1:16	12.04	≤ 13.8
1:32	15.05	≤ 17.0
1:64	18.06	≤ 20.5

Table 2. Insertion loss of fiber splitters: (a) theoretical value calculated for equal division of power between ports and (b) actual data of commercial splitters for the 1260–1650 nm band [18].

Splitter loss rises by approx. 3.5 dB with doubling of the split ratio and is the largest part of loss of the OLT-ONT path. In a typical PON with a 1:32 split, the combined loss of splitter and drop fiber with connectors is approx. 19 dB, while 10 km of feeder fiber with 0.35 dB/km attenuation at 1260 nm, plus a 0.1 dB splice every 2 km, contributes only 4 dB. In a network built in a dense urban environment, splitter loss consumes up to 75% of OLT-ONT loss budget, which, depending on the version of transceivers, is between 20 and 33 dB, of which 2–3 dB shall constitute a margin allocated for equipment aging, repair splices, etc.

Loss calculations are made for the shortest operating wavelength in a given PON (**Table 4**), at which currently manufactured single-mode fibers exhibit the highest attenuation.

As a consequence, upgrade of existing PON to a higher split ratio is hard because there is usually no adequate loss margin, and replacement of transceiver modules with another type having higher loss budget is very costly and requires visits to all subscribers.

Sharing of OLT bandwidth between multiple ONTs is arranged by time division multiplexing (TDM) of data packets sent in both directions. Duration of frame is 125 μs in GPON, XG-PON, and XGS-PON systems and 2 ms in EPON.

Typical PON exhibits large differences in distances from OLT to each ONT and corresponding transmission delays. An exception to this rule is a PON covering one apartment block, with a single splitter and all connections to apartments made with patch cords of equal length. Standards require FTTH-PON equipment to tolerate (compensate) differential fiber distance up to 20 km, corresponding to 200 μs of round-trip differential delay. Transmission delay between OLT and ONT is measured when ONT is activated for the first time; this procedure is known as “ranging.”

5.2.2. Development and standards

Standardization of FTTH-PON systems is carried out by IEEE—as part of Ethernet technology and the Technical Standardization Section of International Telecommunication Union (ITU-T). ITU-T traditionally caters for needs of large operators, with considerable attention given to network monitoring and management. ITU-T Recommendations G.988 [19] and G.997.1 [20] cover management of FTTH-P2P, FTTH-PON, G.fast, VDSL2, and ADSL systems in a uniform way, which is important for incumbent operators using both copper and fiber infrastructures. In contrast, the voluminous IEEE 802.3 Ethernet standard [7] is aimed at fast, low-cost implementation in diverse environments, including core and metro networks, data centers, LANs, etc. Management and maintenance of EPON and 10G EPON networks are covered by a separate, relatively new 1904.1 standard [21].

In agreement with demand evolution presented in Section 2, the main direction of development is the increase of system capacity by periodic introduction of new active equipment (**Figure 9**), while the passive fiber infrastructure is retained. A summary is shown in **Table 3**. While high split ratios up to 1:256 are included in several standards, both the need to reuse existing passive infrastructure and rising demand for 1 Gb/s services while OLT operates at 10 Gb/s or 2.5 Gb/s keep them between 1:16 and 1:64.

http://dx.doi.org/10.5772/intechopen.71785

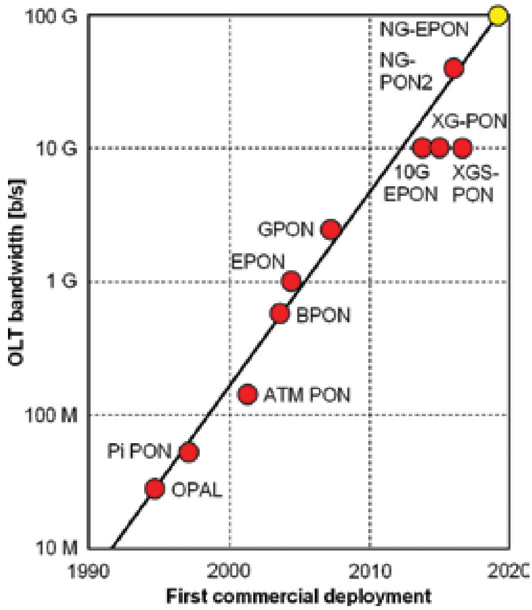

Figure 9. Evolution of OLT bandwidth with time. Date of NG-EPON introduction is estimated.

System	EPON	GPON	10G EPON	XG-PON	XGS-PON	NG-PON2
Standard(s)	IEEE 802.3 [7]	ITU-T G.984 [6, 24, 25]	IEEE 802.3 [7]	ITU-T G.987 [26, 27]	ITU-T G.9807.1 [28]	ITU-T G.989 [29, 30]
Standardized	2004	2003	2009	2010	2016	2010
First deployed	2004	2007	2014	2015	2017	2016
Multiplexing	TDM	TDM	TDM	TDM	TDM	TWDM
OLT downstream bit rate (Mb/s)	1000	2488	10,312	9953	9953	≥4 * 9953
OLT upstream bit rate (Mb/s)	1000	1244	1000 / 10,312	2488	9953	≥4 * 2488
Rate asymmetry	1:1	2:1	10:1/1:1	4:1	1:1	4:1
Reach (km)	20	20	20	40	40	40

Notes: (a) XG-PON is also known as NG-PON1; (b) ITU-T Recommendations are issued as series, e.g., G.987., G.987.1, etc.; (c) all IEEE standards, first separate, are now part of 802.3; (d) date of standardization applies to issue of the first document; (e) GPON capacity data refer to dominant “best practice” variant.

Table 3. Comparison of FTTH-PON systems in use or planned for deployment.

XG-PON, XGS-PON, and NG-PON2 systems are based on a common set of technologies, including management, TDM multiplexing, and ranging; differences are in OLT capacity and optical interfaces. XG-PON was standardized in 2010, when subscribers downloaded a lot of data, watching movies, downloading video clips and music, or browsing, but uploaded much less (except for users of peer-to-peer networks, but ISPs did not like it). At the same time, the XGS-PON project was halted due to lack of interest among major telecom operators.

Development work at ITU-T shifted to further increase of OLT capacity. However, increase of bit rate at a single-carrier wavelength to 40 or 100 Gb/s for transmission over a standard single-mode fiber 20–40 km long requires multilevel coherent modulation and detection, complemented by digital compensation of fiber dispersion [22]. This technology has been successfully implemented in core and metro networks after 2006 but is too complicated and costly for access networks. The dense wavelength division multiplexing (DWDM) technology was adopted instead, with a low number of wavelengths: 4 or 8. The combination WDM and TDM multiplexing is abbreviated as TWDM. To save energy, NG-PON2 includes the option of deactivating some wavelengths during periods of low traffic. However, while WDM transmission in a PON with spectrally non-selective splitter enables "stacking" of multiple physical PONs into a single logical PON of larger capacity, the full use of this feature requires expensive and not yet technically mature transceivers with tunable transmitters and receivers. As in case of WDM-PON (5.2.3), network operators became wary of high costs, not supported by profits achievable in the price-sensitive access market.

By 2015, big operators like Verizon (the first large customer for GPON) requested XGS-PON back, as the demand for symmetrical Internet access grew large and symmetry was important to compete against cable TV operators, whose DOCSIS 3.0 networks could not support it (5.1).

In 10G EPON, the option of having two upstream bit rates mitigated the symmetry issue, but relatively high cost of 10 Gb/s transceivers in comparison to 2.5 Gb/s devices for GPON equipment delayed its deployments, and GPON dominates since 2010. The commercial use of 10G EPON was limited to linking switches in FTTB networks in China until 2016 [23].

IEEE began to work on NG-EPON standard including TWDM multiplexing and Nx25 Gb/s bit rate in 2015. This requires replacing the simple two-level (on/off and NRZ) modulation with more complex solutions to achieve a 20 km reach. A prototype 4x25 Gb/s symmetric PON system developed at Huawei Technologies was tested together with BT Openreach in 2017.

5.2.3. Overlay operation

Upgrade of FTTH-PON running out of capacity or not supporting gigabit services may include gradual conversion to a new standard with "overlay" operation of OLTs and ONTs belonging to two generations in the same ODN. Purchases of new equipment are spread over time, initially to fulfill orders from new customers or for new services, without replacing working equipment at customers unwilling to upgrade.

Overlay operation is ensured by allocation of separate wavelength bands, shown in **Table 4**, for each of the three system generations: (1) GPON, (2) 10G EPON, XG–PON, and XGS–PON, (3) NG–PON2 (see **Table 3**), except for old GPON devices. The 1550–1560 nm band is reserved for analog TV distribution. WDM multiplexers and wavelength blocking filters must be installed at OLT and ONT locations [30], adding combined loss up to 2 dB.

http://dx.doi.org/10.5772/intechopen.71785

System	EPON	GPON	10G EPON, XG-PON, XGS-PON	NG-PON2 (TWDM-PON)
Downstream (OLT to ONT)	1480–1500	1480–1500	1575–1580	1596–1603
Upstream (ONT to OLT)	1260–1360	1290–1330[1]	1260–1280	1524–1544[2]
Distribution of TV programs	1550–1560	1550–1560	1550–1560	1550–1560
Monitoring of fiber network	1625–1675	1625–1675	1625–1675	1625–1675

(1) Initially 1260–1360 nm. (2) ITU-T G.989.2 [30] sets also narrower bands: 1528–1540 nm and 1532–1540 nm.

Table 4. Wavelength bands allocated to FTTH-PON systems (values in [nm]).

This arrangement does not cover 1 Gb/s EPON systems, still having a very wide band allocated for upstream transmission. To avoid collisions, 1G and 10G EPON upstream data must be separated in time domain, but implementation of this feature by equipment manufacturers is not universal.

5.2.4. WDM–PON

In this type of PON, each ONT uses a unique wavelength or a pair of them, and splitter is replaced with a WDM multiplexer. WDM-PON is logically a set of P2P connections between OLT and each ONT at different wavelengths. This eliminates issues [(a)–(c)] in PON networks (Section 5.2.1), and each subscriber can use full bandwidth of OLT port.

Despite trials in South Korea [31], Malaysia, and the Netherlands, WDM-PON was neither standardized nor commercially adopted. The main reason is the cost of transceiver modules, because DWDM technology with 100 GHz channel spacing is required to ensure similar split ratio (1:16 or more) as in FTTH-PON networks and reuse passive infrastructure. The fiber split ratio is doubled with reuse of the same wavelength for transmission to and from ONT and is made possible by adding a high-power broadband light source at the OLT site and re-modulation of spectral slice of this radiation filtered by the WDM multiplexer with upstream data at ONT. However, cost and stability of such system during tests were not acceptable [32].

5.2.5. Comparison of P2P and PON networks

Despite the dominance of FTTH-PON technology in literature, which can be explained by its relative complication and many standards shown in **Table 3**, worldwide deployments are, in fact, split almost 50–50% between P2P and PON networks. Each of them has its merits, and investor must carefully choose the best option in local conditions. Properly installed passive fiber infrastructure, while expensive, has a very long lifetime estimated at 40 years, so the selection of network technology has lasting consequences.

Relative advantages of FTTH-P2P technology include:

– Flexibility: subscriber loops are physically separate; may work with different equipments, wavelengths, bit rates, etc.; and be selectively upgraded as needed.

– Security: data are sent from OLT to one ONT only and are not accessible to others.

– Unbundling: individual fiber loops can be leased to any operator and for use with any type of transmission equipment without restrictions.

– Fibers can be diverted to other uses by the operator, like backhaul and fronthaul links to wireless base stations, Ethernet links to business customers, etc.

– Low loss of subscriber loops, because there is no splitting.

– Distances to customers limited only by power and dispersion budget.

– Simple diagnostics of subscriber loops.

The FTTH-PON option is superior in the following respects:

– Construction costs are lower due to smaller mileage of optical fibers, cable installation costs (pulling, splicing, etc.), and fewer OLT ports; the cost of P2P network can be up to 30% higher in comparison to PON [16], but estimates depend considerably on characteristics of deployment area and local labor costs.

– Lower demand for space in ducts to lay optical fiber cables; additionally, a P2P network in a large city may require cables with very high fiber counts, up to 1000.

– Less space required for OLT equipment and optical distribution frame at the CO.

– Considerably lower power consumption of OLT equipment.

P2P network is not necessarily more reliable in service. The dominant type of network fault in urban or suburban area is a cable cut affecting all fibers in the cable and all customers in the area fed by it, regardless of fiber splitting. However, the repair of large fiber count cable typical in the feeder segment of P2P network takes more time and money. The advantage of P2P network is that a defective, continuously transmitting "rogue" ONT does not jam any other, so only one subscriber loses a service, while in a PON, all customers are affected.

In terms of access speeds, both types of networks (with up-to-date active equipment) can support 1 Gb/s services; upgrade to 2–10 Gb/is possible, although expensive.

Market regulators in Western Europe often display preferences for P2P networks due to possibility of fiber unbundling, support for multiple service providers, and compatibility with the idea of separating wholesale infrastructure operators from retail service providers. However, cost comparisons and shortage of duct space favor the FTTH-PON option, as far as large incumbent operators like Orange, Telefonica, NTT, or Verizon are concerned.

5.2.6. Passive optical LAN (POL)

This is an optical fiber LAN based on GPON, XG-PON, or EPON technology (Section 5.2.1 and 5.2.2). Advantages of POL over LAN with twisted-pair cabling include:

– Elimination of the 100 m distance limit.

– Total immunity to electromagnetic interference.

– Better security (fiber tapping is relatively difficult).

– Light, compact cables (fiber cable, 2–3 mm; UTP cable, 6–7.5 mm).

While the NT in LAN is built into a PC, printer, etc., the ONT in fiber network is usually a mains-powered stand-alone device, typically having four twisted-pair RJ45 data ports (10/100/1000 Mb/s) and a telephone socket.

5.2.7. Cost issues

Provision of consumer Internet access and video services in fixed access networks in most countries is a highly regulated and competitive business, with exceptional price sensitivity and lower profit margins in comparison to mobile or business services.

As shown in **Table 5**, the relative cost of building an FTTx network in area already having a copper network rises with the degree to which the copper loops are replaced with fibers.

The FTTH option involves full replacement of existing copper plant with fiber optics. While the conversion greatly reduces demand for space in cable ducts and buildings, construction is costly and disruptive. A temporary overlay of old and new infrastructure is inevitable, especially when rapid switchover is hampered by regulations protecting customers and alternative operators making use of local loop unbundling. Most money is spent on passive network, with design, cables, accessories, and installation costs constituting 60–75% of the total [33], similarly as was the case with copper telephone network 50 or 100 years ago. The proportion rises with labor costs and restrictions on digging and building access (6.1), putting Western European countries at considerable disadvantage.

Total cost per home passed strongly depends on characteristics of deployment area: population density, terrain, availability of existing ducts or poles, type of buildings, CO locations, rights of way, etc. [34, 35]; published estimates range from 200 up to 6000 EUR. This dependence is much stronger than in wireless networks (7.2), precluding commercial deployment of FTTH networks in sparsely populated areas, despite their excellent performance with loop lengths up to 40 km and even more (5.1, 5.2).

Cost component	New cabling	Remote units	OLT, ONT	Remarks
FTTN	—	**	*	Reuse of existing copper drops, few remote units
FTTC	*	***	*	Many remote units—power and permit issues
FTTDp (G.fast)	*	****	**	Many small remote units, nonstandard locations
FTTB	**	**	*	Switches and new cables in buildings (≤ 100 m)
FTTH-P2P	****	—	****	Full replacement of copper with fiber, very disruptive
FTTH-PON	***	—	****	As above, but less fibers and equipment at the CO

Table 5. Relative cost components of different FTTx access networks (see **Figure 4**). Remote units include Ethernet switches of FTTB networks (4.2) located in apartment blocks.

The FTTH-P2P option, most flexible, supporting multiple operators and easy to upgrade (5.2.5) is also the costliest because no fibers are shared between multiple subscribers. Therefore, market regulators shall support joint investments and cost sharing between operators, taking into account long life (approx. 40 years) of and low operating costs of passive fiber plant. On the contrary, heavy-handed unbundling requirements applied to incumbents in several European countries have stopped FTTH investments and created conditions for proliferation of HFC networks with inferior technical performance (7.1) and hard-to-break monopoly on provision of broadband services.

Cost of passive infrastructure can be significantly (20–30%) reduced with optimized design of cable routes, splitting ratios and points, etc. after analysis of service area [34]. Investors must also analyze demand for services and assume realistic take-up rates to avoid overinvestment, especially in areas with competing service providers. Average take-up rates in Europe range from 10 to 35% and differ widely.

Contrary to big upfront investment in passive infrastructure, most expenditures on active equipment in FTTH networks result from installation of OLT and ONT devices after orders from subscribers are received and are roughly proportional to service take-up rate. Here, the typical cost is between 100 and 400 EUR per subscriber, depending on the system chosen.

Equipment costs tend to dictate selection of FTTH-PON technology. While trials of new systems are reported within months of its standardization, mass introduction often waits before prices drop and demands of customers overwhelm capabilities of older systems. An example is delayed, and selective adoption of 10 Gb/s FTTH-PON systems, standardized in 2009–2010 (**Table 3**), but waiting for large deployments till 2016–2017, when subscribers finally wanted 1 Gb/s services the EPON and GPON networks could not deliver.

Now, this mechanism works in favor of XGS-PON, and against the more advanced and earlier standardized, but some 30% costlier NG-PON2. In the future, it will significantly delay rise of bit rate on a single wavelength above 10 Gb/s (2.2, 5.2.2).

6. Passive fiber technologies for access networks

Here, we focus on changes and progress in this field during the last decade.

6.1. Fibers and cables

FTTx access networks, particularly of FTTH-P2P type (5.1), require large volume of optical fiber cabling, frequently installed in challenging conditions:

a. In crowded underground ducts.

b. In old buildings, where laying cables is considered invasive and restricted.

c. In apartments, where cables are subject to rough handling: sharp bends, stapling, etc.

The primary solution to problem (a) is reduction of fiber and cable diameter. Because single-mode fibers must retain standardized 125 μm cladding diameter [4, 8] for compatibility with existing tools, fusion-splicing machines, connectors, etc. and adequate handling strength, savings are made on protective coatings and cable elements, by:

– Cutting coating diameter from 250 to 200 μm (56% more fibers in the same space).

– Using microcables with thin sheath (approx. 0.5 mm), tight-fitting tubes, marginal strength member, and outer diameter reduced to 1.5–9.6 mm.

– Subdivision of secondary cable ducts (32–40 mm) into 3–12 "microducts," intended exclusively for blowing of fiber microcables.

– Replacement of cable pulling with blowing, using special machines for microcables.

– Using thin (1.2–2 mm) indoor cables where conditions allow.

In 2013, a 9.6 mm microcable with 288 fibers appeared, while an equivalent "traditional" cable with stranded loose tubes 20 years ago had a diameter of 18–20 mm. As a result, four such microcables can be installed in a secondary duct instead of one conventional cable.

Most of these changes are made possible by development and mass adoption of "bending-tolerant" and "bending-insensitive" single-mode fibers. They retain stable attenuation while subjected to macrobending at radius reduced (depending on fiber category and loss limits) to 3–15 mm [8] rather than 30 mm in telecom fibers intended for other applications [4]. This enabled also significant, approx. 50%, size reduction of splice cassettes, joint closures, termination boxes, or distribution frames with LC connectors (4.2).

Problem (b) is partly alleviated by making cables thin or alternatively flat, and in color matching the one of wall, usually white or cream. Installation in "sensitive" areas requires the use of suitable attachment methods to follow the shape of wall corners, door frames, etc. and, if necessary, masking accessories. Still, some 30% of managers of historic properties in Western Europe say "no" to any drilling. But how were the much bulkier telephone and power cables installed there 80 or 100 years ago? Or the water and sewer pipes?

Problem (c) appears when fiber cables entering houses and apartments are hastily installed by poorly trained personnel, stapled to wall corners, or passed through working surfaces of doors or windows (no drilling, please!!!). Because of multiple sharp bends, crushing by staples or pulling, the use of "bending-insensitive" (ITU-T G.657.A2 or B3 [8]) fiber is a must, together with a special 4.5 mm cable of design allowing the fiber to smoothly bend *inside* the cable when its jacket is firmly pressed against a 90° corner or squeezed by a staple.

6.2. Connectors, splitters, and filters

In this area changes are less disruptive, but ones worthy of note include:

– Proliferation of LC connectors with 1.25 mm ferrule (**Figure 10**), smaller, more durable, and less sensitive to the use of lower-grade plastics than SC connectors.

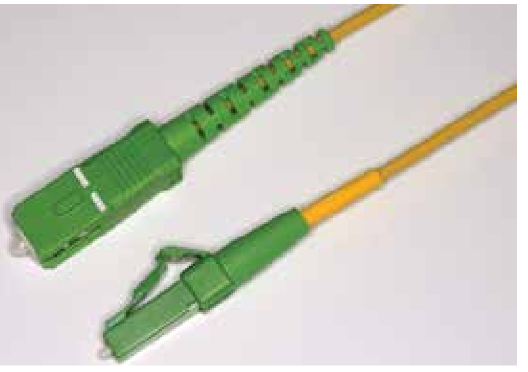

Figure 10. Comparison of fiber connectors: SC (top) and LC (bottom) (couplers not shown).

– Mandatory inspection of connector endfaces before making a connection, with availability of portable multifunction test instruments including ferrule inspection cameras and adoption of IEC 61300–3-35 standard for endface quality [36].

– Steadily rising demand for WDM multiplexers and bandpass blocking or reflecting filters for FTTH-PON networks and monitoring of fibers with a 1650 nm OTDR.

– Abortion of plans to introduce very high-fiber split ratios like 1:256 or more; commercial splitters are available with up to 64 or (rarely) 128 ports (5.2.2).

7. Alternative broadband access technologies

7.1. Coaxial cable TV networks

These originally served for one-way distribution of analog TV programs (since 1948), but were later upgraded, in particular by introduction of return channel for telephone services (1980s) and data services (1990s). Coaxial cable has two main advantages over twisted pairs:

– Wide bandwidth: the frequency range is up to 550–862 MHz in most networks, and 1218 MHz in early implementations of DOCSIS 3.1, being limited rather by available amplifiers than characteristics of the cable.

– Effective shielding against cross talk and external interference.

The attenuation of coaxial feeder cable is high: 40–130 dB/km at 800 MHz vs. 0.25–0.50 dB/km for a single-mode fiber with splices, and periodic amplification is required. Therefore, cable network includes more and more optical fibers in the trunk segment, being known as hybrid fiber-coaxial (HFC) networks.

Typical network in Europe uses the 85–862 MHz (777 MHz wide) band downstream (most of it for distribution of TV and radio programs) and 20–65 MHz (45 MHz wide) upstream.

Standards for data segment of cable TV networks, known as data over cable service interface specification (DOCSIS), are developed by Cable Television Laboratories (CableLabs). Current implementations conform to DOCSIS 3.0 standard, first issued in 2006 [37], while the latest is DOCSIS 3.1 [38] published in 2013.

In DOCSIS 3.0, data are modulated on carriers replacing selected video channels in both directions. The resulting downstream data rate per channel is 38 Mb/s and 50 Mb/s in the USA (6 MHz channels) and Europe (8 MHz channels), respectively, and 27 Mb/s upstream; channel bonding is possible. Maximum capacity is 1216 or 1600 Mb/s downstream and 216 Mb/s upstream. It is shared between all subscribers using the same feeder cable, as in FTTH-PON network.

In DOCSIS 3.1, widening of the upstream band to 204 MHz and downstream band to 1218 or 1794 MHz and spectrally effective QAM-1024 modulation increased the capacity to 10 Gb/s (downstream) and 1 Gb/s (upstream), as in the asymmetric variant of 10G EPON (**Table 3**).

Internet access in cable TV networks is impaired by low upstream capacity; the ratio of upstream and downstream bit rates offered by operators of DOCSIS 3.0 networks is 1:20 or more compared to between 1:10 and 1:1 in FTTH and FTTB networks (5.1, 5.2).

The fastest Internet access provided in DOCSIS 3.1 networks in 2017 was no exception: 35 and 1000 Mb/s or 1:28.6.

This problem is to be eliminated in Symmetrical DOCSIS 3.1 system announced by Intel in 2016, where echo cancelation in a fully passive coaxial node without amplifiers shall allow the use of full frequency range to transmit in both directions at 10 Gb/s [39]. Corresponding CableLabs standard, known as Full Duplex DOCSIS 3.1 (FDX), was issued on October 2017.

Despite upgrades made to DOCSIS, optical fibers and EPON, 10G EPON, GPON, or XG-PON equipment are introduced in many cable TV networks. However, the provisioning of FTTH equipment is performed using DOCSIS-based systems and policies, enabling uniform management of fiber-based and coax-based network elements and services. In such case, the DOCSIS service layer interfaces to media access control (MAC) and physical (PHY) layers of FTTH-PON network. This architecture was developed and standardized since 2013 by CableLabs and known as DOCSIS Provisioning of EPON (DPoE) or GPON (DPoG).

7.2. Broadband wireless access

7.2.1. 4G: LTE and LTE-A

This is the first generation of cellular networks providing a (moderately) reliable Internet access at rates up to 120 Mb/s (LTE, 2 x 20 MHz spectrum) or approx. 300 Mb/s (LTE-A), but such performance is available only for a single user in area served by one sector of base station. With more users, capacity is shared between them, with up to 60% of it being lost.

Despite modest (by the standards of fiber networks) performance, LTE made large impact when introduced around 2012. An interesting case is Japan, a country where FTTH networks exist since 2001 and pass majority of population. However, as can be seen in **Figure 11**, LTE networks eclipsed all fixed networks within 3 years, with a slowdown of demand for new FTTH connections, particularly from young, single, and mobile persons [40, 41]. Growth in new FTTH subscriptions resumed only after the incumbent carriers NTT East and NTT West

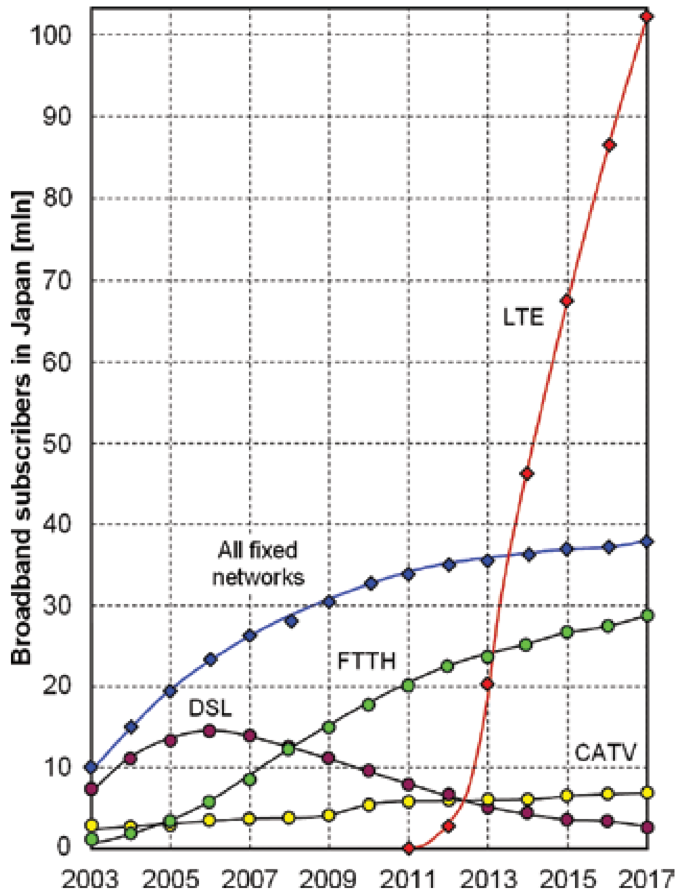

Figure 11. Broadband subscribers in Japan by access technology [41] (data for march of each year).

took the unprecedented step of expressly *inviting* other companies to use their fiber networks within the "Hikari Collaboration Model," and offering dedicated application programming interfaces, technical support, service test environment, etc. in 2014. In 2016, the same companies were the first to introduce software-defined networks (SDN) technology to FTTH networks in a drastic attempt to reduce costs.

7.2.2. 5G: The network of the future?

The set of technologies constituting the fifth generation (5G) of mobile systems, to be standardized in 2020, is not fully defined yet, but performance goals include 1 Gb/s access for a mobile user (with multiple users served by a single base station) and 10 Gb/s for a fixed one [42]. Interestingly, the first use of 5G technology announced by the US operator Verizon for 2018 will be in fixed networks in areas where laying fibers to subscribers is not economical; data rates up to 3.6 Gb/s were reported after trials in 2017.

In the full version, 5G networks will offer mobility and support for unlimited number of user devices without setting up a home network. Lessons from LTE deployments (7.2.1)

http://dx.doi.org/10.5772/intechopen.71785

suggest that after 2025, 5G wireless and its successors can permanently dominate access market, providing universal broadband service after demise of copper networks. The main difficulties in commercial deployments will be high cost and construction of backhaul and fronthaul links; the necessary capital will likely be reassigned from FTTx projects. FTTH and FTTB networks built before will remain in service due to superior quality of service and excellent handling of large volumes of data (8 K video, business applications, etc.), but their expansion is likely to be curtailed.

8. Transition from copper to fiber

8.1. Service continuity in emergency situations

Access to voice services in emergency situations (fire, hurricane, blackout, etc.) is of critical importance for public safety. In a traditional telephone network, central office has backup batteries and generators, while telephone sets are remotely powered over the copper loop.

In FTTx network, the termination installed at subscriber's premises (NT or ONT) is locally powered from the mains. The same applies to cable modem (CM) in a cable TV network.

Several major disasters, in particular hurricanes, recorded in the twenty-first century included loss of mains power for up to 15 days [1], so the communications infrastructure providing basic voice services shall be able to operate without mains power as long as possible.

Remote units in FTTN/FTTC networks (usually located in street cabinets) have batteries providing backup power for 2–8 hours, and traditional telephone sets still work being powered over the twisted pair. However, this duration is not enough in many emergencies.

The most acute problem is powering of equipment at customer's premises: NT, ONT, or CM.

These devices usually consume 5–25 W and are typically powered by 12 V DC from a mains adapter. Backup power can be provided by a rechargeable (sealed lead-acid) or non-rechargeable (alkaline) battery; ONT provides battery charging and monitoring. A typical 5 Ah lead-acid battery can provide power for approx. 8 h. A bank of eight D (LR20) alkaline batteries may power the ONT for 24 hours and can be stored for approx. 10 years.

To prolong operation on backup power, ITU-T G.988 standard for management of GPON, XG-PON, XGS-PON, and NG-PON2 networks [19] includes several options for selective deactivation of nonessential functions to reduce energy consumption after loss of mains power. Alas, most network operators treat backup power as a cost item rather than a safety feature. In the USA, the country frequently affected by hurricanes, twisters, and earthquakes, the telecom regulator (Federal Communications Commission) imposed a requirement for an 8-h backup power at ONTs since 2015, and 24 h beginning from 2018 [43], but installation is not mandatory. It is made at the request of a customer, and charges apply.

Replacement of rechargeable battery is required after approx. 5 years. It usually has to be done by the subscriber, after the operator remotely detects signs of battery aging and mails a replacement unit; the old battery is sent back for recycling.

All this provides continuity of voice service, if the subscriber still has a traditional, wired phone attached to the ONT. The portable, wireless phone does not work during mains failure, as the docking station has no backup.

8.2. Remote powering

DSLAMs and other active equipments located in remote units, usually street cabinets, can be powered from the central office or other large operator's sites via a bundle of free copper pairs in existing telephone cable, working at DC voltages up to 800 V with automatic current limiting to 60 mA for safety (RFT-C system). System working on a bundle of ten pairs with 0.5 mm wires can deliver 250 W of power at distances up to 5 km [44].

Another option is laying a special hybrid cable, comprising both copper power conductors of larger size and optical fibers, with a metallic screen over the core. In this case, a dedicated DC power system is used, with power conductors insulated from the ground, known at Isole Terre (IT) or remote feed for telecommunications with limited voltage (RFT-V) and operating voltage up to $\pm$200 V. For electrical safety, the system can detect a leakage between power conductors and ground, e.g., when the cable is damaged [45]. The RFT-V system normally uses twisted pairs from a telephone cable and has power output limited to 100 W per circuit.

Remote powering has attracted considerable interest of network operators in Europe after 2010, as it allows to avoid costly and cumbersome local powering from commercial AC power network and greatly improves network resilience to power failures. However, there is no consensus on the best technology yet.

Remote powering is implemented in G.fast systems (4.1), where the DPU is jointly reverse powered by all active NTs. However, this arrangement is aimed at simplifying DPU installation and does not protect against mains failure.

Power over Ethernet (PoE) is standardized for LANs [7], where a DC power (up to 57 V and 960 mA) and data share the same copper pairs, with each power circuit utilizing the difference between common mode voltages in two pairs ("phantom circuit"). PoE technology can, in principle, be used to provide interruptible power to NTs in FTTB networks (3.4, 4.2), ensuring continuity of voice service.

8.3. Dismantling of legacy copper plant and regulatory issues

The fate of copper cable plant and conventional telephone switching systems is sealed:

- Telephone cable networks in developed countries are old (40–80 years) and deteriorated, with frequent failures and rising maintenance costs.
- Most existing copper loops do not support true broadband services (4.1).
- The number of subscribers to traditional telephone services has been falling since 2000 due to migration to mobile networks or VoIP services; in the USA, the proportion of households with wired phone fell from 93% in 2003 to 25% in 2013 [46].
- Servicing and spare parts for TDM telephone switching systems are often no longer available, the equipment is fully deprecated, and technical specialists are retiring.

- Wired networks are saddled with extensive regulatory obligations, primarily universal service and local loop unbundling the incumbent operators want to get rid of [47].

Retaining copper network is rejected due to added maintenance costs and the need to vacate cable ducts for optical fiber cables. From the operational point of view of incumbent operator, the migration toward FTTH or LTE/5G wireless is inevitable. The only issues are:

- Deadline, estimated as approx. 2025.

- Type of new infrastructure: fiber or wireless (or both, chosen depending on area).

- Response of market regulators and consumers.

Large-scale dismantling of copper network has already begun in Spain, Portugal, and the USA. In the last country, the largest wireline operators AT&T and Verizon have reportedly stopped maintenance of copper networks around 2005, including open refusals to perform any repairs [48]. As a result, deteriorating quality of service and frequent outages force customers to migrate to a fiber or wireless network.

9. Summary

There are several competing fixed broadband access technologies, and the case for FTTH networks is not made simpler by existence of two standard bodies, ITU-T and IEEE. The winner may ultimately be the 5G wireless, offering the mobility most customers want.

Author details

Krzysztof Borzycki

Address all correspondence to: k.borzycki@itl.waw.pl

National Institute of Telecommunications, Warsaw, Poland

References

[1] Johnston D. Sandy Aftermath: How to fix a broken system. Newsweek, 11-11-2012. http://www.newsweek.com/sandy-aftermath-how-fix-broken-system-63815

[2] Visual Networking Index (VNI) - Forecast Highlights 2016–2021. San Jose, CA, USA: Cisco Systems, Inc.; 2016. https://www.cisco.com/c/m/en_us/solutions/service-provider/-vni-forecast-highlights.html

[3] Cloonan T. Bandwidth trends on the Internet. A cable data vendor's perspective. IEEE 802.3 Ethernet Bandwidth Assessment Teleconference, 08-09-2011, http://www.ieee802.org/3/ad_hoc/bwa/public/sep11/cloonan_ 01a_0911.pdf

[4] ITU-T G.652: Characteristics of a single-mode optical fibre and cable. ITU Telecommunication Standardization Sector. Geneva, Switzerland. (11/2009)

[5] Borzycki K. Development of FTTH-PON technologies: Market reality check 2010-2014, Proc. Networks-2014, Funchal, Madeira, Portugal, Sept. 17–19, 2014, thu.s16. 4

[6] ITU-T G.984.1: Gigabit-capable Passive Optical Networks (GPON): General characteristics. (03/2008)

[7] IEEE Std 802.3-2015: IEEE standard for Ethernet. New York, NY, USA: IEEE Computer Society LAN/MAN Standards Committee; Sept. 3, 2015

[8] ITU-T G.657: Characteristics of a bending-loss insensitive single-mode optical fibre and cable for the access network. (10/2012)

[9] ITU-T G.993.5: Digital sections and digital line system – Metallic access networks: Self-FEXT cancellation (vectoring) for use with VDSL2 transceivers. (01/2015)

[10] ITU-T G.9700: Fast access to subscriber terminals (G.fast) – Power spectral density specification (04/2014)

[11] ITU-T G.9701: Fast access to subscriber terminals (G.fast) – Physical layer specification (12/2014)

[12] Maes J, Spruyt P, Vanhastel S. G.Fast breaks through the gigabit barrier. Nokia Insight, 09-07-2014. https://insight.nokia.com/gfast-breaks-through-gigabit-barrier

[13] Coomans W, et al. XG-FAST: The 5th generation broadband. IEEE Communications Magazine. 2015;**12**:83-88

[14] ITU-T G.986: 1 Gbit/s point-to-point Ethernet-based optical access system. (01/2010)

[15] Faulkner DW, Payne DB, Stern JR. Passive optical telephony networks and broadband evolution. Proc. GLOBECOM '88, Hollywood, FL, USA, 28 Nov.-1 Dec. 1988

[16] Soria B, Hernández-Gil F. The Competitive and Social Implications of Mandating P2P Technologies for FTTH Networks. Madrid, Spain: Telefónica, S.A.; May 2011 http://dx.doi.org/10.2139/ssrn.1798002

[17] Jay S, Neumann K-H, Plückebaum T. Comparing FTTH Access Networks Based on P2P and PMP Fibre Topologies. Bad Honnef, Germany: WIK-Consult; April 2011

[18] Passive Optical Network Products 2012/13 (catalog), Fibrain

[19] ITU-T G.988: ONU management and control interface (OMCI) specification. (10/2012) + Amendment 1 (05/2014) + Amendment 2 (06/2016)

[20] ITU-T G.997.1: Physical layer management for digital subscriber line transceivers (11/2016)

[21] IEEE Std 1904.1-2013: IEEE standard for service interoperability in Ethernet passive optical networks (SIEPON). New York, NY, USA: IEEE Communications Society; 14-06-2013

[22] Roberts K, et al. Performance of dual-polarization QPSK for optical transport systems. Journal of Lightwave Technology. 2009;**27**(16):3546-3559

[23] Shen Ch. FTTx in China – Current status and future prospects, Joint ITU/IEEE Workshop on Ethernet - Emerging Applications and Technologies, Geneva, Switzerland, Sept. 22, 2012

[24] ITU-T G.984.1: Gigabit-capable passive optical networks (GPON): General characteristics. (03/2008)

[25] ITU-T G.984.2: Gigabit-capable Passive Optical Networks (G-PON): Physical Media Dependent (PMD) layer specification. (03/2003) + Amendment 1 (02/2006)

[26] ITU-T G.987.1: 10-Gigabit-capable passive optical networks (XG-PON): General requirements. (03/2016)

[27] ITU-T G.987.2: 10-Gigabit-capable passive optical networks (XG-PON): Physical media dependent (PMD) layer specification. (10/2010)

[28] ITU-T G.9807.1: 10-Gigabit-capable symmetric passive optical network (XGS-PON). (06/2016)

[29] ITU-T G.989.1: 40-Gigabit-capable passive optical networks (NG-PON2): General requirements. (03/2013)

[30] ITU-T G.989.2: 40-Gigabit-capable passive optical networks (NG-PON2): Physical media dependent (PMD) layer specification. (12/2014)

[31] Lee C-C, et al. WDM-PON experiences in Korea. Journal of Optical Networking. 2007;**6**(5): 451-464

[32] Kim B, Kim B-W. WDM-PON development and deployment as a present optical access solution. Proc. OFC'2009, San Diego, CA, USA, 22–26-03-2009, OThP5.5

[33] Feng W. The requirements and evolution to next generation optical access network. Proc. OECC-2009, Hong Kong, 13–17-07-2009, MWD8

[34] FTTH Handbook. 7th ed. Fibre to the Home Council Europe, 16-02-2016

[35] FTTH Business Guide. 5th ed. Fibre to the Home Council Europe, 16-02-2016

[36] IEC/EN 61300–3-35: Examinations and measurements – Fibre optic connector endface visual and automated inspection. (2009)

[37] Data Over Cable Service Interface Specifications: DOCSIS 3.0 – Physical Layer Specification CM-SP-PHYv3.0-I08–090121. Louisville, CO, USA: Cable Television Laboratories, Inc.; 2009

[38] Data-Over-Cable Service Interface Specifications DOCSIS 3.1 – Physical Layer Specification CM-SP-PHYv3.1-I08-151210. Cable television laboratories, Inc.; 2015

[39] Shulman S, Kurtz B. Multigigabit Speed and Symmetrical Data Service with Full Duplex DOCSIS. Santa Clara, CA, USA: Intel Corporation; 2017. https://www.intel.com/content/dam/www/public/us/en/documents/white-papers/full-duplex-docsis-cable-paper.pdf

[40] Inoue A, Tsuchiya Y, Saito M, Iwashita M. Demand analysis of internet access Services in Japan: Fixed-line broadband services become needless?, Proc. Networks-2014, Funchal, Madeira, Portugal, 17–19-09-2014, thu.s14.2

[41] Statistics published by Ministry of Internal Affairs and Communications of Japan. http://www.soumu.go.jp./main_sosiki/joho_tsusin/kyousouhyouka/data.html

[42] Wang C-X, et al. Cellular architecture and key technologies for 5G wireless communication networks. IEEE Communications Magazine. February 2014:122-130

[43] FCC 15–98 Report and Order: In the Matter of Ensuring Continuity of 911 Communications, Federal Communications Commission, PS Docket No. 14–174, 08-06-2015

[44] Thiry S, et al. Adaptive remote powering solution for fast changing requirements of telecom networks: A real case experience, Proc. 36th International Telecommunications Energy Conference (INTELEC), Vancouver, Canada, 28 Sept.- 2 Oct. 2014

[45] Stokke BH, et al. High efficiency remote power distribution for telecom infrastructure, Proc. IEEE Telecommunications Energy Conference (INTELEC), Osaka, Japan, 18–22 Oct. 2015

[46] Werbach K. No Dialtone: The end of the public switched telephone network. Federal Communications Law Journal. 2014;**66**(2):203-262

[47] Knutson R. AT&T's plan for the future: No landlines, less regulation. The Wall Street Journal. 07-04-2014

[48] Brodkin J. How Verizon lets its copper network decay to force phone customers onto fiber. Ars Technica, 15-08-2014

Section 3

Practical Aspects of Broadband Networking

Metrics for Broadband Networks in the Context of the Digital Economies

Salman M. Al-Shehri, Pavel Loskot,
Tolga Numanoğlu and Mehmet Mert

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72035

Abstract

In a transition to automated digital management of broadband networks, communication service providers must look for new metrics to monitor these networks. Complete metrics frameworks are already emerging whereas majority of the new metrics are being proposed in technical papers. Considering common metrics for broadband networks and related technologies, this chapter offers insights into what metrics are available, and also suggests active areas of research. The broadband networks being a key component of the digital ecosystems are also an enabler to many other digital technologies and services. Reviewing first the metrics for computing systems, websites and digital platforms, the chapter focus then shifts to the most important technical and business metrics which are used for broadband networks. The demand-side and supply-side metrics including the key metrics of broadband speed and broadband availability are touched on. After outlining the broadband metrics which have been standardized and the metrics for measuring Internet traffic, the most commonly used metrics for broadband networks are surveyed in five categories: energy and power metrics, quality of service, quality of experience, security metrics, and robustness and resilience metrics. The chapter concludes with a discussion on machine learning, big data and the associated metrics.

Keywords: digital transformation, metrics, measurements, performance, broadband networks

1. Introduction

The digital transformation of telecommunication industry alone will likely create $2 trillion new business opportunities and values for the industry as well as the society [1]. This transformation is already bringing profound changes to how telecommunication services are delivered and managed along with changes in the corporate organizations and cultures [2]. The digitalization will also require adoption of new policies and regulatory models. Today's hardware components provide sufficient computing and storage enabling to abstract many processes entirely in software. For instance, software-defined networks (SDN), network function virtualization (NFV), and network analytics provide unprecedented flexibility in configuring the communication services while optimizing the utilization of network resources. The future networks will be completely autonomous, that is, self-organizing, self-healing, self-secure, and self-optimizing. Thus, the broadband networks have become much more than just a telecommunication infrastructure. They are now the backbone of the digital economy, and are envisioned to be enabler of the new business models. The telecommunication industry is developing new solutions including [2]:

- **Application programming interfaces** (APIs) are offering a standardized, controlled, trusted, and secure access to telecommunication services. They support unified access and **over-the-top** (OTT) **services** for different types of users, and can be used for wholesale of connectivity and other applications.
- **Digital platforms** are universal marketplaces connecting digital producers with consumers. Examples use cases are smart city, autonomous vehicles, and ultra-high resolution video streaming services. They resemble operating systems, and are usually hosted in clouds.
- **Digital business models** are often cloudified and formed as XaaS (**anything as a service**) where X can be a platform, infrastructure, network, software, or anything else.
- **Digital ecosystems** are complete solutions and strategies for delivering digital services. They orchestrate architectures, infrastructures, interfaces, policies, and service definitions. They need to be trusted by the stakeholders, even while being highly autonomous and self-configuring.
- **Operations and business support systems** (OSS and BSS) are various, increasingly integrated, policy-driven autonomous systems supporting the key processes, and business models in the enterprises. They include systems for predictive analytics, business intelligence, and customer relationship management (CRM).
- **5G networks** aim to support different application areas (verticals) with vastly different requirements such as enhanced mobile broadband (eMBB), ultra-high reliability and low-latency (uRLLC) networks for the Internet of Things (IoT) devices, and massive machine-type communications (mMTC). The cyber-resilience of these networks is assumed from the onset. However, as of now, the 5G standards are still a work in progress.
- **Digital roadmap** is a digitalization strategy of an enterprise toward fully digital monitoring and control of its processes and operations.
- **Internet of Everything (IoE)** is interconnecting digital platforms, IoT and everything else within different service and industry verticals. The key driver is the customer experience and efficient utilization of the infrastructure and other resources.

- **Edge (fog) computing** distributes computing and digital content centers in the network edges, closers to the consumers in order to reduce latency and load in the core network.
- **Cyber-physical systems** intelligently exploit the connectivity and digitalization to significantly enhance efficiency of the underlying physical systems, and also to improve the business and customer experiences.

The share of revenues by communication service providers (CSPs) within the telecommunication industry has been constantly declining as evidenced by decreasing **average revenue per user** (ARPU) numbers [1]. Large portions of the voice and message revenues have been transferred to the digital content provides offering popular OTT applications. Thus, providing the connectivity and infrastructure as the legacy Internet service providers (ISPs) once did is no longer sufficient. CSPs today need to leverage the connectivity to offer services in new businesses and consumer markets (B2C and B2B). This may be considered to be much more a business transformation than a technology transformation. In fact, the value in connectivity is estimated to be an order of magnitude less than in the digital services [1]. This creates great opportunities for the CSPs to exploit connectivity, agility, cognition, and wealth of data to drive the innovation, customer experiences, and generate new revenues by becoming both digital service providers and enablers. Only then, the CSPs will be able to compete with the traditional content and application providers.

Winning and retaining the customers is and will be even more critical for the business survival. The interactions with customers are now omnichannel. Properly integrating and managing all the interaction channels is crucial for the superior customer experience. The agility and rapid prototyping of new applications (and, as a matter of fact, of the whole businesses) is achieved by modularization and microservices which are accessed via the published APIs. The microservices and APIs are fundamental in building digital platforms in the clouds. The digital platforms substantially reduce the barriers, costs, and times to market for the 3rd party producers to offer goods and OTT services in the web-scale economy as highlighted in **Figure 1**.

The digital transformation is non-trivial, but inevitable for CSPs to remain competitive next to the digital natives such as Google and Amazon who may possibly expand their future activities from computing to also provisioning communication services. The

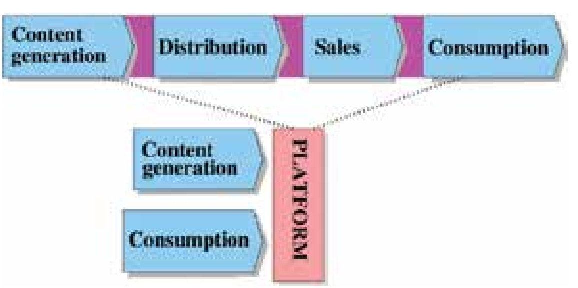

Figure 1. The legacy chain versus the centralized platform business models.

advantage of CSPs is their previous experience in building communication infrastructure, managing the connectivity, and analyzing large volumes of customer and network data. However, the rapid developments of telecommunication systems create many business and technological challenges. It is then important, more than ever before, to carefully consider how to efficiently evaluate, validate, optimize, and orchestrate rather than control these increasingly complex (eco-) systems. Choosing the right metrics and the measurements strategies is critical for this task. The need for standardized metrics for 5G systems is perceived by 21% of CSPs, and 33% of suppliers [2]. The metrics moderate interactions among different stakeholders including equipment manufacturers, subcontractors, infrastructure providers, service providers, content providers, network operators, end-users, governments, and regulatory bodies. The metrics enable to make informed decisions, and to define long-term strategies.

An overview of IT investments, productivity, staffing, and other key business data for enterprises in 21 vertical industries are provided in [3]. The reports such as [1, 3] clearly uncover the cultural changes in high technology companies as their primary focus is moving onto the needs of the customers whereas technology is leveraged for the business growth.

In general, it is desirable that the metrics and measurements exhibit these characteristics [4, 5]:

- **Accuracy:** the measurement errors and biases need to be within acceptable limits.
- **Validity:** the measurements and their evaluations need to be checked for correctness.
- **Feasibility:** the measurements have to be collected as often as desired.
- **Robustness:** the measurements quality must not be affected by changing conditions.
- **Efficiency:** the measurements should not consume too much of system resources.
- **Desirability:** the measurements collected are required for the design and operation.
- **Viability:** the measurements being collected can clearly provide measurable benefits.

The rest of this chapter is organized as follows. Section 2 reviews metrics for digital platforms and ecosystems, websites, and computing systems. Section 3 provides metrics for broadband networks and 4G/5G systems, lists the metrics which have been standardized and which are used for the Internet measurements. Section 4 covers other commonly used metrics for telecommunication networks. Section 5 discusses big data, machine learning, and associated metrics. Finally, Section 6 concludes the chapter.

2. Measuring the digital economy

One of the key objectives in designing technology systems in order to deliver services to end-users is performance. The key measures for evaluating the system performance are: revenues and profits, customer experience, and operational efficiency.

A digital maturity model assesses readiness and effectiveness of the enterprise on its digital journey [2]. It evaluates the digital strategy, understanding of customers, human resources and other assets, processes and operations, and the availability of required technology. It is usually given a score between 1 (initial, ad-hoc improvements) and 5 (a highly optimized digital enterprise) in each of these dimensions. Since the cost of reaching the highest score in all dimensions may be prohibitive, prioritization of objectives, and balancing the costs against the benefits is important. TM Forum [6] developed a sophisticated digital maturity model evaluated over 5 main dimensions (customer, strategy, technology, operations, and organization) with 28 sub-dimensions, and additional 175 criteria questions.

2.1. Metrics for computing systems

Many computing systems are implemented in clouds, and their access is governed by security and demand policies as well as admission rules. There are three main groups of shareholders in this space: the cloud infrastructure and application providers; the 3rd party application and content providers; and the service consumers. The following metrics can be used to measure the technological and business performance of computing systems [7].

Service and system availability is the percentage of time the system is operational, so it can deliver services without degradation. It can be equivalently expressed as the average downtime over a given time period such as month, or year, as the average outage, or as the average time between failures. When the failure occurs, the average **time to recovery** may be useful.

Response reliability is a fraction of the satisfactorily handled requests or service outcomes.

Response time is the average time until the response is received after generating a request. Since the response time may be greatly increased during high demand conditions, the load balancing is critical to provide sufficient scalability of the service provisioning. A similar measure is concerned with the delay to create and configure a new computing instance.

Security threats and incidents detected per a unit of time, for example, in a month, are the indication of both the service attractiveness for an unauthorized use as well as the level of security detection and prevention mechanisms deployed in the system.

Throughput or bandwidth is the number of transactions or requests handled per unit of time, usually a second. It is particularly important for systems and services operating in real-time or at large scale. For real-time services, the average **latency** for repeated requests is also useful.

Capacity or maximum utilization is capability of the system to concurrently handle all or most of the workload requests without a delay, or it is the maximum available computing power or storage space for a single user workload. It can be also expressed as **scalability**, that is, the maximum number of requests served at any given time. On the other hand, **system elasticity** allows scaling the resources to match the total current workload and service demand.

Computing and storage capacity is measured either as the number of processor cores and the memory size available, or at the level of computing units such as the number of virtual servers.

Cost per request, per workload unit, or per user accounts for all supporting, operational, and business processes and any other recurring costs required to provide the agreed services to users. For a well-designed system, this cost is decreasing over time as the system scales up, so the revenues and profits are sustained or improved.

Return on invested capital (ROIC) can be expressed as the number of years until the total profit generated by the offered services exceeds all costs accumulated from the capital investments and the operational expenditures (OpEx).

Market share growth can be indicative of the business viability. If it is declining, the business is normally unsustainable over long-term horizons.

Furthermore, technical metrics for Google Cloud and Amazon Web Services monitoring are listed in [7]. Google also developed the corresponding API to access these clouds and collect the metric values from running application instances. Although the users can define their own metrics to observe, the data can only be collected for the user private projects. The main metric attributes included are the metric name, type, value type, units, and description.

2.2. Metrics for websites

The web analytics provide indication how the website content is perceived by online visitors. It measures a success rather than performance of a website. It plays important role in effectively disseminating information, online marketing, and optimizing traffic and web hosting [8].

Website traffic indicates the trend of online visits, whether it is growing, stagnant, or declining. It can be used to evaluate recent changes in the website content and its efficacy to attract the visitors. It is also useful to breakdown the visitors as new, repeated, returning, and unique ones.

Traffic sources reveal where the visits are coming from. They can arrive via search engines, from another referral, social or other websites, or directly. This metric is often important to evaluate the search engine optimization (SEO) strategy.

Bounce rate is the percentage of visitors only seeing a single page, and *not* exploring other pages on the website except the initial or landing page.

Number of shares on social media is an indirect indication of the page or posted article popularity.

Conversion rate is a ratio of the unique visitors to the number of conversions, that is, those visitors performing a desired action such as visiting a recommended site, subscribing to a service, or purchasing a product. It is therefore one of the most important measures of the website usefulness. Tracking the conversions allows us to evaluate other metrics such as **value per visit** and **cost per conversion.**

Many other more detailed metrics are provided by the website analytics engines, for example, by Google. For instance, **visit duration** is calculated as a time difference between the first and last activity of the visitor, **click through** is the number of times a link was clicked, **exit page rate** tracks the last page visited on the site, and so on. More detailed description of the metrics used for website analytics can be found in [8].

http://dx.doi.org/10.5772/intechopen.72035

2.3. Metrics for digital platforms and ecosystems

Traditionally, Zachman framework has been very popular to manage the complexity of modern enterprises since the late 1980s [9]. It enables a systematic and consistent approach to model complex enterprises from the perspective of different stakeholders. For a set of five principal viewpoints representing different stakeholder interests, there are six metric or model attributes as shown in **Figure 2**. The model of enterprises is then represented as a two-dimensional matrix.

The complexities of emerging digital ecosystems are overwhelming. The number of metrics which need to be considered for complex systems is usually very large. It is then crucial to develop a management system to systematically record, categorize, update, search, and otherwise maintain these metrics. A comprehensive framework covering business processes, applications, and information management is being developed by TM Forum [6]. It is a complete suite of standards and best practices to assess and optimize the performance of digitalized businesses. It is a service-oriented framework which strives to support extensive automations of business processes. It is built on two cornerstones: almost 3000 standardized metrics embraced by the industry, and open-source APIs to support the integration across platforms. This approach to digital ecosystems drives innovation while reducing the risks and shortening time to markets. The improvements in system maturity are directly transformed to a better customer experience, and the reduced costs. The framework metrics are organized in several categories: business, customer experience management, cyber operations, fraud management, and cable operations. The business metrics categories are summarized in **Figure 3** [6]. Note that some metrics have additional attributes, for example, the shopping awareness under the customer experience has the attributes access, time, and quality.

The Frameworx digital ecosystem by TM Forum defines rich metadata for each metric assuming many attributes. The attributes are organized in several sections, and each section contains a number of attribute fields. Examples of the attributes provided for each metric are:

- Overview section gives a summary of the main attributes such as the metric ID, business value driver, capability, reporting details, accuracy, responsible entity, capture period, full name, value type, value range, and metric type.
- Description section is a textual summary of the metric whereas general comments are inserted into a separate section.

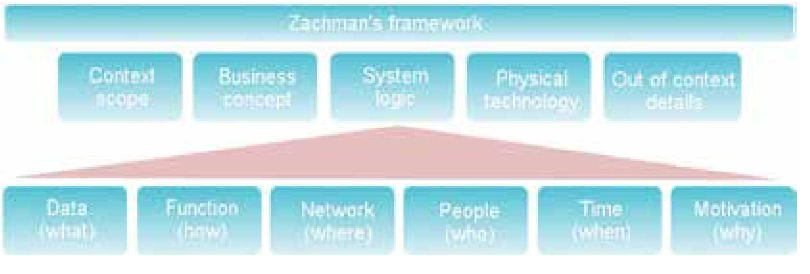

Figure 2. The Zachman modeling framework of enterprises.

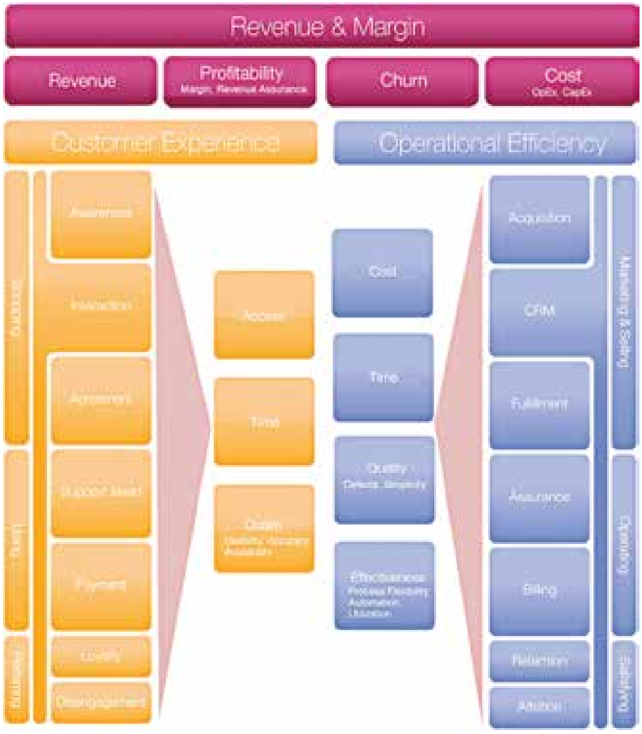

Figure 3. The business metrics categories defined within the Frameworx digital ecosystem by TM Forum.

- Category & Topic section classifies the metric according to its main type, and then three sub-categories depending on the metric objective, specific type, and purpose. Other attributes recorded in this section are the metric level, state, and topic.
- There are other sections to capture the metric mathematical formula and the symbols used.
- Associations section gives details about where the metric is primarily being used. This includes diagrams, other metrics, domains, and documents.

3. Metrics for broadband networks

OECD defines a **broadband service** as having the baseline speed of at least 256 kbit/s [10]. This value has been selected to provide a minimum acceptable quality of service (QoS) for typical broadband applications such as web browsing and VoIP. It is a loose definition which may be less appropriate to compare different broadband technologies, and to improve network performance characteristics. For instance, many countries are introducing mandatory minimums for the broadband access (e.g., 1 Mbit/s in Finland, and 600 kbit/s in downlink and 100 kbit/s in uplink in Switzerland). In practice, broadband measurements provide other data than the access speed, so the definition of broadband may be modified in the future [11]. It is generally

http://dx.doi.org/10.5772/intechopen.72035

agreed that IoT and machine-to-machine traffic is excluded from broadband network measurements, since their traffic profiles are very different from consumers driven traffic, and their traffic is normally restricted and processed locally, for example, within the IoT platforms, so it does not traverse the Internet core or larger parts of the access networks [11, 12].

The roll-out of broadband networks has a direct and measurable impact on the economic development of countries and regions [10]. It stimulates innovations, and has other social impacts such as improved level of the economic productivity, public administration, business practices, health and education. For these reasons, OECD carefully observes developments in the broadband coverage and penetration. OECD defined the broadband checklist including the items directly related to the broadband metrics and measurements such as [10]:

- define broadband services by speed tiers while reflecting national specificities;
- measure the deployment of broadband networks by exploiting the interactive Internet mappings;
- measure the investments in broadband infrastructure, market share, competition among the broadband providers, and also compare the broadband pricing;
- develop a harmonized methodology to measure the broadband speeds, coverage, and capacity;
- explore reliability of the Internet-based statistics to infer traffic flows and web usage patterns;
- develop indicators of the mobile broadband uptakes based on traffic and usage patterns;
- develop new indicators of the broadband demands and of intensity and sophistication of the broadband technology usage patterns;
- develop automated data mining methods to improve learning form the collected broadband statistics; and
- develop appropriate metrics and improve reporting procedures to monitor issues of the security, privacy and consumer protection.

The **speed tiers** of broadband assumed by OECD are: less than 1.5/2 Mbit/s, 1.5/2–10 Mbit/s, 10–25/30 Mbit/s, 25/30–100 Mbit/s, 0.1–1 Gbit/s, and above 1 Gbit/s. Other important actions driven by OECD include measuring the technical and other skills which are required in the digital economy. Moreover, OECD advocates development and use of analytical frameworks and statistical approaches to assess the issues and impacts of the broadband networks on the digital economy and society. In particular, the security and privacy should be assessed across many dimensions including threats, vulnerabilities, incidents, impacts, prevention, and responses. In addition to security and privacy, data on the Internet-related activities incur the costs in order to be of sufficient statistical quality.

In the 2012, OECD workshop on broadband metrics [11], it was recommended to consider the following four categories of metrics for the broadband networks:

- broadband availability metrics and mappings;
- broadband infrastructure investment metrics;
- broadband performance metrics; and
- broadband competition metrics.

We will focus mainly on the broadband performance metrics in the sequel of this chapter. The key performance indicator (KPI) for the broadband Internet access is a **connection speed**. The connection speed affects the quality of experience (QoE), usefulness of applications, and also impacts the broadband policies. However, measuring the connection speed is not straightforward, and common methodologies and best practices to measure the broadband access quality are still subject to many discussions [12, 13]. For now, it has been agreed to classify the connections as wireline and wireless rather than fixed and mobile, and consider whether the broadband users also use or not voice services while their monthly data allowances are less than 500 MB, less than 1GB, less than 5GB, and being more than 5GB. It is further recommended to distinguish between the demand-side and supply-side metrics, and between households, individual consumers, and businesses.

The **demand-side metrics** are obtained from surveys among the regulators and other governmental organizations, and from the data provided by the mobile service providers and other industry stakeholders. From surveys, one obtains **penetration or adoption of broadband services**, applications used, their frequency and usage patterns, socio-economic profile of users (age, gender, education, and income), and the usage issues (e.g., security, pricing, and performance). CSPs can provide data on the number of connections, traffic volumes, usage patterns, and quality of service (QoS) whereas industry stakeholders (e.g., Akamai, Cisco, and Alexa) collect various web statistics. However, the number of subscribers may not be a good indication of the actual broadband usage. Especially for wireless access, the subscribers can have data usage plans which may not be fully utilized. The pricing, tariff plans, and service bundles such as monthly allowances, flat-rates or pay-per-use have significant impact on the broadband usage [14]. It is also possible to exploit crowdsourcing measurements, for example, by downloading a specialized application [15, 16].

The **supply-side metrics** report on the broadband capacity, availability/coverage, speed, and competition. More specifically, **broadband coverage** reflects either the whole region with all broadband providers considered, or it is evaluated for each provider separately. **Broadband network capacity** is the total communication spectrum available, possibly averaged over the regions, or the population while conditioned on the air interface (e.g., 3G/UMTS, long-term evolution (LTE), HSPA, and WiMAX) the technology used (fiber to the premises/cabinet/home, coax, twisted copper cable), and the duplexing method and frequency band used. **Access speed** is one of the most commonly reported metrics on the quality of broadband networks, so we discuss it here in more detail. Similar discussion is generally valid also for other types of performance metrics.

The **access speed** can be either the actual speed measured during the tests, delivered in a day to day use, or it is the advertised (headline) peak or average speed. The advertised and delivered speeds can be vastly different, and each of these speeds can have large variations across the regions, even when the same technology is used. There are obvious speed differences for wireline and wireless access. For the latter, the speed varies greatly as the devices can use one of their wireless interfaces (WiFi and 3G/4G) depending on the context such as mobility. The measurements should be segmented based on the device portability and device type. A typical household has multiple devices sharing the same outgoing connection to the broadband network. The advertised speeds tend to be the theoretical peak vales (the capacity) whereas the measurements are more likely to be the average values. Thus, it is not sufficient to just know the broadband speed, but one has to know what exactly is being reported, that is, peak versus average value, mean versus median average, theoretical versus measured value, end-to-end (Internet) versus the first router (access) speed, when the measurement was performed, and so on. More sophisticated measurement methods can be adopted to average out the measurement noise, for example, using long-term averages of the short-term peaks. Whereas the download speeds were traditionally much larger than the upload speeds, these speeds are becoming more balanced as the users are generating a lot more of their own contents [12]. The speed measurements are usually done with the TCP protocol which can be setup to use one or multiple simultaneous or parallel connections, since a typical webpage generates about 90 requests for content including HTML, CSS, javascript, images, and advertising [17]. However, many of these requests are small pieces of data which are unlikely to reach the full transmission speed of the connection, unlike the transfer of large files over longer time durations which are much more likely to reach the full capacity speed. Thus, the access speed can vary for different applications being used. The broadband speeds for a certain region are typically the values averaged over selected subscribers. How to choose a representative set of these subscribers is another important issue. For instance, sharing of broadband lines between residential customers and businesses significantly influences the speeds observed. The broadband speeds are also strongly affected by the traffic management and traffic shaping policies of CSPs. The fair use policies are especially enforced in the wireless environment. The policies vary the capacity of broadband networks throughout the day, and so they also affect the access speeds. The collected broadband data are often used to compare the offerings of different CSPs, and to formulate the broadband policies, so a consistency in reporting the broadband data is important [16, 18].

Broadband availability in a region is a geographical mapping of the broadband service levels to mainly capture the service providers, the broadband technology in place, and the advertised download and upload speeds. The data can be expressed in different granularity, for example, from whole regions to local municipalities. The most important regional differences in broadband availability are among the urban, sub-urban, and rural areas due to their vastly distinct densities of broadband subscribers. However, the European Commission in its Digital Agenda envision complete (100%) broadband coverage in Europe at 30 Mbit/s or higher by the year 2020.

In order to assess usage of a specific application or service, the user-defined metrics appear to be more common. For mobile internet applications, the following metrics have been defined in [17] to understand the service adoption within a population:

- **Service penetration rate** (SPR) is a fraction of service users among all users in a given time period.
- **Busy hour service attempt** (BHSA) is a fraction of service users during the busy hours among all service users in a given time period.
- **Concentration factor of service attempt** (CSA) measures a concentration of service uses throughout the day.
- **Monthly service activity** (MSA) is a concentration of service uses within days over a month.
- **Service holding time** (SHT) is the average time duration of each service use.
- **Service throughput per usage** (STPU) is the average traffic volume generated during each service use.
- **Time interval of service attempts** (TISA) is the average time interval between two consecutive service attempts by the same user.
- **Net data rate** (NDR) is the average data rate of a service measured at the application layer.

3.1. Metrics for 4G/5G systems

The optimization of broadband networks in 4G/5G systems is driven by an agreed set of KPIs. The KPI monitoring is used both for real-time management of the network as well as for longer term planning of the capacity. The network optimization is dependent on the operator strategy, and the number of systems and sites involved. It targets traffic management, the user experience, and the capital and operational costs. The user experience includes uplink/downlink data rates, and **connectivity** in terms of the coverage, connection setup time, and connection outage or rejection rate. The network planning involves the network coverage and accessibility, **level of congestion** and **traffic volume or density**. The LTE (long-term evolution) standard recognizes at least the following three categories of the KPIs.

Radio-frequency (RF) KPIs are assumed especially during the network roll-out and initial optimizations to achieve the desired coverage with the planned levels of the RF signal strength. These KPIs are usually measured as distributions of the received reference signal strength.

Service KPIs are used to ensure the long-term quality of service and quality of experience targets for different data and voice services as well as to attract and retain the subscribers. These KPIs are usually groups depending on the specific network service such as radio channel access mechanism and type of control messages. The **subscriber retainability** is affected by the handover rate, call drop-off rate, call rejection rate, and other.

Operation KPIs are continuously observed to fine tune the network performance for the current service or traffic demands. These KPIs include the measures of service quality such as data rates, packet loss rates, and packet delays, and the measures of **utilization of the network resources** such as the fraction of resource blocks allocated, module and signaling loads, **cell and channel occupancy rate**, scheduling rates, and other.

The KPIs for 5G systems are defined by the 3GPP standardization body. The aim is to achieve the following improvements against the existing 4G networks: 1000× larger traffic volumes per geographical area, up to 100× more connected devices, up to 100× higher data rates, 10× smaller energy consumption, end-to-end latency at most 1 ms, and ubiquitous coverage also in areas with low density of users. More specifically, the KPIs for 5G networks are defined in three main categories in accordance with the main design goals of the 5G networks.

For eMBB services, the main KPIs are **peak data rate** (20 Gbps in downlink and 10 Gbps in uplink), **expected data rate** (100 Mbps in downlink and 50 Mbps in uplink), and **spectral efficiency** (30 bps/Hz in downlink and 15 bps/Hz in uplink).

For uRLLC services, some of the main KPIs are **maximum latency** (10 ms in control plane, and 0.5 ms in user or data plane), **reliability of packet delivery** (1 lost packet per 100 million transmitted packets), and **connection interruption time** due to mobility (0 ms).

For mMTC services, the important KPIs are **area traffic capacity** (10 Mbits/m^2), **connection density** (1 million/km^2), **energy efficiency** (90% reduction of energy consumption compared to the 4G), **coverage** in terms of the received signal strength (−164 dBm), and the user equipment **battery lifetime** (15 years).

Other metrics for 5G networks which are under development are related to advanced network functions such as network security and network virtualization.

3.2. Standardized metrics for measuring quality of broadband networks

In practice, what really matters is the QoS of broadband networks for CSPs and the QoE for the consumers [19]. We will discuss QoS and QoE metrics more generally in the next section.

ITU-T Y.1540 standard assumes the following metrics to measure QoS over the end-to-end heterogeneous connections:

- **IP packet transfer delay** (IPTD) is the time difference between the ingress and egress packet events when such packet is successfully delivered without errors.
- **IP packet delay variations** (IPDV) are affected by the TCP retransmission mechanisms and may cause undesirable overflow and underflow of packet buffers.
- **IP packet loss ratio** (IPLR) is a fraction of lost packets.
- **IP packet error rate** (IPER) is a fraction of erroneously received packets.
- **IP packet reordered ratio** (IPRR) is a fraction of reordered but otherwise successfully received packets.
- **Spurious IP packet ratio** (SIPR) is a number of spurious packets observed during a specified time interval.
- **IP packet severe loss block ratio** (IPSLBR) is a fraction of the severe loss block outcomes.
- **IP packet duplicate ratio (IPDR)** is a ratio of duplicated packets to the successfully received packets minus the number of duplicated packets.

- **Replicated IP packet ratio** (RIPR) is a ratio of replicated packets to the successfully received packets minus the number of replicated packets.
- **Service availability** is IPLR < 0.75 for at least 5 min duration.

IETF proposes the following five metrics for the end-to-end QoS reflecting the TCP retransmission mechanisms:

- **Link/path bandwidth capacity** (RFC5136) is the overall link/path bandwidth capacity.
- **Bulk transport capacity** (RFC3138) is the bandwidth capacity at the transport layer.
- **One-way and two-way packet losses or connectivity** (RFC2680) is simply the number of packets lost.
- **Packet one-way and two-way delay** (RFC2679, RFC2681) is the end-to-end packet delay.
- **Delay variation** (RFC3393).
- **Packet reordering** (RFC4737).
- **Duplicated packets.**

ITU-T defines six service classes corresponding to the specific values of the metrics listed above. These classes can be mapped to named classes defined in IETF as follows: Class 0/1: telephony, real-time interactive, and multimedia conferencing services; Class 2: signaling; Class 3: low-latency data and high-throughput data; Class 4: broadcast video, multimedia streaming, and low-priority data; and Class 5: standard services. Note that both ITU-T and IETF standards are operator oriented.

The subscriber-oriented QoS may assume additional metrics such as [15, 18]:

- upload and download speed;
- round-trip time (RTT) delay/latency;
- delay jitter;
- packet loss;
- DNS failure rate is a proportion of failed translations of a website name to the IP address;
- DNS resolution is a delay to translate a website name to the IP address;
- web browsing speed is a time it takes to fetch a complete content of a website;
- average daily disconnection is a number of interrupted broadband services per day lasting more than 30 s; and
- distance from the digital exchange indicates the anticipated delays and access speeds.

More importantly, Ofcom evaluates not only the absolute values, but also statistically the probabilities that these metrics are above or below a given threshold:

- the probability of download/upload speed greater than 2 Mbit/s;
- the probability of web browsing loading speed below 1 s; and
- the probability of latency less than 0.1 s.

In addition, Ofcom regularly performs a series of video streaming tests (Youtube, Netflix, and BBC iPlayer) displayed in the standard definition, high definition, and ultra-high definition in order to assess the quality of broadband networks by different CSPs in the UK. In determining the average broadband speeds, Ofcom published the statistical methodology it uses for processing data including the weighting factors to account for different broadband technologies, rural versus urban locations, and varying distances from the network exchange.

Finally, FCC evaluates the broadband networks in states and cities in the US by deploying measuring equipment and using SamKnows methodology [16, 20].

- **Download/upload speed** is throughput in Mbit/s utilizing three concurrent TCP connections.
- **Web browsing** reports the times to fetch a webpage and all its resources from a popular website.
- **UDP latency** is the average RTT of a series of randomly transmitted UDP packets distributed over a long time frame.
- **UDP packet loss** is a fraction of UDP packets lost during the UDP latency test.
- **Video streaming test** measures the initial time to buffer, the number of buffer under-runs and the total time for buffer delays.
- **Voice over IP test** measures the upstream packet loss, downstream packet loss, upstream jitter, downstream jitter, and RTT latency.
- **DNS resolution** is the time for CSP recursive DNS resolver to return a record for a popular website domain name.
- **DNS failures** are a fraction of DNS requests performed in the DNS resolution test that failed.
- **ICMP latency** is RTT of five regularly spaced ICMP packets.
- **ICMP packet loss** is a fraction of packets lost during the ICMP latency test.
- **Latency under load** is the average RTT for a series of regularly spaced UDP packets sent during the sustained downstream/upstream tests.
- **Consumption** is a simple record of the total bytes downloaded and uploaded by the router.

The measurements are obtained for both IPv6 and IPv4 protocols separately. FCC details their statistical data processing methodology including treatment of outliers, adjustment of peak hours to local time, testing setup to avoid congestion and traffic shaping effects, and accounting for the speed-enhancing services.

3.3. Metrics for measuring internet traffic

Measuring Internet traffic is one of the key monitoring tasks for managing and monitoring broadband networks [21]. It is used by CSPs in real-time operations as well as for long-term planning to optimize the network resources, identify anomalies and security issues, and establish traffic control policies, and set the correct levels of service pricing. The processing of traffic measurements is used to classify traffic type, extract more detailed characteristics, and evaluate the statistics [22]. The accuracy and, in some cases, also completeness of such information is critical aspect of CSP business intelligence to remain competitive. There are many challenges in measuring particular traffic. The measurements can be done at or across the flows, or at the level of individual packets, and the measurements can be partially, fully or not at all aware of the underlying protocol used. Unless the measurements are done at the end-points, the router collecting data may not see all packets from the flow as some packets are likely to be routed through a different path. Similarly, many applications use multiple connections to deliver content. The measurements to be meaningful need to check whether more than one connection exists, and then identify which connections belong to the same traffic flow. Moreover, it may be also important to identify different sessions which stemmed from the same application. At a deeper level, examining control information of protocol packets enables to track the protocol state which significantly improves the traffic identification accuracy. Some applications rely on the protocol encapsulation and tunneling, for example, the IPv6 protocol routed through IPv4 sub-network. In this case, the packet inspection needs to look at the payload of the encapsulation protocols. Other challenges in measuring Internet traffic are traffic encryption, for instance, there is increasing preference for using the HTTPS protocol, and traffic is also affected by the deployed proxies.

Some of the **Internet traffic attributes** which are extracted during the deep packet inspection are [21, 22]:

- service tier, content provider, operating system, browser, website;
- IP addresses, MAC addresses, client device, client device type;
- application protocol, media stream type, session protocol, transport protocol;
- video codec, audio codec, media container, video resolution;
- over-the-top application; and
- control versus data content.

4. Other metrics for telecommunication networks

There are many sources of metrics for telecommunication networks [5]. The largest pool of metrics can be found in technical literature. Some of these metrics are much more widely adopted than the others. There are many advantages to create metrics standards in order to ensure consistency and fair comparison of products and services. The metrics standardization is usually driven by industry consortia and telecommunication regulators. In this section, we will review the most commonly used metrics for telecommunication networks. In general, there are often trade-offs between different metric values, so the metrics have to be selected carefully when they are used as KPIs, for instance, to achieve fairness. At a system level, **fairness** evaluates how the network resources are shared and the network capacity utilized among multiple users. We can also evaluate whether a single user is provided a fair access to the network resources and capacity. The most popular measures of fairness are:

- **Jain's index** is independent of the network size, and independent of the measure quantifying how the resources are used by the individual users [23].
- **Max-min fairness** does not allow to increase the resource utilization of one user, if that user has larger utilization of resources than other users.
- **Proportional fairness** is useful in scenarios which incorporates utilization of multiple resources.

In the following subsections, we will review five categories of the most commonly used metrics for telecommunication networks [4].

4.1. Energy and power metrics

The main driver for reducing the energy consumption in telecommunication network is to reduce the operational costs, and to increase the battery lifetime of handheld devices. The measurement procedures including exact location of the measurements, time interval, and conditions such as network load, QoS constraints, and applications are the key factors influencing the measured values of energy consumption.

- **Total energy consumed** is a sum of operational and embodied energies.
- **Operational energy** is the energy consumed during the operation, and it corresponds to the RF power and the overhead power. The overhead power accounts for the baseline circuit consumption at a zero load.
- **Embodies energy** summarizes the energy over the whole life-cycle of the equipment. It includes the energy for manufacturing, transport, installation, decommissioning, and disposal.
- **Energy consumption rating** (ECR) is a ratio of the expended power and the maximum data throughput.
- **Variable-load ECR** is measured as a ratio of weighted averages of the power and the data rates at several values of network load.
- **Energy efficiency rate** (EER) is the inverse value of ECR.
- **ECR for radio access networks** (ECR-RAN) is a ratio of the total expended power in the cell to the cell surface area.

- **Power ratio** of equipment is simply a ratio of the output power to the input power.
- **ATIS energy metrics** are defined as ratios of logarithm of the expended power to the capacity or throughput.
- **ITU metrics** are ratios of the expended power and the product of throughput and distance, for wireline networks, and throughput and area, for wireless networks, respectively.
- **Key power** (not performance) **indicator** (KPI) is a ratio of coverage for rural areas or the number of subscribers for urban areas to the total cell site power, respectively.
- **Transceiver energy consumption** is given by the powers used in transmitting, receiving, scanning, idle, and sleep modes.

4.2. Quality of service

CSPs do not explicitly sell QoS to the subscribers and businesses, but it is included in service packages as the service level agreements (SLAs) at different costs. CSPs may change or update the performance indicators, and still target the same QoS. Guaranteeing QoS is, in general, difficult due to dynamic channel allocation, dynamic routing, energy saving, and fault tolerance mechanisms used. Both passive and active traffic monitoring can be used to assess the level of QoS provided. Since the network resources are shared among many traffic flows, traffic prioritization, and engineering is a dominant method to manage QoS in broadband networks. In both wired and wireless networks, the QoS metrics can be classified as **application-oriented QoS** (AQoS) and **network-oriented QoS** (NQoS). The former is concerned with end-to-end QoS as required, for example, by real-time applications while the latter is concerned with optimizing a core network of routers and switches. Hence, the AQoS metrics can be used to assess the user satisfaction with applications, whereas the NQoS metrics measure the network capability to deliver services while efficiently utilizing the network resources. The application requirements are usually translated to corresponding network characteristics, and they often vary over time.

Most QoS metrics considered in the literature are NQoS metrics such as:

- **Throughput** usually refers to a single flow, and is expressed in bits/s. The throughput can be determined for a single hop, end-to-end connection or aggregated for the whole network.
- **Computing or service throughput** is the ability of a system to process requests and to deliver the work in a given amount of time.
- **Packet delivery ratio** (PDR) is a ratio of the successfully delivered packets to the total number of generated packets. However, measuring the end-to-end packet losses is not straightforward, so the packet losses are usually inferred indirectly.
- **Packet latency** is normally defined as the average end-to-end delay.
- **Delay jitter** is the variance of random end-to-end packet delays. It can be measured for delays in one direction only, or the delays are considered as RTTs. Alternatively, the jitter can

be calculated as the difference between the maximum and minimum RTT values. The jitter can be classified as random/deterministic, correlated/uncorrelated, and constant/transient/short-term.

From the user point of view, **availability** is one of the most important QoS metrics. The users expect that the network is resilient to failures and can provide the agreed services most of the time. More precisely, CSPs standardly aim to achieve 99.999% ("five nines") of the service uptime. Evaluating the availability is not straightforward, since the reliability of many components, systems, and their interactions is at best estimated. Therefore, SLAs usually specify the acceptable downtime and outage periods within a given time interval. The availability can be assumed end-to-end, or at network level. The most commonly used availability metrics are:

- **Mean time to failure** (MTTF) is the expected time to the next failure.
- **Mean time to repair** (MTTR) is the average time to return to operational state after a failure.
- **Mean time between failures** (MTBF) is the average time between two successive failures.
- **Impacted user minutes** (IUM) is a product of the number of users affected by the failure and the failure duration in minutes.
- **Defects per million** (DPM) is a measure of defects of a component or equipment.
- **Point availability** is the probability the system is operational at some future time given the last repair time.
- **Average uptime availability** is a proportion of time the system is ready to deliver the service.
- **Steady-state availability** is expressed as a long-term probability for given rates of failure and repair.
- **Inherent availability** is the steady-state availability considering only corrective downtime.
- **Achieved availability** is the availability assuming only the planned shutdowns.
- **Operational availability** is the average availability assuming all expected downtimes.

4.3. Quality of experience

The QoE metrics evaluate the user satisfaction with the provided level of service. Among challenges to define QoE metrics are: non-linear perception processing by the human senses, lack of accurate models of the human perception, and fast-pace of development of new technologies and services. As the new networks such as the 5G are transforming from the network-centric to user-centric designs, there is a shift from QoS-oriented to QoE-oriented network management. The QoE can be used to determine the required QoS of the network, and how to set adequate pricing levels. However, the relationship between QoS and QoE is complicated, since the improvement in QoS does not guarantee any improvement in QoE. In practice, QoE is evaluated using either subjective or objective metrics. The **subjective QoE** metrics take the human perspective on perception of quality differences through measured statistics, and mainly through the user surveys. However, the surveys are laborious and slow to obtain. Even though the subjective QoE metrics are not generally considered by standardization bodies such as ITU and ETSI for new real-time applications, they are becoming popular among CSPs to accurately forecast the consumer satisfaction.

- **Mean opinion score** (MOS) is a numerical QoE index evaluated through subjective tests, but it ignores some other important aspects such as applications interactivity.
- **Double stimulus continuous quality scale** (DSCQS) is index of video quality which is less sensitive to context, but it is also inefficient for real-time evaluations.
- **Single stimulus continuous quality evaluation** (SSCQE) is more representative for quality monitoring of real-time applications.
- **Absolute category rating**, also referred to as **single stimulus**, optionally with **hidden reference removal** (HRR) is efficient, reliable, and standardized method permitting a great number of test conditions during a single test period.
- **Double stimulus impairment scale** (DSIS) is a paired evaluation of an unimpaired reference video against the impaired video.
- **Single stimulus continuous quality evaluation** (SSCQE) uses a slider device and no standard video.
- **Just noticeable difference** (JND) is a scale obtained by a series of comparison tests on two samples while intensity in one sample increases or decreases.
- **Maximum likelihood difference scaling** (MLDS) measures a relative difference in quality to represent the utility of the tested parameter on visual quality.

The **objective QoE** utilize algorithms, data, and mathematical models to infer the user satisfaction. The data are often QoS measurements which is attractive, since they can be used adaptively and in real-time. The objective QoE can rely, to a different degree, on a reference signal such as a video or an image. The **full-reference** (FR) metrics calculate deprivation of the encoded and subsequently decoded signal pixel-by-pixel. The **reduced-reference** (RR) metrics restrict the comparison to some parts of the signal. In order to completely remove the dependence on a reference signal, the **no-reference** (NR) metrics relies on the ability of human observers to determine the quality of the observed images. The challenges in devising the objective QoE metrics is an unknown dependence on the system parameters, unknown non-linear nature of the human perception, and varying satisfactions of users over time requires that QoE is monitored and updated regularly. The selected QoE metrics for objective evaluation of quality are:

- **E-model** estimates MOS in real-time by computing a so-called R-factor.
- **Perception evaluation of speech quality** (PESQ) model estimates MOS by comparing the observed signal with a reference.

- **Application performance index** (APDEX) evaluates consumer satisfaction on a scale between 0 (no users are satisfied) to 1 (all users are satisfied).
- **Peak signal-to-noise ratio** (PSNR) measures similarity between two different images assuming **mean square error** (MSE).
- **Moving picture quality metric** (MPQM) is a video quality index obtained using a mix of content dependent factors and the network impairments such as packet losses and delays.
- **Motion-based video integrity evaluation** (MOVIE) evaluates video impairments jointly in space and time.
- **Structural similarity index** (SSIM) measures degradation of structural information in video or image such as luminance and contrast.
- **Video quality metric** (VQM) detects human perceivable artifacts in images and video for given codec type, block, and color distortions.
- **Pseudo subjective quality assessment** (PSQA) is a real-time evaluation of quality of video or audio communications over packet-based networks.
- **User satisfaction index** (USI) exploits rigorous analysis of the network level QoS metrics during the call duration.

4.4. Security metrics

The security metrics are ill defined due to lack of sufficiently accurate mathematical models of security. Unlike QoE metrics which can be inferred from the objective QoS measurements, there are no such established baseline metrics which can be objectively measured to infer the security of a system. This is also the reason why there are many more user-defined security metrics than commonly used or even standardized security metrics. Even if the standardization bodies recommend some security metrics and security assessment frameworks, they do not specify any security protection or detection methods, and it is difficult to predict how effective these methods would be. If broadband networks can be made more secure by adopting security policies and procedures, then the security metrics can indicate a relative difference of the perceived security rather than providing the absolute measures. Hence, it is useful to consider a baseline system to evaluate the effectiveness of the implemented security mechanisms, to compare various security strategies, and to decide whether the security policies should be updated. However, comparing security vulnerabilities of the IT technology involving human behavior is challenging, and there are currently no established security metrics combining these two perspectives. A large body of research work considers how to identify the most appropriate security metrics.

Attack graphs are a common tool for modeling system vulnerabilities [24]. They visualize possible progression of the attack from one vulnerability to another. As the more vulnerabilities exist in the system, the more opportunities the attackers have to devise and launch an attack. Due to the size of attack graphs, they can be formed automatically, and they are usually generated for a given host in the network which needs to be protected.

Common vulnerability scoring system (CVSS) is a universal language to describe system vulnerabilities, their urgency, and then prioritize the response and defenses [24]. More importantly, it can be considered as industry standard on defining and accessing the security of systems and products. However, CVSS does not score threats, real-time attacks, or it can manage the security risks. Instead, it combines base, temporal and environmental metrics, and formulas to produce a single security score for the whole system. The base metrics are: access complexity, authentication, and confidentiality, integrity and availability (CIA). The temporal metrics consider time dependency of vulnerabilities, and the environmental metrics are concerned with implementation issues leading to vulnerabilities. The base and temporal scores reflect the severity and urgency, respectively. They are computed and published by equipment vendors whereas the environmental score is computed by users. There are also attempts to further optimize the CVSS scoring to specific systems being considered. Since there are, generally, no widely adopted user-defined security metrics in the literature, we provide a brief list of those which are used somewhat more frequently.

VEA-bility metric (vulnerability, exploitability, and attackability) is a 3D score assessing the network security [25]. It combines impact and temporal assessment of vulnerability expressed as CVSS scoring with exploitability, network topology, and possible attack paths from the attack graph. The scores are evaluated for each network host, and then combined into one final value.

Mean time-to-compromise (MTTC) is the average time required by the attacker to compromise the network [26]. The network compromise event is defined by a set of conditions. The security modeling in this framework assumes attack graphs and probabilities of security-related events.

Relative cumulative risk (RCR) of vulnerability is evaluated as a score combining the individual and neighboring network risks, and measuring the proximity to the untrusted hosts [27].

Hazard metric evaluates the security risks of a network by assessing the network maturity level, frequency of exploits, exploitability impacts, and amendment and authentication levels. The values are then combined into one final score [28].

Security of intelligent electronic devices (IED) calculates susceptibility to each known threat given its countermeasures. The values for each threat are combined into single final score [29].

Critical Vulnerability Analysis Scale Ratings (CVASR) by SANS institute is based on collecting security data using questionnaires [30]. The data are then processed to produce a final 3-level ranking of the potential security threat.

Weakest link security is a popular industrial methodology to evaluate the system security.

There are many other user-defined security metrics such as **Mean-Time-To-Problem-Report** (MTTPR), **Mean-Time-To–Problem-Correction** (MTTPC), problem exposure rates, problem correction rates, problem exploits rates, and so on. The security metrics and the underlying measurements are currently the most active areas of research among all other metrics.

4.5. Robustness and resilience metrics

Robustness is ability of a network to withstand failures, that is, it is a level of **fault tolerance**. For broadband networks, it often means the ability to reroute traffic in case of topology changes or congested links. In general, robustness is achieved by an appropriate network design to create topology having a rich connectivity, and by traffic engineering to utilize this connectivity effectively. The failures can be occasional and random, or targeted and at large scale. For large scale attacks, one may consider the network **survivability**. The most common type of failure dynamics are cascading failures and over-loading attacks. The ability of a network to recover from these adverse events can be quantified by **quality of recovery** (QoR) as an indication of the outage duration.

Resilience is ability of a network to provide the acceptable level of service despite failures. The failures here are either structural related to interactions of the network components or functional which is related to dysfunction of components. The network resilience is intended to be a long-term measure whereas robustness is concerned with short-term conditions. Interestingly, self-organization and autonomy of networks increase the network complexity but also their vulnerability, so their robustness and resilience decrease. Hence, the network robustness and resilience can be also used to infer the level of the network security.

In general, the broadband networks are designed to deliver the desired QoS, and more recently also QoE, and their robustness and resilience are addressed as a subsequent issue. Many resilience metrics are based on the graph-theoretic measures of networks [31]:

- **Node connectivity** is either the smallest number of paths between any two nodes, or the smallest number of nodes whose removal will disconnect the network.
- **Average neighbor connectivity** is the average degree of neighbors of a k-degree node.
- **Heterogeneity** is a measure of robustness of the network topology.
- **Average node degree** is another measure of robustness of the network topology.
- **Symmetry ratio** is a measure of the network functionality response to various attacks.
- **Clustering coefficient** is a measure of the network density as the number of triplets.
- **Average hop-count** is the average shortest paths between all pairs of nodes.
- **Radius** is the length of the shortest path among all shortest paths in the network.
- **Closeness** is measure of node centrality as the mean distance from node to all other nodes.
- **Betweenness** is defined as the number of the shortest paths through a node or a link.
- **Diameter** is the longest of all the shortest paths between all pairs of nodes.
- **Average shortest path length** (ASPL) is the average of all the shortest paths between all node pairs.

- **Algebraic connectivity** is the maximum number of node or link failures a network can tolerate before it becomes disconnected.
- **Natural connectivity** is the redundancy of alternative paths.
- **Weighted spectrum** can be used to identify geographically vulnerable links and nodes.
- **Network criticality** is the surviving ability of the network against topology changes.
- **Effective graph resistance** depends on the presence and quality of backup paths between a given pair of nodes.
- **Path diversity** is the number of disjoint alternative paths between two communicating nodes.
- **Assortativity coefficient** is the correlation of the node degrees.

It should be noted that these graph-based metrics assume a static topology, which may not be the case even for fixed broadband networks, particularly with dynamic route assignment, function virtualization, and slicing of resources.

5. Big data and machine learning metrics

Management of complex heterogeneous telecommunication networks utilizing advanced techniques such as virtualization and network slicing requires more sophisticated strategies to devise the appropriate metrics. The key objective is to automate decision-making in real-time supported by a deluge of data. The big data requires machine learning to enable data analytics. Big data are sourced from many different sub-systems into a common data pool to produce insights which can be utilized by whoever stakeholders need them. Big data is also an enabler of **deep learning**, the most powerful machine learning strategy developed so far. Deep learning is particularly effective for tasks which are difficult to clearly define such as security auditing, business analytics, predicting faults, revenue maximization, configuration and performance optimization, and other. The models of telecommunication networks in machine learning tasks are not explicitly defined, but they are learned and evolve through inflow of data. Ultimately, the network can forecast faults and correct them before they occur, making the broadband network completely self-healing. Another appealing application is the network self-configuration to optimize the performance and utilization of resources.

The popular **5V's** measures for big data are:

- **Data volume** is the amount of data generated per a unit of time.
- **Data velocity** is the speed at which the data are being generated and moved around.
- **Data variety** refers to different sources and types of data.
- **Data veracity** is a level of trustworthiness of data.
- **Data value** is the potential monetary or other valuation of data.

In most tasks, the machine learning algorithms perform prediction from previously learned cases. The quality of these predictions can be evaluated using these metrics:

- **Estimator variance and bias, mean squared error, and scoring function** are the average measures of the estimator quality, provided that the estimator is being used repeatedly.
- **Classification metrics** evaluate loss, score, and utility functions. For instance, **accuracy score** measures the proportion of correct identifications.
- **Binary classification metrics** are usually concerned with false-positives (false alarm, type I error) and false-negatives (type II error). There is usually a trade-off between these two types of error, that is, improving one will deteriorate the other and vice versa. **Sensitivity** (probability of detection) is the proportion of correctly identified positive outcomes whereas **specificity** is the proportion of correctly identified negative outcomes.
- **Regression metrics** measure the performance of regression algorithms such as **mean absolute error**, and **r2 score**.
- **Clustering metrics** measure the performance of clustering algorithms. In particular, they measure whether the clustering defines separations of data similar to some ground truth such that members belonging to the same class are more similar than members of different classes assuming some **similarity metric**.
- **Distance metrics** measure the distances of objects. However, the true distance metrics such as the Euclidean norm must satisfy certain basic properties, which is not always the case.
- **Kernels** are measures of similarity of objects. They are less restrictive, that is, more general, than the distance metrics.

A good overview of various machine learning algorithms with the corresponding performance metrics can be found in the documentation for the Python machine learning library *scikit-learn* [32].

In general, the most powerful machine learning algorithms such as deep learning do not provide justification or explanation for their outcomes. Hence, the network operators are left in the dark to either trust these algorithms or reject their decisions. A partial remedy to this problem is to visualize and evaluate the intermediate outcomes of these algorithms. Development of reliable machine learning algorithms which are provably trustable is a subject of very active research. It is possible that the designers will need to use some machine learning algorithms to validate other machine learning algorithms and their decisions, preferably in real-time.

6. Discussion

In this chapter, we reviewed many different metrics how to evaluate the technical and business performance of broadband networks and the associated technologies. It is clear that the subject of metrics and sensing for the next generation broadband networks has become rather complex. The situation further complicates that broadband networks are a backbone of the digital economy, and that the broadband networks can now rarely be considered separately form the underlying and overlaying technologies. Even though many metrics have been proposed in the literature, only some of them are used much more frequently than the others, so they can be considered de-facto standards. The actual standardization of metrics is mainly driven by large industry consortia. The digital platforms in a production (users-facing) environment require to define and maintain thousands of various metrics which is a managing challenge beyond the resources available even at large university laboratories.

The most active areas of research on metrics for broadband networks and other systems are:

- The digital transformation of broadband networks and CSPs is happening fast. The legacy metrics are often insufficient while new metrics are being introduced at unprecedented speed. The systems and processes are needed for the management of thousands of metrics including maintenance, validation, updating, defining reference values, and so on.
- The broadband network design and management is transiting from the QoS-based metrics to the QoE-based metrics to better align the performance objectives with a customer-centric focus.
- The next generation networks will exploit advanced techniques such as virtualization, softwarization, and network slicing for the maximum flexibility. Furthermore, the mobile broadband access and heterogeneous nature of the next generation broadband networks will require new metrics and management strategies. The traditionally used metrics must be updated to remain relevant, since the underlying assumptions of their use have changed.
- There are a few complete security frameworks while many researchers are proposing their own security metrics to quantify the vulnerabilities and security risks of systems. With better availability of data on threats and vulnerabilities, a set of established simpler security metrics can be expected, for example, as it occurred with the QoS metrics.
- Machine learning algorithms are very powerful, and are being trialed for automated management of broadband networks, and even customer interactions and management. However, the concern is that machine learning algorithms do not justify or explain their outcomes. Active research is ongoing to develop machine learning algorithms which can indicate how their decisions were reached.
- More generally, there is also a need for high performance processing systems for big data analytics.

Author details

Salman M. Al-Shehri[1], Pavel Loskot[1*], Tolga Numanoğlu[2] and Mehmet Mert[2]

*Address all correspondence to: p.loskot@swan.ac.uk

1 College of Engineering, Swansea University, Swansea, United Kingdom

2 Communications and IT Division, Aselsan A.S., Ankara, Turkey

References

[1] Digital transformation initiative: telecommunication industry [white paper]. World Economic Forum; 2017. 44 p

[2] TM Forum. https://www.tmforum.org/

[3] IT key metrics data. 2017: executive summary. Gartner. URL: https://www.gartner.com

[4] Al-Shehri SM, Loskot P, Numanoğlu T, Mert M. Common metrics for telecommunication networks. Technical Report. June 2016. 65 p

[5] Eusgeld I, Freiling FC, Reussner R, editors. Dependability Metrics. Lecture Notes in Computer Science, Springer; 2008. 300 p. ISBN: 978-3-540-68947-8

[6] TM Forum. Frameworx. https://www.tmforum.org/tm-forum-frameworx/

[7] Google. Google Cloud Platform Monitoring. https://cloud.google.com/monitoring/api/metrics

[8] Burby J, Brown A, WAA Standards Committee. Web Analytics Definitions. 2007. 34 p. https://www.digitalanalyticsassociation.org

[9] Zachman JA. A framework for information systems architecture. IBM Systems Journal. 1987;**26**(3):276-292. DOI: 10.1147/sj.263.0276

[10] OECD. Measuring the Digital Economy: a New Perspective. OECD Publishing; 2014. 161 p. DOI: 10.1787/9789264221796-en

[11] OECD. Workshop on broadband metrics. Summary of Recommendations. London, June 2012. 88 p

[12] Loskot P, Hassanien MA, Farjady F, Doran N, Payne DB, Ruffini M, Nesset D, Seton J. Long-term socio-economical drivers of traffic in next generation broadband networks. Annals of Telecommunications. 2015;**70**(1-2):10 p. DOI: 10.1007/s12243-014-0424-9

[13] Sundaresan S, Donato W, Feamster N, Teixeira R, Crawford S, Pescapè A. Measuring home broadband performance. Communications of the ACM. 2012;**55**(11):100-109. DOI: 10.1145/2366316.2366337

[14] Ofcom. UK Home broadband performance: A consumer summary of fixed-line broadband performance provided to residential consumers. Research Report. 2016. 18 p

[15] Ofcom. Measuring mobile broadband performance in the UK. Research Document. 2015. 66 p

[16] FCC. Measuring broadband America. Fixed broadband report. Technical Report and Appendix. 2016. 78 p

[17] Li K, Xu X, Swamy MNS. Modelling and analysis of regional service behavior properties of mobile internet applications. IEEE Access. 2017;**5**:4795-4807. DOI: 10.1109/ACCESS.2017.2684135

[18] Ofcom. UK Home broadband performance: The performance of fixed-line broadband delivered to UK residential consumers. Research Report. 2017. 82 p

[19] Lehr W, Smith-Grieco T, Woo GR. Broadband metrics best practices: Review and assessment. Technical Report. 2008. 89 p

[20] Samknows. UK Broadband Availability. https://www.samknows.com/

[21] Sandvine. Identifying and measuring internet traffic: Techniques and considerations. White Paper. 2015. 20 p

[22] Sandvine. Internet traffic classification. White Paper. 2015. 17 p

[23] Jain R, Chiu D, Hawe W. A quantitative measure of fairness and discrimination for resource allocation in shared systems. Technical Report. DEC-TR-301. 1984. 38 p

[24] Mell P, Scarfone K, Romanovsky S. A Complete Guide to the Common Vulnerability Scoring System Version 2.0 [Report]. NIST; 2007. 23 p

[25] Tupper M, Zincir-Heywood AN. VEA-Bility Security Metric: A Network Security Analysis Tool. In: Proceedings of ARES, Barcelona, Spain; 2008: 8 p. DOI: 10.1109/ARES.2008.138

[26] Suh-Lee C, Jo J. Quantifying security risk by measuring network risk conditions. In: Proceedings of ICIS, Las Vegas, USA; 2015. 6 p. DOI: 10.1109/ICIS.2015.7166562

[27] Singh UK, Joshi C. Quantifying security risk by critical network vulnerabilities assessment. International Journal of Computer Applications. 2016;**156**(13):26-33. DOI: 10.5120/ijca2016912426

[28] Premaratne U, Samarabandu J, Sidhu T, Beresh R, Tan JC. Security analysis and auditing of IEC61850-based automated substations. IEEE Transactions on Power Delivery. 2010;**25**(4):2346-2355. DOI: 10.1109/TPWRD.2010.2043122

[29] Anbalagan P, Vouk M. An empirical study of security problem reports in Linux distributions. In: Proceedings of ESEM, Florida, USA; 2009:481-484. DOI: 10.1109/ESEM.2009.5315985

[30] SANS Institute. Implementing a vulnerability management process. Technical Report. 2013. 24 p

[31] Cholda P, Mykkeltveit A, Helvik BE, Wittner O, Jajszczyk A. A survey of resilience differentiation frameworks in communication networks. IEEE Communication Surveys and Tutorials. 2007;**9**(4):32-55. DOI: 10.1109/COMST.2007.4444749

[32] Scikit-learn. Python machine learning library documentation. http://scikit-learn.org/stable/

Chapter 16

Smart Connected City for Holistic Services

Hyun Jung Lee and Myungho Kim

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.79988

Abstract

The construction of a smart city is based on broadband networks and high-tech under consideration of city infrastructure with holistic city service systems. Digital city was started to connect computing devices using network-based technologies in 1990s. In the beginning of 2000s, many cities were interested in the construction of city infrastructure based on the broadband networks. With the developing high-tech like wireless network, the ubiquitous city was introduced as a new type of an urban city infrastructure to satisfy citizens' needs. These days it would become more important for citizens to provide holistic city services using the transferred data as generated resulting traffics from massive number of end-devices through broadband networks. Smart city has been constructed with multifaceted sectors like high-tech device-based physical and service-based social sector. The integrated sectors are creating new tremendous values based on embedding intelligence in the hyperconnected city. Finally, the smart city should be evolved by centering on people and the creative market is growing up rapidly.

Keywords: smart city, broadband networks, holistic services, embedding intelligence, digital city, ubiquitous city

1. Introduction

The concept of smart connected city has been developed through the era of digital city and ubiquitous city as a new type of city infrastructure. The construction of smart city for providing city services is grounded on high-tech using generated data from smart devices and Internet of Things, which are transferred through broadband networks and processed by information and communication technology. In addition, intelligent and smart systems support to provide holistic civil services for citizens.

According to the development of network-based technologies in 1990s, the digital city based on wired network was introduced and implemented as a kind of a sustainable eco-city, like Amsterdam and Kyoto. At that time, the online commercial environment was also popularized. In addition, the electronic government was started to provide online civil administration services to citizens, such as e-government. In the digital city, unlike in the past, the concept of the city was expanded from the physical to the cyber world. As an instance of services in the digital city, e-government was popularized to provide more comfortable and reliable civil administration services for citizens.

In 2000s, according to the innovative development of mobile technologies, citizens' needs for civil services were more specified, customized, and refined in a variety of fields for their more comfortable urban life using network-based mobile services. Many cities started to build the ubiquitous city based on mobile and using information technologies. In the ubiquitous city, citizens can be connected to networked things using city broadband networks by mobile technology, which allows citizens to access the given and required services whenever and wherever. In the USA and Europe, the ubiquitous city was integrated with the concept of open innovation and expanded to the concept of a "living lab" because it was used to experiment and evaluate the newly introduced concept in the city infrastructure. A project related to the ubiquitous city was introduced as "Smarter Planet" by IBM in 2008. In 2009, IBM announced again the expanded concept of the city as a kind of smarter city to helping cities to efficiently use city resources, to provide comfortable civil services and to improve the quality of life for citizens. The evolving progress of city development is illustrated in **Figure 1**.

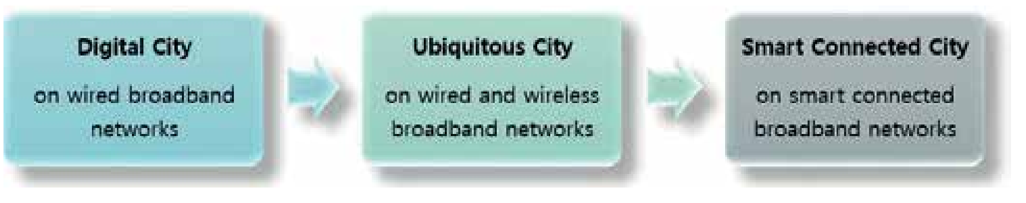

Figure 1. The evolution of cities based on broadband networks.

During the step of the evolution of a city, the digital city was a test bed to apply innovative technologies like online services to the city. The ubiquitous city was a living lab as an infrastructure to support urban services using interconnected high-tech. Finally, the smart city has been developed as a platform of the provision of high-tech, the city infrastructure, and an ecosystem to provide holistic services for citizens as in **Figure 2**.

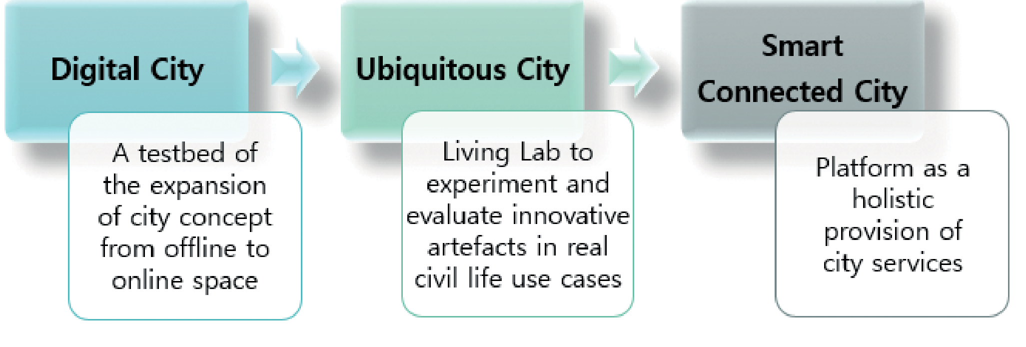

Figure 2. The transitions of city concepts.

http://dx.doi.org/10.5772/intechopen.79988

Depending on the evolvement of a city, many countries are naturally focusing on the construction and development of smart connected cities that can provide holistic services to citizens using all developed high-tech and city infrastructure. Sometimes, a smart city can be a good test bed to implement innovative technologies and a living lab to interoperate and apply the developed high-tech to the city infrastructure to develop and provide useful services for citizens. It is important to implement and apply the developed technologies to citizens' real lives in the city. To achieve this, it is necessary to connect the high-tech infrastructure and to develop and apply appropriate services for the citizens. Finally, the smart city pursues a kind of ecosystem as a platform that brings together a variety of things with hyperconnectivity.

This chapter is organized as follows. In Section 2, we review the smart city trends and market size. In addition, we overlooked the broadband network trends and business impacts with the IoT in Section 3. In Section 4, we introduce the concepts of a smart city, and in Section 5, the smart city types are illustrated. In Section 6, regarding the smart city, there are discussions on smart infrastructure with data and services and the social sectors and values that are created and managed in the city are mentioned. In Section 7, smart city is discussed as a test bed and a smart platform. In Section 8, we illustrate smart city cases in several countries. Finally, we present conclusions.

2. Smart city trends and market size

Many cities are currently concentrating on constructing a smart and connected city with holistic services on broadband networks using high-tech. Over time, high-techs have been rapidly and progressively developed, especially in network-based technologies such as the Internet of Things (IoT), cloud computing, big data, artificial intelligence (AI), and information and communication technology (ICT). According to the customized and specified citizens' needs, the technologies help to serve more comfortable and reliable city services to the citizens in all areas of their lives. The requested and needed services are related to smart city infrastructures such as smarter building, government, work, transportations, information and communications, distribution, and so on that are constructed using high-tech. For instances, as artifacts of the constructed smart city, there are smarter street lighting, trash cans, parking lot services, climate and water services, waterways, smart grid and energy systems, and so on. They generate creative values for citizens using the generated tremendous data through the networked Internet of Things in the city. As in **Figure 3**, McKinsey expected that, until 2025, the allocation level of the Internet of Things will be in a minimum 3.9 trillion and a maximum 11.1 trillion dollars for a year [1]. The Internet of Things will be used to collect data from applied smarter services to the city infrastructure. As major components to construct smart city infrastructure, the illustrations of smarter artifacts in smart city are smart factory, smart health, smart retail, smart worksites, logistics, transportations, home, and so on [1].

These kinds of services need to be integrated, collaborated, and coordinated to provide appropriate services to customers using high-tech. The required services should be interconnected and intertwined to provide holistic services to citizens. The holistic services in the smart city are based on high-tech like the embedded intelligence.

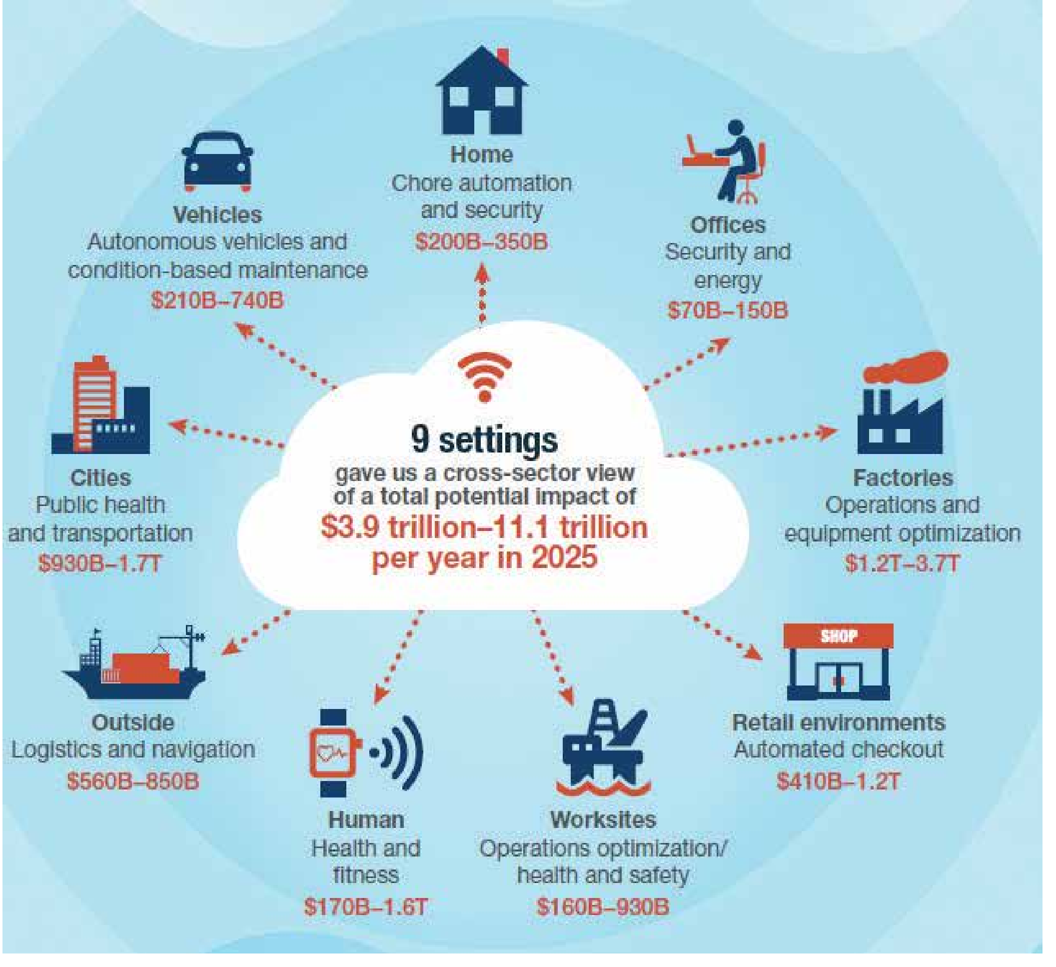

Figure 3. The Internet of Things: the value beyond the hype, McKinsey Global Institute (2015).

By Hewlett Packard (2016), as progress has been made in the high-tech sector, public institutions to provide city smarter services are already linking building security systems (57%), street lighting (32%), and automobiles (20%) to create an interoperable technological environment that will support the smart city of the future. For instance, the most widely deployed IoT applications in this sector are remote monitoring and the control of urban devices (27% responded that this is the main application) [2]. The major issues with the IoT are build costs (50%), maintenance (44%), and integration with legacy systems (43%). Within the city, the limitations of legacy technology are becoming the biggest challenge. Nearly half (49%) of IT departments in the public sector are having difficulty integrating existing technologies into the system. Smart cities that are implementing effective the IoT strategies, however, actually demonstrate why the IoT initiative is worthwhile. Seven out of 10 public-sector IoT adopters (71%) said they are saving money, and 70% said that the IoT has improved visibility across the smart city. This is an essential step toward realizing the smart city's integrated infrastructure [2].

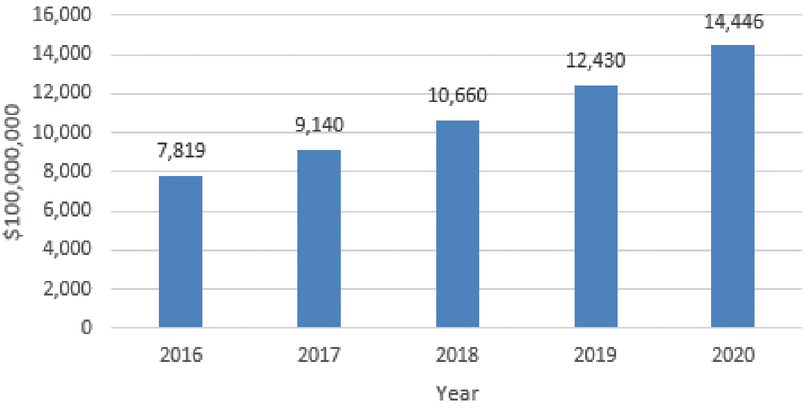

Figure 4. The forecast of smart city market size (source: Research and Markets, 2017).

The global market for the smart city is growing rapidly. As in **Figure 4**, by Research and Market, the market size of smart city was $781.9 billion in 2016. It forecasted that it will be $1.4 trillion in 2020 with an annual growth rate of 16.6% [3]. By Frost and Sullivan [4], the estimated smart city market is likely to be worth a cumulative $1.565 trillion by 2020. Grand View Research [5], a US-based market research and consulting company, estimates this will be $1.423 trillion by 2020. MarketandMarkets [6], a B2B research company, estimates this will be $1202 trillion by 2022.

Regarding the view of high-tech, according to Gartner [7], as the market size of the smart city is growing, the utilization of the IoT is also increasing significantly. Gartner predicted that 1.6 billion IoT devices will be used in a smart city in 2016, up to 39% from 2015. Smart commercial buildings were expected to be the top usage application of IoT devices in 2017, while smart homes were expected to rise to first place (more than 1 billion IoT devices) in 2018.

According to the United Nations (UN) [8], in 2016, 1.7 billion people (23% of the world's population) lived in a city with at least 1 million inhabitants. By 2030, a projected 27% of people worldwide (2.3 billion people) will be concentrated in cities with at least 1 million inhabitants.

3. Broadband networks

In technical view point, smart connected city is based on broadband networking. In advance, the real-time networks are used to transfer and communicate the generated and created data. The data are usually collected from the IoT devices equipped with sensors. Now, new 5G cellular standard is introduced to enable the IoT. In addition, the cutting edge technology like the IoT in **Figure 5** is tremendously emerging and they are applied to develop embedded intelligent systems for providing efficient services as optimal solutions to solve urban problems in city life with processing and transmission of big data through the broadband networking.

Figure 5. Growth in the Internet of Things (the number of connected devices will exceed 50 billion by 2020), source: WEF (assets.weforum.org).

By experts, by 2022, more than 1 trillion sensors will be embedded in networked devices and there are the embedded 45 trillion sensors in the world within 20 years [9]. By 2020, the estimated IoT market size, the number of connected devices will be around 50 billion by Cisco, 32 billion by IDC, 30 billion by McKinsey, and 26 billion by Gartner [10].

In the smart city, physical things embedding computing systems can be hyperconnected by IoT technology. The IoT was developed for the purposes of connecting various things to exchange information and realizing value-added information services. If the IoT is effectively applied to cities' facilities, management and surveillance for city functions could be performed faster and more efficient than before. In addition, it is possible to provide a variety of services that are required by citizens for their smart living and lives.

According to the development of the IoT, by 2020, the estimated economic effect is $19 trillion by Cisco, $7.1 trillion by IDC, and 1.9 trillion by Gartner. By 2025, the economic impact depending on the IoT-based created business is estimated between $2.7 and $6.2 trillion [10]. Machina Research forecasted the extreme growth of the IoT-based business market, especially in platform and services as given in **Figure 6** [11]. By Machina Research (2015) [12], the global IoT market opportunity will reach $4.3 trillion by 2024.

In **Figure 7**, by ETRI, the world market size of 5G mobile communications is estimated from $52 billion in 2020 to $1.2 trillion in 2026 at a CAGR of 148%. In 2026, the market size of "communication services" is forecasted for $731.9 billion, accounting for 60.6% of the total, while

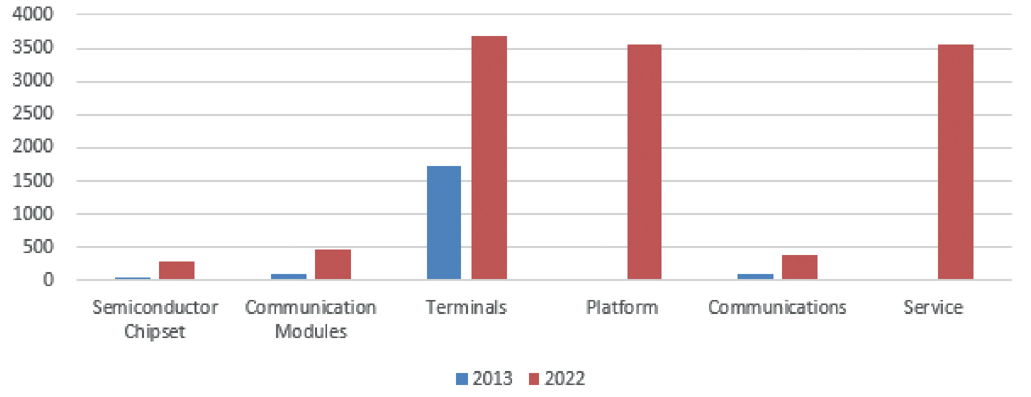

Figure 6. The estimated market size by the Internet of Things in the world (millions), source: Machina Research (2013).

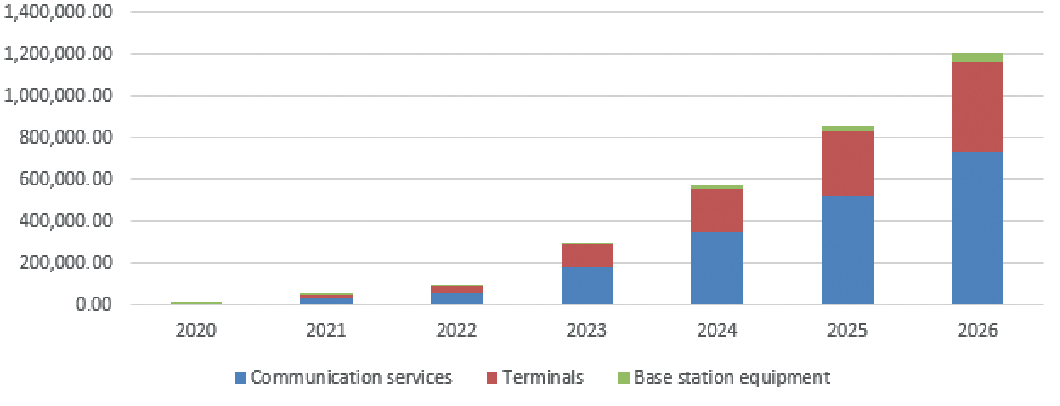

Figure 7. 5G mobile market size in the world (millions).

“terminals” and “base station equipment” are estimated by $438 billion at 36.3% and $38.4 billion at 3.1%, respectively [13]. It is assumed that the market transition will be started from 3G and 4G to 5G networks from 2020. The terminals are including the market size of sales of feature and smart phones except for the wearable devices [13].

According to ETRI, the world 5G mobile communication market is predicted to grow at a CAGR of 148% from US $5.2 billion in 2020 to US $1.2 trillion in 2026 [13].

4. Smart city concepts

According to research by the International Telecommunication Union (ITU) in 2014 [14], there are 116 definitions of the smart city. Keywords used in definitions vary from “environment” and “sustainable growth” to “information and communication technology” (ICT) and “intelligence.” According to this research, 26% of 116 definitions of the smart city are related to keywords such as ICT, communication, intelligence, and information. This means that ICT is a core concept of the smart city. The United Nations Conference on Trade and Development (UNCTAD) [15] defines the smart city as smart mobility, smart economy, smart living, smart governance, smart people, and smart environment.

In **Figure 8**, there are three components like physical services, holistic services, and broadband networks. Recently, the definition of the smart city has been extended from the physical aspects, such as city infrastructure and technologies, to the social aspects to serve holistic services including social services and impacts, as results of the interoperated processing of smart financing and capital, government, health care, welfare, and environment.

In the initial stage of the digital city, the implementation of the cities was based on fixed broadband networking technologies to connect among physical devices with computing and connectivity functions. In the city, the implemented digital services were online commerce, information service providers, electronic governments, and so on.

With the entrance of mobile computing devices, the developed ubiquitous city was spotlighted. It focused on the building of dynamic connections among mobile devices as well as physical things as components of infrastructure of the city. Ubiquitous city began to serve city public and private services using mobile devices in mobile broadband networking environment. For instance, pedestrians with mobile devices can easily access electronic services such as e-government and e-markets whenever and wherever they want even if they are walking down the street. In the ubiquitous city, it was focused on the transferring and deliberating data among

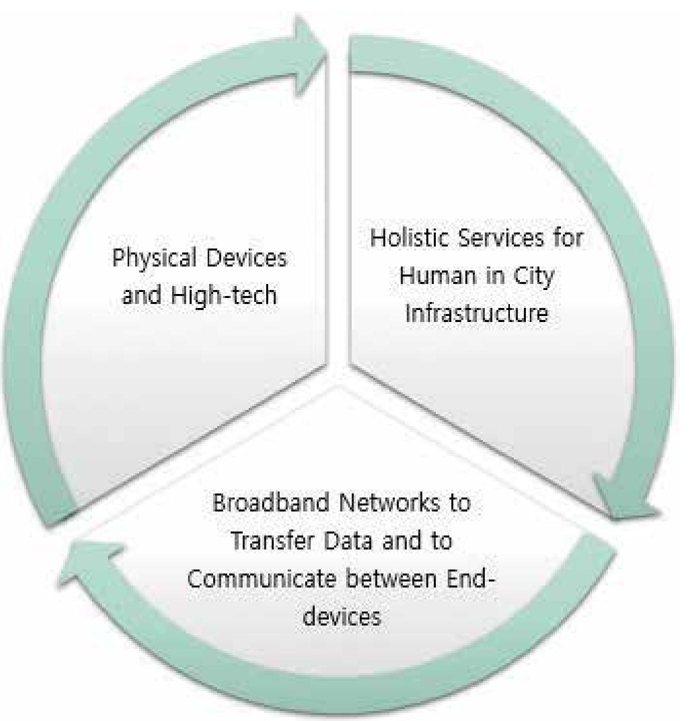

Figure 8. The components of smart city.

http://dx.doi.org/10.5772/intechopen.79988

discrete objects through broadband dynamic networking. It started to build dynamic connections among the physical computing devices such as products, devices, sensors, vehicles, and drones as components of city infrastructure. However, the developed services were centered on city service providers like e-government services and construction of city infrastructure.

In smart connected city, it is important to provide citizen-centric services which are based on smarter connected networking according to the appearance of the fourth industrial revolution. The developments of these kinds of services depend on the embedded intelligent technologies and created customer centric services based on dynamic connecting network systems. The smart connected city is constructed by integration of the intelligent techs, dynamic and interactive connections, and city infrastructure. It orchestrates the physical techs, social services, and broadband services to provide city holistic services. It connects and internetworks among the physical devices such as products, drones, devices, sensors, and vehicles that are components of city infrastructure to serve smarter administration, smarter energy, smarter home, smarter office, smarter security, smarter vehicles, smart factories, and so on as in **Figure 9**. So smart connected city becomes a kind of a platform to create and provide the holistic city services to citizens, and it is possible to provide a variety of services that are required by citizens for their smart living and lives. As a platform, in smart connected city, services for social aspects are implemented depending on citizens' desires to solve a variety of urban problems, such as housing, health care, capital, energy, transportation, and public services. Finally, in the smart city, the concept of space is expanded from physical space to connect physical things to social space to provide citizen centric services, social impacts and effects, and so on. The construction of the smart city is based on the holistic orchestration of physical and social states for social efficiency.

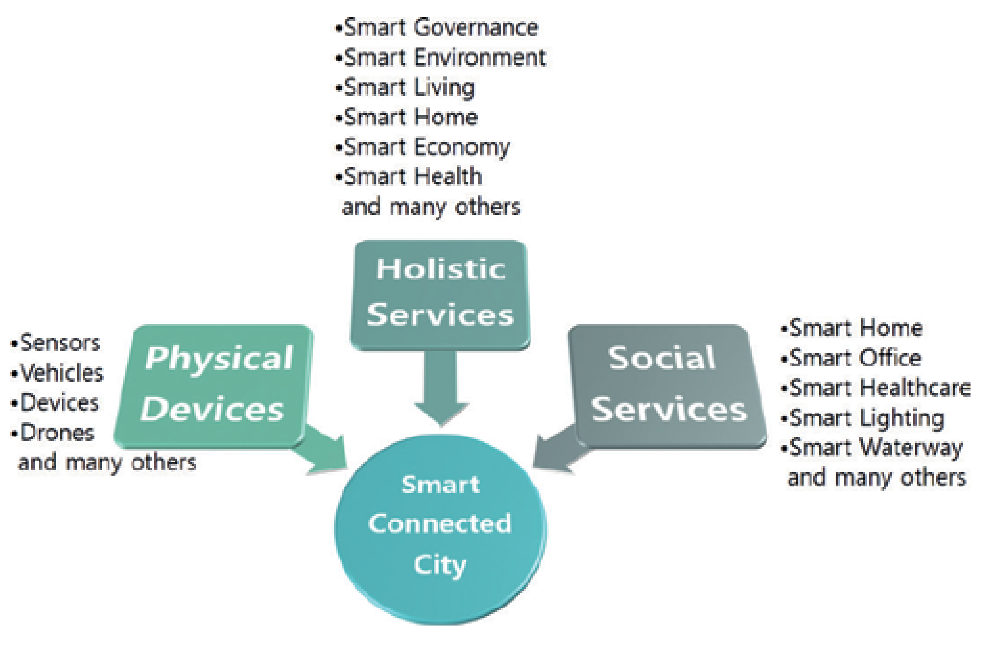

Figure 9. The holistic services in smart connected city as a platform.

5. Smart city types

The world is faced with several problems due to the increase in energy consumption and the urban population. The hyperconnected city could solve these problems. It is clear that the future city, with autonomous driving cars, the use of renewable energy, and the spread of the shared economy, will not be the same as the present metropolitan areas. The future smart city will be a new city that pursues sustainable development based on advanced citizen consciousness and overcomes the problems of industrialization based on digital technology. Smart city types can be clustered by several types of purposes, such as urban problem solving, environmental and ecological, and the ICT-based smart city.

5.1. Urban problem-solving smart city

The smart city is designed to systematically solve urban problems for the sustainability and resilience of the city. These problems are increasing according to urban population growth. Urban problems include a lack of infrastructure, traffic congestion, the carbon problem in transportation, parking problems, and air pollution, which are related to roads, waterworks, electricity, schools, etc., as parts of the public sector in the city. Thus, urban problem-solving is closely coupled to solving problems in the public sector for citizens in the city.

5.2. Environmental smart city

This is specifically focused on the environment in urban problems because it is directly related to improving the quality of citizens' lives and the sustainability of the city. From the environmental perspective, the environmental smart city can concentrate on a green-smart city, smart environment, climate-smart city, zero-energy city, smart grid city, carbon-free smart transportation, carbon capital creation and management, and so on.

5.3. Ecological smart city

If a hyperconnected society that connects things and cities is realized in the near future, we will be able to truly experience a smart city that can integrate the city management system that has been operated individually. For the improvement of the efficiency of the city, certain actions should be taken, such as avoiding wasting time and the reduction of energy and materials. In the ecological smart city, considerations include the enhancement of environmental quality, fostering system effectiveness, and the optimization of resource yields for resource circularity in the city. To construct a smart city, it is necessary to consider several sectors, such as the connected, data, public, social, business, and environment sectors.

5.4. ICT-based smart city

In the smart city, physical things are hyperconnected by smart and intelligent technologies, such as the IoT using sensors, big data analytics, cloud computing, and AI. The IoT generates a huge amount of data called big data. It is necessary for a city to create, manage, and control services using big data for citizens, because the data include a significant amount of information on the city that are generated yearly, monthly, and daily. However, it was difficult to collect and manage such data before the appearance of high-tech. In the ICT-based smart city, by using the collected and generated data and information, it can be focused on types of cities, such as data-, knowledge-, and network-driven cities. The accumulated, classified, and processed data can be stored in knowledge management systems, which can be applied to create and improve citizens' services in the city.

5.5. Technology-driven smart city

In the technology-driven smart city, three major facts are considered. First, there is hardware, such as computing and mobile devices, sensors, equipment, controllers, and Wi-Fi. Second, there are data and software applications usually used to develop services for the citizens or for controlling and managing the resources of the city. There are big data analysis, cloud computing, the digital industry and capital, integrated operational and control centers, and so on. Finally, they are connected by a network that plays the role of connecting all things (e.g., machine to machine, machine to people, people to people, physical things to social value, and the public and private sectors) [16]. This is for the development of the sustainable city and to provide smart services for the citizens. Therefore, we can consider hardware-driven, data- and knowledge-driven, software-driven, and network-driven smart cities.

6. Smart infra with data, services, and social sectors

In the smart city, there is the coexistence and organization of infrastructure, data, and services based on digital and smart tech. The Korea Transport Institute (KOTI) defines the smart city as consisting of infrastructure, data, and services. Infrastructure refers to physical and technical implications, including cities and ICT technologies. Data originates from all infrastructure and things in the city based on IoT technology. Services are developed for utilization by citizens based on the collected data.

According to the urban development stages, the digital city was based on data and network-related technologies and focused on the development of software applications. The ubiquitous city was based on mobile-based technologies and concentrated on the construction of the city infrastructure. As an advanced city model, the smart city is introduced, which is constructed from data, software applications, networking technologies, and citizens' services. It has been available for all of citizens using the developed digital technologies, hyperconnected network, and high-tech, including big data as citizens' data, cloud as the network, and AI for customized services, in the smart city. In addition, the smart city is a kind of holistic system with infrastructure, data, and services to efficiently integrate the city resources. It is constructed from smart infra, which is organized and controlled by smart operating, smart management, a smart trade system, and so on based on the collaboration and coordination of resources.

In the smart city, the data sector is related to collecting, storing, and analyzing data. From the big data collected by sensors, the smart city can construct the smart cloud as a network using the IoT. In the digital sector for the collection of data, there are sensors for the collection of data, data storage, data platforms, and network devices. Data truly makes the smart city smart, and data visualization is needed to facilitate data access.

The service sector is related to solving citizens' problems and providing more convenience in citizens' lives. Factors related to the sustainability of the city include low-carbon systems, the reduction of pollution systems, and climate systems, and those related to improving citizens' convenience include street lighting, traffic congestion systems, welfare services, and so on. In addition, it is necessary to intertwine and integrate the individually developed services to provide services to citizens. The services are also specifically developed for the improvement and support of the citizens' community in the municipal government, advanced public sector provision, planning, permits, and operational entities. It is necessary to demonstrate the developed services and applications in the city. The city as a kind of platform contains good infra. Therefore, it is sometimes called smart infra for the integration and interoperation of urban sectors to build the smart city, such as the physical, public, private, business, economy, capital, social, environmental, and ecological sectors.

Ultimately, the smart infra pursues the creation and support of social sectors in urban areas with social goods, effects, and value in the fields of national health care, welfare, climate, low-carbon resources, economy, social community, funding for innovation, and so on. Social facts are closely related to the public sector, which is developed to solve urban problems. For the construction of the smart city, it is important to consider social facts to provide the society with social goods, effects, and value using innovative digital technologies. The use of digital and high technologies to generate positive social impacts as "social tech" has emerged in a number of fields, from the provision of health care to addressing financial exclusion. In the smart city, the purpose is the creation of social facts for citizens who need customized services. To achieve this, it is necessary to analyze big data and to customize services that can be implemented using AI.

Even if the social sector is related to the public sector to improve public services in the city, through the progress of the city and the development of innovative technologies, citizens' service needs have become very specified and customized. Therefore, public service support processes have also been individualized for the improvement of citizens' satisfaction and can be customized depending on an individual citizen's needs. To provide customized city services, it is necessary to dynamically interoperate social things under the collaboration and coordination of smart services. In addition, there are some significant sectors, such as the economy, capital, business, environmental, and sustainable sectors, for the ecological smart city. The business sector is related to financing and loan systems, the market and trade system, and so on. The environmental sector supports climate systems and the energy gird and so on. The sustainable sector is related to a low-carbon and green city. As an instance of the economy and capital sector, this is related to carbon capping and trading for dual benefits, such as the improvement of sustainability and enjoying the prosperity of the city.

7. The smart platform as a holistic ecosystem

In the smart city, there are many components of the physical and social sectors to be interconnected and integrated with interactions. To achieve this, there are two significant aspects. The first is a test bed to implement the developed physical and high technologies. The other is for the completion of a holistic and ecological city using digital and social tech. To achieve this, it is necessary to have a test bed. It is called a variety of names, such as a Digital Living Laboratory [17], platform, big data hub, and ecosystem. This means that it is indispensable to pass through the test bed for the completion of the construction of the smart city as a kind of platform. In the platform, the components can be assembled and integrated as interoperable blocks to create social value for collaborative and coordinated services for citizens in the smart city.

It is important to serve and share optimized services with citizens. This is based on the optimization of the resources of the city and improvement of citizens' satisfaction. For instance, the energy grid system optimizes energy resources to improve citizens' satisfaction by trading energy among citizens. In smart trash can systems, cleaners collect the cans when the cans are full. This is also an instance of the optimized practical use of city resources. Using the collected data, it is possible to make a correct decision using the optimization systems. For instance, we can make a decision about how many trash cans are needed and collected in the city and how much idle energy exists and can be traded in the city. It is possible to rationally allocate the resources of the city as a kind of holistic system. Through this process, we can construct necessary optimized services to improve citizens' satisfaction for a sustainable, efficient, effective city.

To complete the holistic smart city, it is necessary to evolve thinking. When technologies are developed, it is necessary to employ system thinking to understand the new developed technologies and find application for the technologies. With ever-changing technological advances, thinking processes need to be evolved from system thinking based on technologies to digital thinking based on the design of services and the developed digital technologies. Digital thinking can positively determine applications using the developed digital technologies. This means that digital thinking can create new services. Finally, the created services using the developed technologies need to be integrated, collaborated, and coordinated to seamlessly provide services. This is called holistic thinking to optimize resources and reduce friction for citizens.

The smart city is a kind of ecosystem because it is not simply completed by only digital technology, smart components, and climate and environmental systems. It needs organic interaction among them and is based on the intertwined components. There are many connections, such as machine to machine, people to people, and people to machine. Thus, it is important to seamlessly operate the components and connections without disharmony.

The smart city is a kind of a platform to trade and share digital and smart technologies and a kind of ecosystem to provide holistic services.

8. Smart city global trends

Therefore, many countries around the world are promoting the smart city for various purposes. In this section, we will briefly look at the smart city trends in the US, EU, UK, Spain, Japan, China, and Korea.

8.1. The United States

In 2013, the US government implemented a project named "Smart America," which was boosted by the White House and confirmed that IoT technology could penetrate into society. Based on this, the White House has invested $1.6 billion into 25 kinds of new technology projects with the "smart city initiative" [18]. The White House is preparing for the future by organizing the skills needed for future cities. Local governments are also making efforts to build smart cities to help ease traffic jams and the fight against crime and for the stimulation of economic growth, response to climate change, and city administrative services.

In 2016, the US Department of Transportation (DOT) launched the smart city challenge. In March, seven finalists were announced, and Columbus, the state capital of Ohio, was chosen as the final winner in June. The smart city challenge was a pilot project to employ advanced technologies in mid-sized cities to create innovative solutions for future city transportation. Columbus received $40 million of funding from DOT along with $10 million of funding from Vulcan Inc. for electric car projects.

8.2. The European Union

In response to changes in economics and technology triggered by globalization, European cities have faced challenges based on the need to enhance competitiveness and transform into sustainable cities. In particular, small- and medium-sized cities, unlike big cities, are gradually approaching a crisis with less competitiveness [4]. Therefore, the EU has established the smart city strategy focusing on small- and medium-sized cities with more than 100,000 people [19]. The characteristics of a smart city in the EU include a smart economy, smart people, smart governance, smart mobility, smart environment, and smart living [20].

The elements that need to be wiped out in the smart city are congestion, air pollution, crime, and high-cost energy. On the other hand, the elements that need to be boosted in the smart city are energy efficiency, a clean environment, and convenient transportation.

8.3. The United Kingdom

The UK is one of leading countries for building smart and digital cities in the world. They are operating the "HyperCatCity project" in Milton Keynes, Manchester, and London. Milton Keynes as a municipal government is focusing on economic growth and building transformative, open, and accessible services with connectivity using sharable data [21].

Manchester is a large-scale smart city demonstrator that shows how IoT technologies and services can improve the quality and efficiency of services in transport, energy, health, and culture. Manchester is focusing on supporting an open platform and opportunities in four key areas. Using technology can enrich the local community activities for residents. Collaboration is essential for open innovation [22]. After adopting the smart traffic system application, there were reductions of 25% of traffic time, 50% of traffic accidents, and 10% of air pollution [23]. Additionally, London, a tech city, has funded the development of innovative digital technologies. For instance, smart city "hackathons" have been run many times and increased the overall level of participation between citizens and business to address city-specific challenges.

http://dx.doi.org/10.5772/intechopen.79988

8.4. Spain

Barcelona's smart city was planned from 2011 to 2015 and was awarded the European Capital of Innovation 2014–2016 title. The model is focused on the alignment of resources and development of a sustainable city for citizens' welfare. For instance, about 30% energy savings are made per year through free Wi-Fi routed via street lighting systems [23] and smart parking spaces have been adopted [24].

As a smart city, Valencia is focused on smart, sustainable, and inclusive growth. Smart growth is based on the knowledge and innovation economy. Sustainable growth is focused on the effective use of resources. Inclusive growth is related to a high level of employment, delivering social and territorial cohesion. In Valencia, the smart city has been implemented based on e-government, smart economy, quality of life, sustainability, and mobility and infrastructure. There is a Valencia smart city platform for smart service management based on intelligent management systems [25].

8.5. Japan

Japan is promoting a "Strategy for becoming an environment and energy power through green innovation" as part of the "The new growth strategy (basic policies) toward a radiant Japan" since 2010 [26]. The Ministry of Economy, Trade and Industry (METI) and New Energy and Industrial Technology Development Organization (NEDO) have been jointly promoting a "Smart Community" strategy since 2010. According to NEDO [27], the Smart Community is "a mechanism to use energy intelligently by sharing data in both directions between the supply and demand sides of the system using ICT. This enables the optimal use of renewable energy such as solar power, wind power, and biomass while limiting the impact on the environment and increasing energy efficiency."

8.6. China

The smart city has become a new urban development trend in China since IBM introduced the concept of a "Smarter planet" in 2009. The Central People's Government of China has been directly managing and promoting China's smart city policy since 2013, which was promoted at the local government level. According to research by Li et al. [28], there were 193 approved smart city pilot projects in China by 2013. The Chinese government has regarded the smart city as a key strategy to promote industrialization, informatization, and urbanization. Therefore, the rapid development of the smart city in China is largely attributed to the cooperation between IT companies and the government.

According to news from SmartCityWorld based on an Arup study [29], China has been recognized as a significant driver of global smart city growth as some RMB500 billion ($75 billion) has been earmarked for the smart city in the country during the 13th Five-Year Plan period (2016–2020).

8.7. Korea

Korea recently launched its Presidential Committee on the fourth industrial revolution in September 2017, and this committee set up a special subcommittee on the smart city that aims to promote the smart city as the innovation engine of the fourth industrial revolution at the national strategic level. The Korean government is accelerating the creation of the smart city, which is the product of the ICT convergence industry [30].

The development of Goyang smart city is based on an open innovation platform to solve pending urban problems using the developed converged and intertwined services employing the IoT. In addition, Goyang smart city is building a citizen-based local community in the municipal government as a demonstration of IoT-based services. Goyang applied high-tech to a smart park service for a street lighting system, a smart welfare service for a children's day-care center, a smart climate-monitoring center, and a smart ecological service by monitoring biomass [31]. In June 2017, Goyang city built a smart garbage collection system based on the IoT along the city streets and in the residential areas. Specifically, the system consists of a load detection sensor, a solar compression garbage can, a garbage collection tracker, and a garbage collection system. The data measured in the smart trash can are transmitted to the Goyang City Demonstration Center server and the environment-friendly smartphone so that the garbage-loading information can be checked in real time. The smart parking enforcement system can perform monitoring and decision-making regarding if something is legal or illegal using the IoT sensors that are attached to garbage collection can in Goyang [32].

Busan smart city is focused on the collection of data as big data that are interconnected with networking. Busan smart city has been developed as a platform for machine-to-machine and city-to-city connections. Busan is focused on the development of urban services, such as safety, transportation, energy, converged services, and local community services, and the demonstration of big data-based services [32].

9. Conclusion

Many cities and countries are interested in building smart cities and have started to construct them. The smart city is an ecological system including many components, such as high-tech and digital sector-based and social and community-based components, and city sustainability related to environments and ecosystems, economy and business, and government. The smart city is a kind of platform to connect physical, social, governmental, public, business, and environmental systems as components of the city. The development of the smart city is closely coupled with newly developed technologies, such as the IoT, big data, cloud, and AI. In the smart city, there are interoperable services and social value for the citizens. Manwaring [33], cofounder of the IoT Living Lab in Amsterdam, mentioned that "We need to empower communities to solve their own problems, provide them with tools to accelerate social impact and make their lives better. Smart People make Smart Cities and technology is useless if it doesn't engage the public." The smart city built with IoT should ultimately serve as a people-centered and equitable space that increases people's quality of life [34].

http://dx.doi.org/10.5772/intechopen.79988

Author details

Hyun Jung Lee[1]* and Myungho Kim[2]

*Address all correspondence to: hjlee5249@gmail.com

1 Department of Economic and Social Research, Goyang Research Institute, Goyang-si, Gyeonggi-do, South Korea

2 Omogary Holdings, Goyang-si, Gyeonggi-do, South Korea

References

[1] McKinsey. The Internet of Things: Mapping the Value beyond the Hype. 2015. Available from: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/The%20Internet%20of%20Things%20The%20value%20of%20digitizing%20the%20physical%20world/The-Internet-of-things-Mapping-the-value-beyond-the-hype.ashx [Accessed: April 24, 2018]

[2] Hewlett Packard Enterprise. Internet of Things, Today and Tomorrow. 2016. Available from: http://www.arubanetworks.com/assets/eo/HPE_Aruba_IoT_Research_Report.pdf [Accessed: April 24, 2018]

[3] Research and Markets. 2017. Available from: https://www.researchandmarkets.com/ [Accessed: April 24, 2018]

[4] Economic Review. 2016. Available from: https://kerala.gov.in/documents/10180/ad430667-ade5-4c62-8cb8-a89d27d396f1 [Accessed: April 24, 2018]

[5] Grand View Research. Smart Cities Market Expected To Reach USD 1,422.57 Billion by 2020. 2016. Available from: https://www.grandviewresearch.com/press-release/global-smart-cities-market [Accessed: April 24, 2018]

[6] MarketandMarkets. Smart Cities Market Worth 1,201.69 Billion USD by 2022. 2017. Available from: https://www.marketsandmarkets.com/PressReleases/smart-cities.asp [Accessed: April 24, 2018]

[7] Gartner. Gartner Says Smart Cities Will Use 1.6 Billion Connected Things in 2016. 2015. Available from: https://www.gartner.com/newsroom/id/3175418 [Accessed: April 24, 2018]

[8] United Nations. The World's Cities in 2016, Data Booklet. 2017. Available from: http://www.un.org/en/development/desa/population/publications/pdf/urbanization/the_worlds_cities_in_2016_data_booklet.pdf [Accessed: April 24, 2018]

[9] Wellers D, World Economic Forum. Is This the Future of the Internet of Things? 2015. Available from: https://www.weforum.org/agenda/2015/11/is-this-future-of-the-internet-of-things/ [Accessed: April 24, 2018]

[10] Koontz R. The Internet of Things "Smart" Products Demand a Smart Strategy Using M&A for a Competitive edge, 2015 Woodside Capital Partners, March 2015. 2015. Available from: https://es.scribd.com/document/349983910/WCP-IOT-M-and-A-REPORT-2015-3-pdf [Accessed: April 24, 2018]

[11] Machina Research. 2013. Available from: https://machinaresearch.com/ [Accessed: April 24, 2018]

[12] Machina Research. 2015. Available from: https://machinaresearch.com/ [Accessed: April 24, 2018]

[13] Information Research Service Global. Market Report 2017-04. 2017. Available from: http://www.irsglobal.com [Accessed: April 24, 2018]

[14] ITU-T Focus Group on Smart Sustainable Cities. Smart Sustainable Cities: An Analysis of Definitions. 2014. Available from: www.itu.int/en/ITU-T/focusgroups/ssc/Documents/Approved_Deliverables/TR-Definitions.docx [Accessed: April 24, 2018]

[15] UNCTAD. Issues Paper on Smart Cities and Infrastructure. 2016. Available from: http://unctad.org/meetings/en/SessionalDocuments/CSTD_2015_Issuespaper_Theme1_SmartCitiesandInfra_en.pdf [Accessed: April 24, 2018]

[16] Ko J, Hong SG, Lee BB, Kim NS. Trends of Converging Smart Devices with IoT Technology, 2013 Electronics and Telecommunications Trends. 2013. Available from: https://ettrends.etri.re.kr/ettrends/142/0905001853/28-4_079-085.pdf [Accessed: April 24, 2018]

[17] Living Lab. Sustainability. Available from: https://green.harvard.edu/series/living-lab [Accessed: April 24, 2018]

[18] The White House. FACT SHEET: Administration Announces New "Smart Cities" Initiative to Help Communities Tackle Local Challenges and Improve City Services. 2015. Available from: https://obamawhitehouse.archives.gov/the-press-office/2015/09/14/fact-sheet-administration-announces-new-smart-cities-initiative-help [Accessed: April 24, 2018]

[19] European Commission. The European Innovation Partnership on Smart Cities and Communities. 2015. Available from: http://ec.europa.eu/eip/smartcities/ [Accessed: April 24, 2018]

[20] Giffinger R, Fertner C, Kramar H, Kalasek R, Pichler-Milanović N, Meijers E. Smart Cities: Ranking of European Medium-Sized Cities. Vienna, Austria: Centre of Regional Science (SRF), Vienna University of Technology. 2007. Available from: http://www.smartcities.eu/download/smart_cities_final_report.pdf [Accessed: April 24, 2018]

[21] Hypercatcity in Partnership with Bristol Is Open & Milton Keynes Collaborative Smart City Innovation, Supported by Mayor of London. Available from: http://www.hypercat.io/uploads/1/2/4/4/12443814/hypercatcity.pdf [Accessed: April 24, 2018]

[22] What is CityVerve? Introducing Manchester's Smart City Demonstrator. Available from: https://cityverve.org.uk/what-is-cityverve/ [Accessed: April 24, 2018]

[23] Orchestrating Infrastructure, for Sustainable Smart Cities. IEC WP Smart Cities: 2017, International Electrotechnical Commission. 2017. Available from: http://www.iec.ch [Accessed: April 24, 2018]

[24] Barcelona: Smart City Revolution in Progress—Financial Time. Available from: https://www.ft.com/content/6d2fe2a8-722c-11e7-93ff-99f383b09ff9 [Accessed: April 24, 2018]

[25] Valencia Smart City Strategy, AJUNTAMENT DE VALENCIA. Available from: https://www.smart-circle.org/smartcity/wp-content/uploads/sites/3/2014/05/Valencia-Estrategia-Smart-City-2014-ENGLISHv4-mayor-interview.pdf [Accessed: April 24, 2018]

[26] Japan Cabinet Office. On the New Growth Strategy (Basic Policies). 2009. Available from: http://www.fj.emb-japan.go.jp/Political/Full%20Text_1230sinseichousenryaku_e[1].pdf [Accessed: April 24, 2018]

[27] Japan New Energy and Industrial Technology Development Organization. Smart Communities Developing Towns of the Future that Coexist with the Environment. 2014. Available from: http://www.nedo.go.jp/content/100642110.pdf [Accessed: April 24, 2018]

[28] Yongling Li, Yanliu Lin, Stan Geertman. The Development of Smart Cities in China. 2015. Available from: http://web.mit.edu/cron/project/CUPUM2015/proceedings/Content/pss/291_li_h.pdf [Accessed: April 24, 2018]

[29] SmartCityWorld. Smart Cities Development Hub Launched in China. 2016. Available from: https://smartcitiesworld.net/news/news/smart-cities-development-hub-launched-in-china-1152 [Accessed: April 24, 2018]

[30] Yung Sil Yoon. Creation of Smart City: 4th Industrial Revolution Committee to Promote Smart City as Export Models, Business Korea. 2017. Available from: http://businesskorea.co.kr/english/news/national/19747-creation-smart-city-4th-industrial-revolution-committee-promote-smart-city [Accessed: April 24, 2018]

[31] Kim JH. Goyang City, IoT Garbage Can Appear, May 17, 2017, e4ds. Available from: http://www.e4ds.com/sub_view.asp?ch=30&t=1&idx=6203 [Accessed: April 24, 2018]

[32] IoT-Based Smart City Progress Status, NIPA. Available from: http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjI-qPPs9faAhVHU7wKHcq6BRQQFggmMAA&url=http%3A%2F%2Fwww.krnet.or.kr%2Fboard%2Finclude%2Fdownload.php%3Fno%3D2249%26db%3Ddprogram%26fileno%3D2&usg=AOvVaw08AQpE4qwzKv9VXpsO-02d [Accessed: April 24, 2018]

[33] Paul Manwaring. The IoT Living Lab at the Smart City Innovation Summit Asia—SCIS Asia 2017. 2017. Available from: https://www.linkedin.com/pulse/iot-living-lab-smart-city-innovation-summit-asia-scis-manwaring [Accessed: April 24, 2018]

[34] Ashton K. Making Sense of IoT. 1999. Available from: http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjz3vj-uIXYAhVIgrwKHYGzC6cQFggrMAA&url=http%3A%2F%2Fwww.arubanetworks.com%2Fassets%2Feo%2FHPE_Aruba_IoT_eBook.pdf&usg=AOvVaw2eucTstKelqZE5OGGRCBRC [Accessed: April 24, 2018]

Chapter 17

Planning of FiWi Networks to Support Communications Infrastructure of SG and SC

Arturo G. Peralta

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71781

Abstract

Nowadays, growth in demand for bandwidth, due to new and future applications being implemented, for services provided from smart grids (SG), smart cities (SC) and internet of things (IoT), it has drawn attention of scientific community, on issues related to planning, and optimization of communication infrastructure resources, in addition is necessary comply with requirements such as scalability, coverage, security, flexibility, availability, delay and security. Another important point is how to find and analyze possible solutions that seek to minimize the costs involved by capital expenditure (CAPEX) and operational expenditure (OPEX), but where it is possible to measure the uncertainty coming from stochastic projections, in order to obtain the maximum benefit expected to give access to users Who benefits from the services provided by SG, SC and IoT, on the other hand, we must look for communications architectures that generate optimum topologies to meet demanded requirements and at the same time save energy, possible alternatives highlight the use of hybrid networks of optical fiber links combined with wireless links (Fiber-Wireless, FiWi). This chapter seeks to provide planning alternatives to network segments linking universal data aggregation point (UDAP) with base stations (BS), this segment joins wide area network (WAN) with metropolitan area network (MAN).

Keywords: FiWi networks, internet of things, planning, scalability, smart cities, smart grids, stochastic programming

1. Introduction

The following chapter proposes a ***new planning model*** for the ***scalability and deployment*** of communications infrastructure that give supports to SG, SC and IoT; countries such as the United States and those that made up the European Union, are carrying out projects with SG motivated by the drawbacks related to the current energy network, such as blackouts, overloads and voltage drops, most of these events were due to a slowness in response times of the devices that control the energy network, in addition, the increase in the population of residential and commercial clients that demand to connect intelligent appliances or the IOT, has caused that the network of supply is obsolete, considering this background, it is urgent to make changes in the infrastructure of electrical and communications systems, so as to adapt to the temporal-spatial evolution of customers and to meet requirements such as: ***scalability, coverage, security, flexibility, availability, delays and latencies*** [1–3].

In order to observe a horizon of temporal-spatial evolution, it is necessary to characterize important parameters such as ***the demand and density*** of users, Who benefit from the services offered by SG, SC and IoT. It is difficult to make accurate forecasts regarding the projection and growth of intelligent electronic devices (IED) given that uncertainty exists because of the number of variables involved, however it is possible to make future projections in a stochastic way, which can serve as a reference for the take of decisions related to the deployment of the communications network, which supports the services provided on SG, SC and IoT, but testing various planning scenarios.

Another point to highlight is how to find and analyze possible solutions that seek to minimize the costs involved by CAPEX and OPEX to maximize the benefits expected by telecommunications operators. Therefore, communication architectures that generate optimal topologies should be sought, in order to meet the requirements demanded by SG, SC and IoT and that at the same time save energy; possible alternatives from the scientific community point to the use of FiWi Hybrid Networks [4–9].

The systems implemented through SG and SC are characterized by important parameters such as user density, types of services provided, spatial and geographical location of resources like communications infrastructure [1, 10–15], which is the backbone of SG, SC and IoT. On which applications and services such as automated meters reader (AMR), or with more extended services advanced metering infrastructure (AMI), which for example help in detecting system failures such as: communications, failures in devices like sensors, actuators and/or controllers or failures due to control system and resources scheduling [16].

As for electricity distribution in terms of a smart grid, the terminology of distributed generation (DG) or distributed energy resources (DER) is introduced. In this way, the der goes from having few generation centers to having a large number distributed generation centers throughout electrical network, which can be renewable and/or traditional, forming interconnected micro-networks [17]. The main advantage of having DER is that distribution network operators (DNOs) can quickly and efficiently reconfigure and redirect power flow in response to events such as failures, changes in demand or even changes in energy generation costs.

Furthermore, storage sources include traditional high-performance batteries such as lead-acid, sodium sulfide, lithium ions, and others, but studies are being made of materials and alloys that will form batteries of greater capacity, durability and more economical that the current ones. In ([18], chapter 1) are mentioned membranes and cells that are in process of investigation like polymer electrolyte membrane (PEM) and hydrogen fuel cells (HFC).

http://dx.doi.org/10.5772/intechopen.71781

On the other hand, in the next years a considerable increase in the penetration of electric vehicles (EV) is expected and the most common will be plug-in electric vehicles (PEV) and plug-in hybrid electric vehicles (PHEV), in [19–21] the requirements are mentioned that must satisfy a SG to meet these challenges.

All these services and applications required by users of SG, SC and IoT, grow over time, like a tree that expands its leaves, in this way services implementation layers provided by SG and SC will be created them across different stages temporal, in addition to all this, the information flow must be conducted in a secure and scalable manner, on the different network segments how are: personal area network (PAN) and Home area network (HAN) see **Figure 1**, Neighborhood area network (NAN), WAN, and MAN see **Figure 2**.

1.1. Scalability of FiWi networks

Figure 2 shows different users who are geographically located in four subregions that form a planning area, these future clients will benefit from the services provided by SG, SC and IoT, such as smart metering (energy, gas, water) demand response, power storage, civil security, community alarms, smart public lighting, smart road signposting, etc.

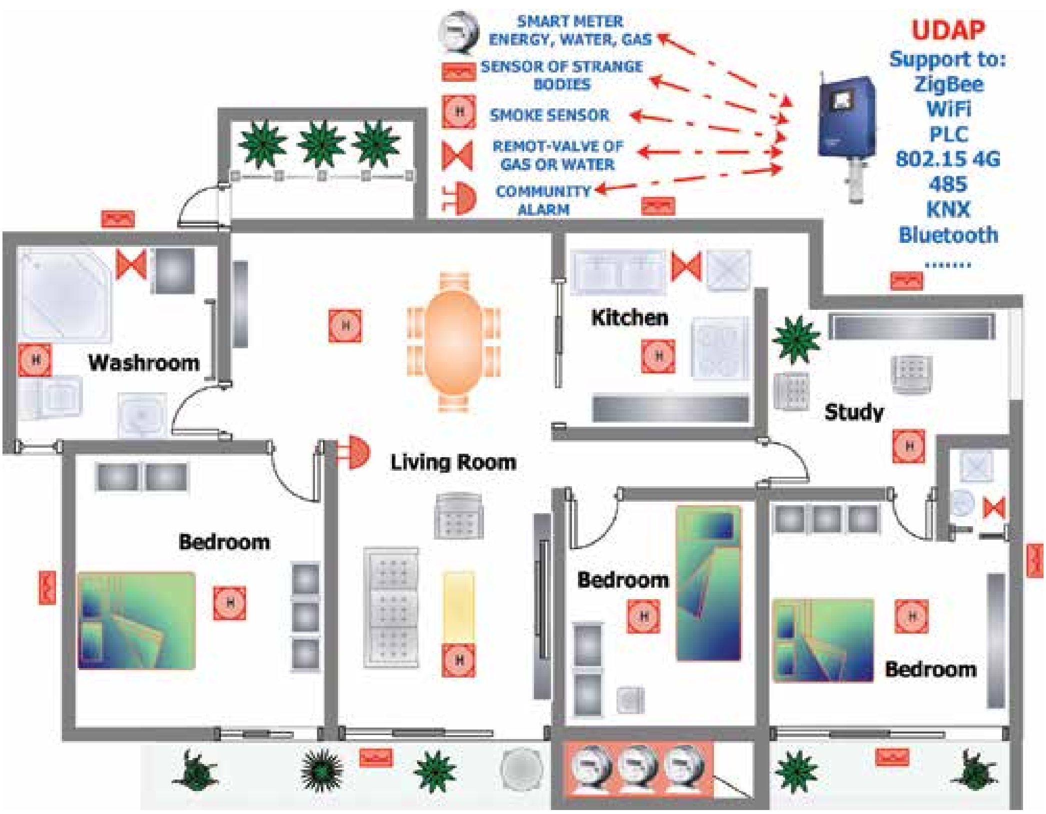

Figure 1. PAN and HAN networks for SG, SC and IoT.

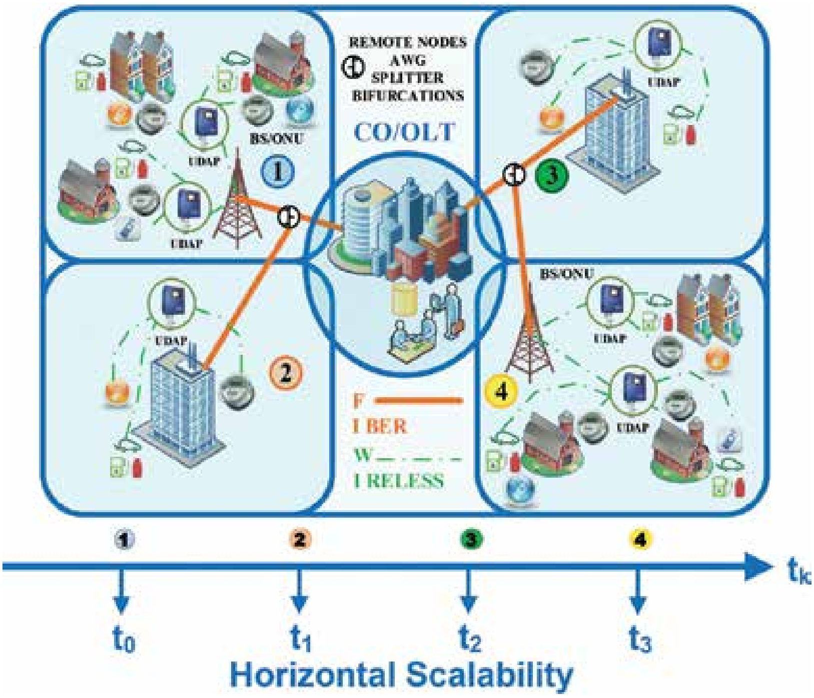

Figure 2. Infrastructure for SG and SC across a horizontally scalable FiWi network in four temporary stages.

To be able to offer these services it is necessary to have an adequate communications infrastructure throughout the region, to manage the flow of information together with the flow of energy. Therefore, **Figure 2** shows how the deployment of a FiWi network to cover and scale horizontally in a timeline to all subregions would be. The configuration of the architecture would be conformed to the wireless access through data concentrators that we have called UDAP. These devices have the ability to carry information from wireless heterogeneous network (WHN) [22], coming from the different wireless sensors network (WSN) to the base stations that function as enhanced node base station (eNodeB), which have a gateway that connects to the optical network unit (ONU) to send the information over a PON that makes fronthaul/backhaul [23, 24] (very high speed) using fiber-optic links in tree topology to optical line termination (OLT), where the co of the public or private service provider is. In the PON network, it is possible to place bifurcations that function as remote nodes (RN) where passive optical equipment such as Splitter (SP) or arrayed waveguide grating (AWG) can be located.

The proposed model seeks to guarantee a horizontal scalability in each stage of time $\boldsymbol{t}_k$, since by passing a time $\boldsymbol{t}_k$ to $\boldsymbol{t}_{k+1}$, **Fi**ber Optic and **Wi**reless resources are designated to the FiWi network, by means of actions and policies that add hardware in an optimal way, trying to give the greatest possible coverage to the users that evolve and grow spatially in a timeline and at the same time, returning the maximum economic benefit to investors represented by public or private companies.

This chapter is organized as follows: Section 2 we present the state of the art and related works; in Section 3 problem formulation with the planning model, and the algorithm MOA-FiWi; in Section 4 result analysis. Finally, in Section 5, we present the conclusions and future works.

2. State of art and related works

Aggregation points (AP) over the NAN play a very important role for the communications network that holds SG, so an adequate AP planning model that links to the HAN, can minimize costs in the deployment of SG and it is proposed in Ref. [25]. In addition to this premise, algorithms based on Greedy and clustering techniques are presented; these proposals presented analysis in power line communication (PLC) and optical fiber.

SG proposes a new concept in which electrical energy is generated, transported, distributed and consumed, thanks to the integration between telecommunications and advanced sensors to provide daily control and monitoring of the operation of the energy network within a WAN. Electricity is the key nucleus for the functioning of society and for the provision of services provided by technologies of information and communication (TIC). The works presented in Ref. [26] investigate the challenges and opportunities that can be achieved through the interaction of SG with green TIC, through efficient use of energy with wireless technologies and wired technologies such as PLC and fiber optics, present in the different domains that SG handles such as HAN, NAN and WAN.

On the other hand, the problem of efficient collection of measured data from AMI by reusing existing communications infrastructures such as the cellular network, but facilitated by a primary or secondary operator, the latter through a mobile virtual network operator (MVNO) or cognitive-virtual network operator presented in Ref. [27], requires to analyze the coverage problem in rural areas and the capacity of channels in urban areas due to the density of cellular telephone users. In other words, there is a need to allocate channels in an equitable way to reduce the costs in the lease of the spectrum of frequency.

Significant contributions have focused on the electric energy reserves, which can be managed by sending the information of the data measured by leased secondary channels at the lowest possible price. In order to reduce both costs of energy and communications, a problem called cost minimization for meter data collection (CMM) is formulated. This problem seeks to find an optimal solution for the minimization in the costs involved in the selection of communication channels and a scheme in the programming for the delivery of energy [28].

Within the AMI concept, the sub-steps that constitute the network topology for infrastructure planning must be determined. Thus, we have NAN [29, 30] delimited from the client meter to the UDAP concentrator with an uplink link [31], For this, conglomerates or clusters of smart meters (SM) are created to form NAN where cellular technology such as general packet radio service (GPRS)/long term evolution (LTE) [32] or WiFi and IEEE 802.15.4 g can be used through multiscales [33]. In this way the first stage is completed. Subsequently the different UDAP of NAN form a MAN, which forms the WAN [7, 34], but among NAN and MAN/WAN two solutions can be proposed for boundary zones; thus, we can continue to maintain a wireless cellular solution, WiFi or IEEE 802.15.4 g [35]. According to the coverage and the capacity of each UDAP that will be the one that finally allows the connectivity with the nearest cellular base station, but when the information demand grows substantially a fiber optic return is proposed [36–40], in addition to the interconnection of cellular base stations normally arranged for telephony, thus forming a hybrid network FiWi.

On the other hand, resources allocation is important for network operator profitability, therefore communications network must be dimensioned to satisfy customers' coverage and demand. Considering that these evolve over time, infrastructure must evolve accordingly. The demand growing is difficult to predict, in consequence it constitutes an important uncertainty source.

Strategic planning of communications network must take account this uncertainty, and network evolution must be able to adaptable to market conditions, therefore, the application of advanced planning methods taking into account the uncertainty can improve network profitability and create a competitive advantage. Wireless network planning demands complex tasks and automated procedures that must adapt and support large data demands that flow from current and future technologies, such as LTE, 4G and in a few years 5G.

There are very few contributions from the scientific community, regarding a planning framework that is suitable for various technologies and that demonstrates practical applicability by performing computational experiments using realistic and wide-ranging planning scenarios, where moreover network evolution start from an initial year and scale toward future years.

A popular method for evaluating investment opportunities in several domains with real options is presented in Ref. [41]. The real options approach treats investment projects as options of the outcome of future cash flows and uses the financial market for a neutral monetary valuation in the presence of risk when there are investment opportunities. The real options have been used as a tool in several applications, including telecommunications [42, 43]. In order to correctly apply the theory of real options, the project has to be embedded in an appropriate market.

Furthermore, stochastic programming can be useful as a tool to evaluate real options in the absence of a market embedding [44]. A discount rate must adjust the risk and be used to arrive to an outcome that is an implicit evaluation of paths that form scenarios over a stochastic decision tree.

Since communications network evolution can be divided into several stages, the multistage stochastic programming (MSP) [45] is an appropriate framework for modeling strategic planning on telecommunication networks. Wireless networks planning for cellular telephony through multistage stochastic programming is modeled in Ref. [46], it is left for future works to get a deeper analysis to be able to do FTTx networks planning.

Considering the aspects reviewed in the State of the Art, we can state that important work has been done on the analysis to save energy and provide greater capacity through the use of FiWi in multiservice networks that support SG, SC, and IoT. However, it is a priority to model mathematically in the presence of uncertainty how to deploy a FiWi network in a scalable way to propose a green field planning tool, or to perform access network upgrades or generate backup networks in case of failures. In addition, it is important to optimize the allocation of wireless and wired resources involved to meet the requirements of scalability, coverage and capacity, which is the main contribution and reason for the work we propose.

In this chapter we present a novel model of scalability of FiWi networks, on *Delaunay Triangulation Spaces*; which to our best understanding, it is the first time the combination of scalability

http://dx.doi.org/10.5772/intechopen.71781

analysis is considered (CAPEX and OPEX) introducing uncertainty in the different time-space stages, by multi-stage stochastic programming. The model presents flexibility in decision making as the time stages progress, and this situation allows the planning of green fields, as well as the updating of networks that already have communication infrastructure.

3. Problem formulation

The investigation problem seeks to make a resource allocation, over a temporal-spatial evolution, for communications infrastructure deployment, which will support provided services by SG, SC, and IoT, fulfilling with requirements of scalability and coverage, through use of wired and wireless mediums.

3.1. Planning model

The model proposed in **Figure 3** is divided into four phases which are described below:

3.1.1. Determination of parameters to be projected

- Characterize demand, population or density of users. In order to do this, it is important to have previous statistical data, which can be obtained by surveys, fieldwork or by comparison with previous projects.
- Then, a large number of projection scenarios are constructed at each stage of time. In order to fulfill this step, we can use Wiener stochastic processes (WSP), also known as geometric Brownian motion (GBM), whose model is represented in (1). This process is characterized by two parameters such as the expected growth rate μ and the volatility σ that generates uncertainty values at each time stage of a projection path. More features, properties and details of this stochastic process can be found in [47, 48].

$$\frac{d\mathbb{S}_t}{\mathbb{S}_t} = \mathrm{m}\ \ dt + \mathrm{s}\ \ d\mathbb{F}_t \tag{1}$$

- The evolution steps that generate a large number of scenarios are reduced by the techniques proposed in Ref. [49]. As a result a multistage stochastic projection tree (MSPT) is obtained.

3.1.2. Region of planning and location of candidate sites

With the data generated by the MSPT, the candidate sites for base stations, fiber-optic links, location of the central office and potential users will be located. These sites will evolve over a time line for each of the routes in the MSPT scenarios.

It is important to note that the coverage radios can be combined for macro, micro and femto cells. In this way, ***horizontal scalability*** (coverage requirements) and ***vertical scalability*** (increase capacity) can be given as users grow spatially in the planning region.

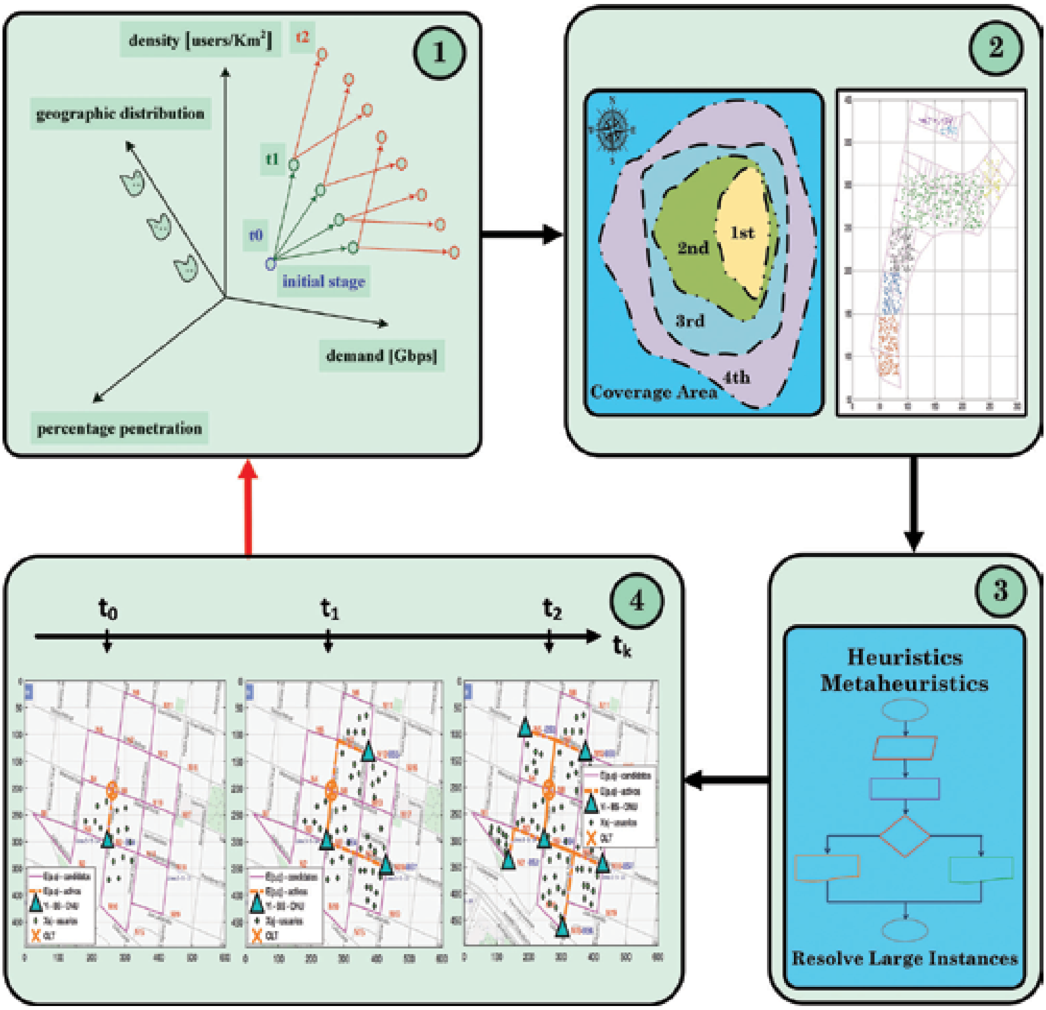

Figure 3. FiWi planning model.

Horizontal scalability, refers to the growth of the FiWi network over time, and in conformity with the evolution and growth of users. Therefore, this type of scalability does not observe the behavior of the process in a single instant of time (as a single photo or image of the scalability process). On the contrary, it is a process that changes, evolves and adapts automatically over time, according to the addition of hardware (base stations, fiber-optic links, etc.), from a time t_k to t_{k+1}.

On the other hand, with the ***horizontal scalability*** capacity is not guaranteed, it can even be very limited. Then, to address the issue of capacity, the issue of ***vertical scalability*** is stated, whose objective is to increase capacity without increasing the deployment of the communications infrastructure.

For ***vertical scalability***, it is important to clarify that it is not part of the scope of this work to perform exhaustive analysis of capacity, interference and topology performance that can result as a possible solution.

Carrier aggregation	Intra-carriers
	Inter-carriers
Spatial multiplexing using techniques	MIMO
	MISO
	SIMO
	SISO
Relay nodes	Microcells
	Femtocells
Fronthaul/Backhaul	NG-PON1
	NG-PON2
	DWDM
	UDWDM

Table 1. Alternatives of vertical scalability.

However, in general, there are alternatives for technological updating, which increase the capacity for each user, which is added to the future in a time $\boldsymbol{t_{k+1}}$.

Moreover, the optimization model proposed by mixed integer linear program (MILP) is very versatile to adapt it to any wireless and fiber-optic technology; for example, chosen as wireless resources to work under LTE-Advanced-4G, and for the wired fiber optic network xPON, the alternatives for technological updating would be those presented in **Table 1**.

3.1.3. Description of the MILP optimization model

There is a planning area $\mathbb{A}$ made up of users coming from services provided by SG and SC conglomerated through UDAP, situations which next will be represented by the binary variable $X_{s_j}^{n}$ whose value is one if one j^{th} UDAP is served and covered within a time stage $\boldsymbol{t_k}$ for a node n of MSPT, or zero otherwise; the information of $X_{s_j}^{n}$ is conveyed toward the base stations forming a $\mathbb{C}$ set of candidate cells for coverage, as long as restrictions are met at the energy thresholds that hold connectivity in wireless links, then, when a i^{th} base station is activated in a stage of $\mathrm{t_k}$ for a node n of MSPT, the variable Y_i^n becomes one or zero; on the other hand, any candidate cell in a parent node $p(n)$ of MSPT remains active when scaling horizontally from time $\boldsymbol{t_k}$ to time $\boldsymbol{t_{k+1}}$. It is possible to carry out the technological upgrades indicated in **Table 1**.

All the active base stations incorporate a Gateway that allows to migrate the information at high speed by ONU through fiber-optic links that are selected from a graph $\mathbb{G}(\mathbb{V}, \mathbb{E})$, configured by a grid of streets, Avenues and intersections that lie within $\mathbb{A}$. Consequently, if a link is activated in a time stage $\boldsymbol{t_k}$ for a node n of MSPT, the variable $Z_{p,q}^{n}$ becomes one or zero otherwise.

The active links form a PON network in tree topology, whose two-way information flows, go from the ONUs to the Central Office where they have OLT. It is not the object of this work to perform an analysis of intermediate passive equipment, such as optical splitters SP and optical routers AWG. This is justified because the major cost in an investment is in the construction of the PON network; that is to say, it is directly related to the laying of fiber-optic cable which also requires civil works such as conduits, pipelines and fittings to guide the laying of the transmission medium. However, if the positioning of bifurcations which act as RN is allowed, after the optimization process and depending on the configuration of the PON network, SP or AWG will be located. These equipment would be part of the technological upgrades indicated in **Table 1**.

Similar to what happens with active base stations, if a $Z^n_{p,q}$ link of optical fiber is chosen in a parent node $p(n)$ of MSPT, it remains active when scaling horizontally from time $\boldsymbol{t}_k$ to time $\boldsymbol{t}_{k+1}$, it is allowed to make branches to add RN, and make fronthaul/backhaul to the new cells that are going to be activated in the future. In this way, the reuse of guided transmission media such as the optical fiber is optimized.

Then, the proposed MILP model seeks to maximize the benefit expected by the investment in the deployment of the FiWi topology, $D^* : argmax\mathbb{E}\{\mathbb{R}\}$, there are two ways to solve it. The first is through the use of mathematical optimization software, which could only treat small instances of the problem, since if the proposed MILP model is simplified and relaxed in some restrictions, then we have the equivalent of a maximum coverage problem (MCP); in this way, it can be stated that the complexity present in the proposed optimization problem is ***NP-Hard*** type.

The second way to deal with the solution is to approximate feasible solutions by means of the heuristics and metaheuristics approach to provide ***computational scalability*** through polynomial models that do not grow exponentially to the size of the system; with this we could treat medium and large instances of the problem. The detailed formulation of the multistage stochastic optimization problem and the algorithms proposed on the basis of policies are discussed in the following subsections.

3.1.4. Appropriate horizontal scalability path

The last phase of the scalable planning model for FiWi networks to get the information optimized by the MILP model, is to perform an adequate analysis to make decisions.

In the last stage of time there are some scenarios, formed by paths that run through the MSPT, where each node n contains the topology FiWi and the UDAPs to be covered.

The scenarios can be classified as ***conservative, realistic and optimistic***, depending on the degree of uncertainty they have. The tools, such as the analysis of real options [44, 50], can help to select which ***horizontal scalability paths*** are suitable within the MSPT.

On the other hand, the model is dynamic and if necessary future scenarios can be reformulated at any stage of time, and the planning model process is the same as described in Section 3.1.1–3.1.4 presented in **Figure 3**.

3.2. Objective function

The $D^* : argmax\mathbb{E}\{\mathbb{R}\}$ is detailed in (2–12), It is important to indicate that all values are carried at net present value (NPV); CAPEX is loaded at the beginning of a year; income and OPEX at the end of a year and since there is no OPEX value for year zero, investment at both the beginning and the end of this year is high, giving negative cash flows in some cases:

$$argmax\mathbb{E}\{\mathbb{R}\} = \mathbb{R}_{profit}^{UDAP} - \mathbb{C}_{capex}^{BS} - \mathbb{C}_{capex}^{OF} - \mathbb{C}_{opex}^{BS} - \mathbb{C}_{opex}^{OF} \tag{2}$$

Subject To:

$$\mathbb{R}_{profit}^{UDAP} = \sum_{n\varepsilon\mathbb{N}} \mathbb{P}(n) \sum_{j\varepsilon\mathbb{A}} \widehat{\mathbb{I}}_j^n \; X_{s_j}^{\ n} \tag{3}$$

$$\widehat{\mathbb{I}}_j^n = \frac{\mathbb{I}_j^n}{(1+r)^{t_k}} \tag{4}$$

$$\mathbb{C}_{capex}^{BS} = \sum_{n\varepsilon\mathbb{N}} \mathbb{P}(n) \sum_{i\varepsilon\mathbb{C}} \widehat{\mathbb{C}}_i^{capex,\,n} \left(Y_i^n - Y_i^{p(n)}\right) \tag{5}$$

$$\widehat{\mathbb{C}}_i^{capex,\,n} = \frac{\mathbb{C}_i^{capex,\,n}}{(1+r)^{t_k-1}} \tag{6}$$

$$\mathbb{C}_{opex}^{BS} = \sum_{n\varepsilon\mathbb{N}} \mathbb{P}(n) \sum_{i\varepsilon\mathbb{C}} \widehat{\mathbb{C}}_i^{opex,\,n} \; Y_i^n \tag{7}$$

$$\widehat{\mathbb{C}}_i^{opex,\,n} = \frac{\mathbb{C}_i^{opex,\,n}}{(1+r)^{t_k}} \tag{8}$$

$$\mathbb{C}_{capex}^{OF} = \sum_{n\varepsilon\mathbb{N}} \mathbb{P}(n) \sum_{p,\,q\varepsilon\mathbb{E}} \widehat{\mathbb{C}}_{p,\,q}^{capex,\,n} \; \widehat{\mathbb{D}}_{p,\,q}^n \left(Z_{p,\,q}^n - Z_{p,\,q}^{p(n)}\right) \tag{9}$$

$$\widehat{\mathbb{C}}_{p,\,q}^{capex,\,n} = \frac{\mathbb{C}_{p,\,q}^{capex,\,n}}{(1+r)^{t_k-1}} \tag{10}$$

$$\mathbb{C}_{opex}^{OF} = \sum_{n\varepsilon\mathbb{N}} \mathbb{P}(n) \sum_{p,\,q\varepsilon\mathbb{E}} \widehat{\mathbb{C}}_{p,\,q}^{opex,\,n} \; \widehat{\mathbb{D}}_{p,\,q}^n \; Z_{p,\,q}^n \tag{11}$$

$$\widehat{\mathbb{C}}_{p,\,q}^{opex,\,n} = \frac{\mathbb{C}_{p,\,q}^{opex,\,n}}{(1+r)^{t_k}} \tag{12}$$

3.3. Wireless coverage restrictions

The constraints in (13–16) control the *horizontal scalability* provided by the maximum wireless coverage. The restriction (13) ensures that each UDAP has service and coverage, restriction (14) prevents base stations from being destroyed from parent nodes to child nodes in the MSPT. It should be noted that the parent root node in MSPT is reflected by the variable $Y_i^{p(r)} = 0$, in (15) The number of base stations constructed that can be added to those already existing from the parent node is limited to control propagated energy and consumed electrical energy; in this way, restriction (16) controls and ensures that the coverage is successful over the planning area $\mathbb{A}$ through the parameters α^n, coefficients $\mathbb{W}_j^n$ and variables $\mathcal{C}^n$.

$$\sum_{i\varepsilon\mathbb{C}} Y_i^n \geq X_{s_j}^n \qquad ;\forall \quad n\varepsilon\mathbb{N}, \quad j\varepsilon\mathbb{A} \tag{13}$$

$$Y_i^n \geq Y_i^{p(n)} \qquad ;\forall \quad n\varepsilon\mathbb{N}, \quad i\varepsilon\mathbb{C} \tag{14}$$

$$\sum_{i\varepsilon\mathbb{C}} \left(Y_i^n - Y_i^{p(n)}\right) \leq \mathbb{K}^n \qquad ;\forall \quad n\varepsilon\mathbb{N} \tag{15}$$

$$\sum_{j\varepsilon\mathbb{A}} \mathbb{W}_j^n \; X_{s_j}^n \geq \alpha^n |\mathbb{A}| \mathcal{C}^n \qquad ;\forall \quad n\varepsilon\mathbb{N} \tag{16}$$

3.4. Fiber-optic restrictions for Fronthaul/backhaul

Restrictions (17–20) are responsible for ensuring a scalable deployment of the fiber-optic fronthaul/backhaul. Restriction (17) prevents fiber-optic links from being destroyed from parent nodes to child nodes in the MSPT. In (18, 19), it is sought to ensure the routing of all flows $\mathbb{F}$ from the m–active cells to the Central Office-OLT by means of fiber paths having a minimum distance. On the other hand, (20) enforces that the active links correspond to each of the m flows.

$$Z_{p,q}^n \geq Z_{p,q}^{p(n)} \qquad ;\forall \quad n\varepsilon\mathbb{N}, \quad p,q\varepsilon\mathbb{E} \tag{17}$$

$$\sum_{q|p,\, q\varepsilon\mathbb{E}_p^{OUT}} Z_{p,q}^{n,m} - \sum_{q|q,\, p\varepsilon\mathbb{E}_p^{INPUT}} Z_{p,q}^{n,m} = \mathbb{R}_{p,i} \; Y_i^n \tag{18}$$

$$\mathbb{R}_{p,i} = \begin{cases} 1, & if \quad i = OLT \\ -1, & if \quad i = m \\ 0, & if \quad i \neq OLT \wedge i \neq m \end{cases} \tag{19}$$

$$\sum_{m\varepsilon\mathbb{F}} Z_{p,q}^{n,m} \leqslant \mathcal{M} \; Z_{p,q}^n \qquad ;\forall \quad n\varepsilon\mathbb{N}, \quad p,q\varepsilon\mathbb{E} \tag{20}$$

3.5. Dimensioning of variables

In (21) we place the dimensioning of all the decision variables involved in the MILP. Finally **Table 2** summarizes all the variables, constants, coefficients and parameters used in the formulation of the MILP model.

$$\begin{aligned} &X_s \varepsilon \{0,1\}^{\mathbb{N}\times\mathbb{A}} \\ &Y \;\; \varepsilon \{0,1\}^{\mathbb{N}\times\mathbb{C}} \\ &Z \;\; \varepsilon \{0,1\}^{\mathbb{N}\times\mathbb{E}\times\mathbb{F}} \\ &\mathcal{C} \;\; \varepsilon \{0,1\}^{\mathbb{N}} \end{aligned} \tag{21}$$

3.6. MOA-FiWi algorithm

In order to treat medium or large instances of the problem, a new **M**ultistage **O**ptimization **A**lgorithm for **Fi**ber/**Wi**reless networks, called **MOA-FiWi**, has been proposed.

Name	Domain	Interpretation
Sets		
$\mathbb{A}$	$\subseteq\Re^3$	Planning area, divided into pixels
$\mathbb{C}$	$\varepsilon\mathbb{Z}$	Set of candidate cells for coverage
$\mathbb{F}$	$\varepsilon\mathbb{Z}$	Set of m flows
Tree scenario		
$\mathbb{N}$		Set of MSPT nodes
$\mathbb{P}(n)$	$\varepsilon(0,1]$	Probability at node n
$p(n)$	$\varepsilon\mathbb{N}$	Parent node at MSPT
Coefficients and parameters		
$\mathscr{M}$	$\varepsilon\Re >> 0$	It is a sufficiently large number $> \|\mathbb{F}\|$
$\mathbb{K}^n$	$\varepsilon\mathbb{Z}$	Construction limit at node n
α^n	$\varepsilon[0,1]$	Coverage requirement parameter
$\mathbb{W}_j^n$	$\varepsilon[0,1]$	Weight on a pixel in node n
$\widehat{\mathbb{I}}_j^n$	$\varepsilon\Re\geq 0$	Revenue per pixel at node n
$\widehat{\mathbb{C}}_i^{capex,n}$	$\varepsilon\Re\geq 0$	NPV of CAPEX in cell i at node n
$\widehat{\mathbb{C}}_i^{opex,n}$	$\varepsilon\Re\geq 0$	NPV of OPEX in cell i at node n
$\mathbb{D}_{p,q}^n$	$\varepsilon\Re\geq 0$	Distance for link $p\leftrightarrows q$ at node n
$\widehat{\mathbb{C}}_{p,q}^{capex,n}$	$\varepsilon\Re\geq 0$	NPV of CAPEX for link $p\leftrightarrows q$ at node n
$\widehat{\mathbb{C}}_{p,q}^{opex,n}$	$\varepsilon\Re\geq 0$	NPV of OPEX for link $p\leftrightarrows q$ at node n
Decision variables		
Y_i^n	$\varepsilon\{0,1\}$	Cell i is active at node n
$X_{s_j}^n$	$\varepsilon\{0,1\}$	UDAP j is covered at node n
$Z_{p,q}^n$	$\varepsilon\{0,1\}$	Link $p\leftrightarrows q$ is active at node n
C^n	$\varepsilon\{0,1\}$	Fulfillment of coverage at node n

Table 2. Variables, coefficients, and parameters of the MILP.

The main optimization base of multistage optimization algorithm for fiber-wireless hybrid networks (MOA-FiWi) is through a set of actions and policies π to provide the maximum coverage to the UDAPs that carry the information of the users that benefit from the services provided by SG, SC and IoT. Therefore, it must be kept in mind that the amount of UDAP grows according to each MSPT path, according to this growth, the resource designation to form the FiWi network must horizontally scale in time and space.

The set ξ represents the universe of possible geographical locations over time for UDAPs within the planning region $\mathbb{A}$. Then, π_k a suitable policy depends on the values taken by the spatial locations of the UDAPs; $\xi : \xi^{'}\subseteq\xi; \pi_k^*\left(\xi^{'}\right)$, consequently the expected maximum benefit depends on the policies (22).

$$D^* : argmax\mathbb{E}\{\mathbb{R}\} = \mathbb{E}\left(\pi_k^*\right) = \tilde{\mathbb{R}}\left(\pi_k^*\right) \tag{22}$$

The policies are in charge of activating and optimally locating the base stations on candidate sites, for this a ***Modified Set Covered*** is used and fiber-optic links form a PON network through the help of a ***Modified Dijkstra*** in tree topology between the ONUs and the candidate site chosen to place the OLT, this location gives the reference point where the evolution of the FiWi network begins, this must be fulfilled for all nodes that form the MSPT, **Algorithm 1** details MOA-FiWi.

Algorithm 1. MOA-FiWi.

Step:1 **Generate:**
$MSPT(n,t)$
Step:2 **Generate:**
$\xi_0' \subseteq \xi$
Step:3 **Generate:**
$\pi_0^*\left(\xi_0'\right)\ \forall\ MSPT(n,t)$
Step:4 **Calculate:**
$\mathbb{E}\left(\xi_0'\right) = \tilde{\mathbb{R}}\{\pi_0^*\}$
Step:5 **Generate new:**
$\xi_{k+1}' = f_1\left(\xi_k', \xi\right)$
Step:6 **Modify:**
$\pi_{k+1}^* = f_2\left(\pi_k', \xi_k'\right)\ \forall\ MSPT(n,t)$
Step:7 **Apply:**
Decision criterion&Stopped
Step:8 **Go to Step 5:**
If criterion does not meet
Step:9 **Return:**
$D^* : argmax\mathbb{E}\{\mathbb{R}\}$

4. Result analysis

To exemplify the operation of moa, a planning region $\mathbb{A}$ delimited by a graph $\mathbb{G}(\mathbb{V},\mathbb{E})$ on a ***Delaunay Triangulation Space*** has been generated, within which a large number of UDAPs will be deployed, providing access to an average of ten to twenty users benefiting from the services provided by SG and SC. The coverage of the region will be distributed over four stages of time $t_k \rightarrow \{0, 1, 2, 3, 4\}$. **Figure 4(a)** shows a geographic distribution of the planning area which is a component of the subset ξ_0'. On the other hand, **Figure 4(b)** presents the MSPT for the four temporal stages.

In each node the projected population of UDAP is indicated and the value of the probability that measures the degree of uncertainty. At the end there are six scenarios, two considered as conservative, two as realistic and two as optimistic, being the point of break from year one.

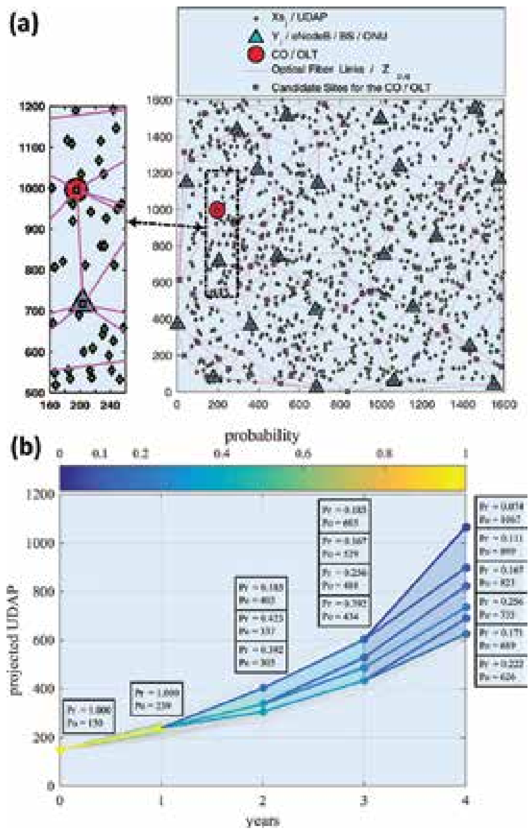

Figure 4. (a) Geographical distribution of an ξ_0' component. (b) MSPT for four stages of time.

Moreover, to obtain the reduced MSPT, one hundred paths were projected with $\mu = \mathbf{0,4}$ and $\sigma = \mathbf{0,1}$.

Table 3 summarizes the incomes and reference costs in US dollars, consulted with three telecommunications operators. These data are considered as input for MOA-FiWi. In addition,

Annual benefit per UDAP	\$ 400,00
CAPEX OLT, with capacity for 1000 users, type XG-PON	\$ 45.000,00
CAPEX eNodeB/ONU	\$ 25.000,00
CAPEX per meter includes optical fiber, supports, pipes, and ducts	\$ 15,00
Annual OPEX of eNodeB/ONU	\$ 200,00
Annual OPEX per meter optical fiber	\$ 1,20

Table 3. Revenues and costs considered in MOA-FiWi.

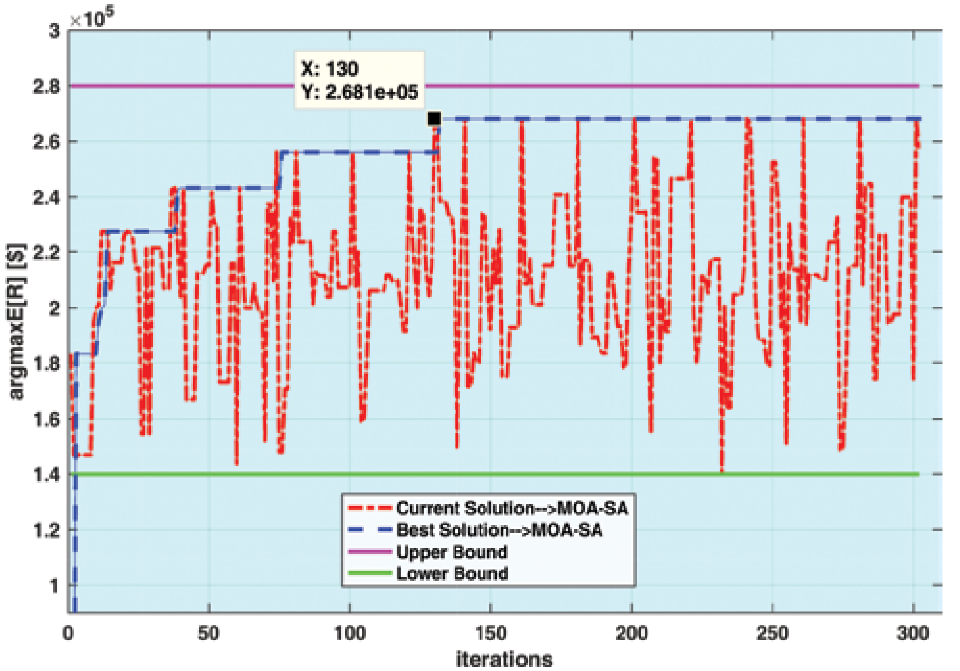

Figure 5. Solutions found by MOA-FiWi-SA.

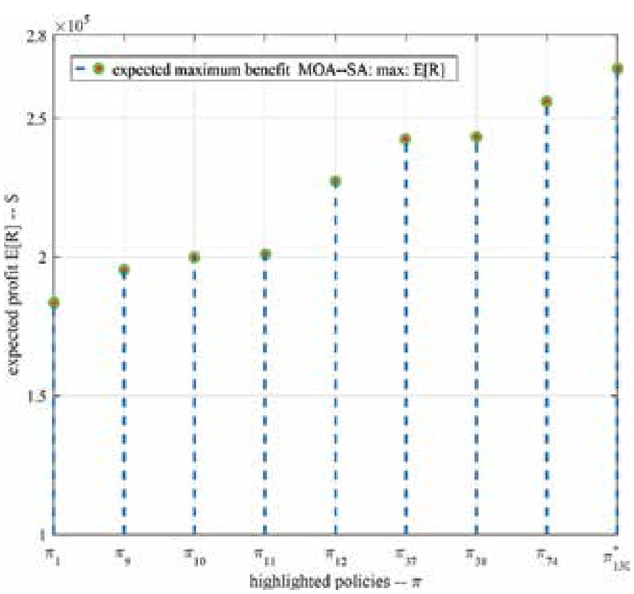

Figure 6. Featured policies found by MOA-FiWi-SA.

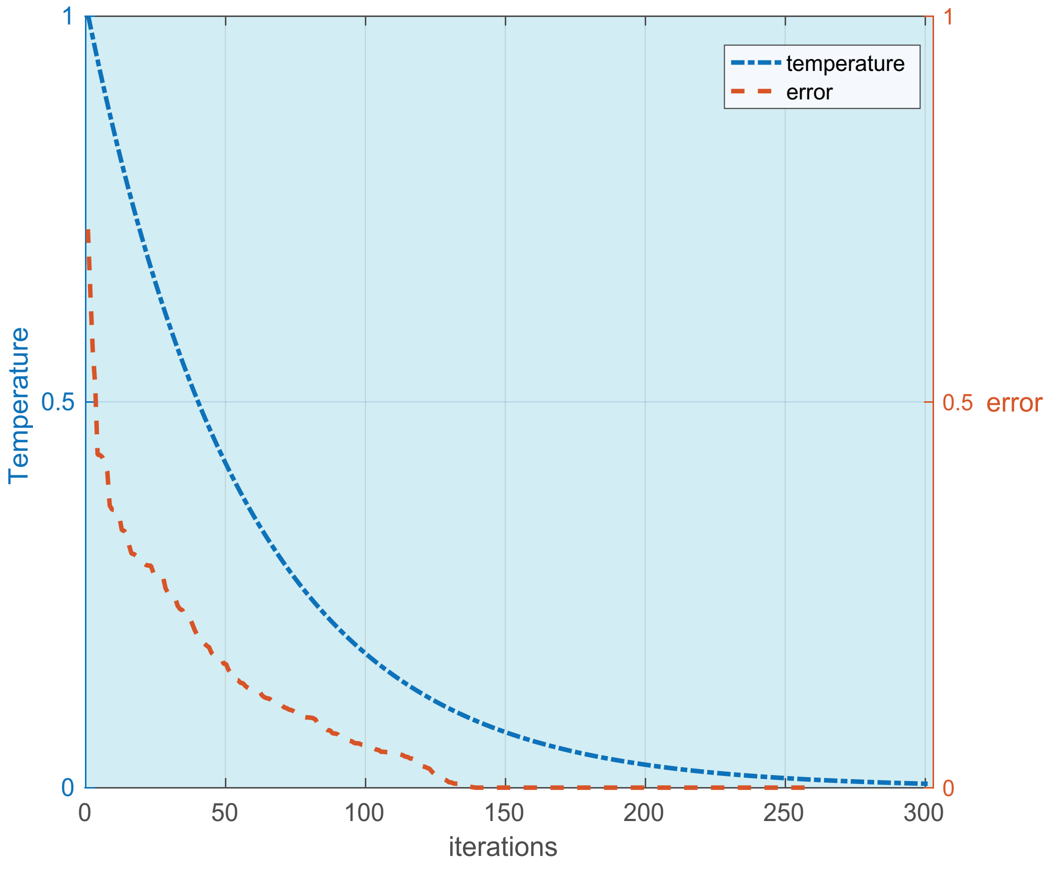

Figure 7. Cooling and error curve for *simulated annealing metaheuristic.*

the simulations were performed with a discount rate $r = 9, 57\%$. **Figure 5** shows the evolution of the value D^*, over the search space by means of a ***simulated annealing metaheuristic***; the main policies are presented in **Figure 6**, and the maximum expected profit was achieved in the 130^{th} policy π^*_{130}, given 300 iterations.

Moreover, **Figure 7** exhibits the behaviors of temperature curve and error curve, in response to the optimization process using ***simulated annealing metaheuristic***; therefore, the behavior of MOA-FiWi is adequate, improving the feasible solutions found in each iteration.

The maximum expected benefit reached in 130^{th} policy on MSPT, where of the six stochastic paths after performing a decision-making analysis, paths with the best result were scenarios one and three. **Figure 8** presents topologies of these two scenarios.

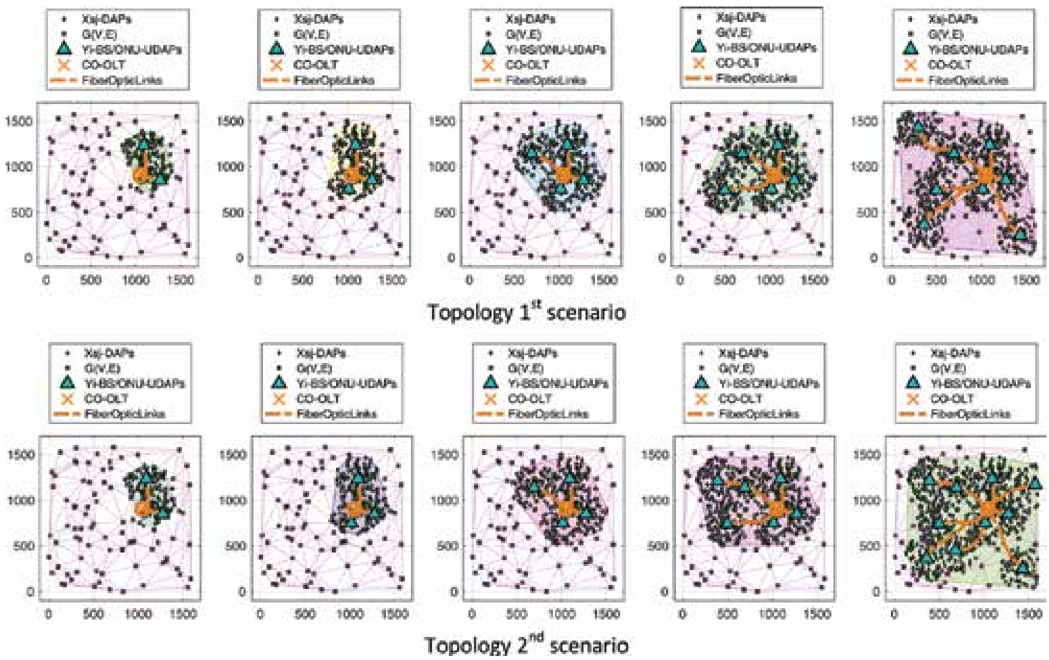

Figure 8. Best solution found by MOA-FiWi, 130^{th} policy on MSPT.

5. Conclusions and future works

- SG, SC, and IoT applications require the power network to support a bidirectional flow of energy, so that users can interact with it to be able to deliver power to the system; in addition, a two-way flow of information between end users and service providers is required. For this reason the communications network that support services provided by SG, SC, and IoT plays a primary role, guaranteeing scalability, coverage, bandwidth, latency, reliability, security and privacy. These requirements must be fulfilled on all segments how are HAN, NAN, MAN and WAN.

- As services provided by SG and SC increase, the demand and coverage of IEDs increase with time, consequently communications infrastructure has to evolve and scale in parallel, to achieve this purpose the application of planning methods advanced under conditions of uncertainty would help to make decisions to network operators and improve their profitability and competitiveness for adapt changing market conditions.

- The main contributions of this work are the proposal of a planning model to treat scalability of FiWi networks, based on four phases, in addition a new mathematical optimization model MILP is proposed, through use of MSP, that it is destined to solve a problem that reaches a degree of complexity NP-Hard, however this is ideal as a tool to make decisions when communications network planning that presents uncertainty in demand growth, according different services that could be anchored over existing wireless networks such as are cellular networks. The optimization model focuses on achieving a scalable planning of fiber-optic network used as fronthaul/backhaul of wireless network, forming a FiWi hybrid network, which evolves over a space-time line.

- Being a stochastic problem, gives possibility and alternative of measuring the risk or benefit playing with actions and policies taken in each projected scenario, therefore, possible solutions can be approached from several points of view and not from one, as is case of deterministic planning model. MSPT allows it to find important breakpoints to take actions and policies that mark new forms of horizontal scalability in the topology of FiWi network that supports the services provided by SG, SC, and IoT.
- In order to deal computationally with multistage stochastic planning, an algorithm called MOA-FiWi has been proposed, where the optimization and stopping criterion were evaluated using simulated annealing metaheuristic. MOA-FiWi is based on optimization actions and policies, which provide horizontal scalability over a timeline and in presence of uncertainty; such situation occurs in real life when projects of expansion, updating or implementation of communications infrastructure are executed.
- On the other hand, obtained results reveal that there is a great sensitivity in maximum expected benefit, according to how the designation of wired and wireless resources in time and space is done to give maximum coverage to the users, with proposed model can be simulated to the problem from different points of view. As a result, a planning tool is available which helps in analysis to make decisions.
- Finally, for future works is intended to treat vertical scalability, with the purpose of improving performance and capacity of the system, in addition, to compare several technologies used in planning of FiWi networks, also try other metaheuristics that would help to explore the search space in a better way, to obtain feasible solutions.

Acknowledgements

This work is supported by the research groups GITEL of the Universidad Politécnica Salesiana, Cuenca, Ecuador, and the research group GIDATI of the Universidad Pontificia Bolivariana, Medellín, Colombia.

Glossary

AMI	advanced metering infrastructure
AMR	automated meters reader
AP	aggregation points
AWG	arrayed waveguide grating

BS	se stations
CAPEX	capital expenditure
CMM	cost minimization for meter data collection
CO	central office
DER	distributed energy resources
DG	distributed generation
DNOs	distribution network operators
eNodeB	enhanced node base station
EV	electric vehicles
FiWi	hybrid networks of optical fiber links combined with wireless links
GBM	geometric Brownian motion
GPRS	general packet radio service
HAN	home area network
HFC	hydrogen fuel cells
IED	intelligent electronic devices
IoT	internet of things
LTE	long term evolution
MAN	metropolitan area network
MCP	maximum coverage problem
MILP	mixed integer linear program
MOA-FiWi	multistage optimization algorithm for fiber-wireless hybrid networks
MSP	multistage stochastic programming
MSPT	multistage stochastic projection tree
MVNO	mobile virtual network operator
NAN	neighborhood area network
NPV	net present value
OLT	optical line termination
ONU	optical network unit
OPEX	operational expenditure
PAN	personal area network

PEM	polymer electrolyte membrane
PEV	plug-in electric vehicles
PHEV	plug-in hybrid electric vehicles
PLC	power line communication
PON	passive optical network
RN	remote nodes
SC	smart cities
SG	smart grids
SM	smart meters
SP	splitter
TIC	technologies of information and communication
UDAP	universal data aggregation point
WAN	wide area network
WHN	wireless heterogeneous network
WSN	wireless sensors network
WSP	Wiener stochastic processes

Author details

Arturo G. Peralta

Address all correspondence to: aperaltas@ups.edu.ec

Universidad Politécnica Salesiana, Cuenca, Ecuador

References

[1] Gao J, Xiao Y, Liu J, Liang W, Philip Chen CL. A survey of communication/networking in smart grids. Future Generation Computer Systems. 2012;**28**(2):391-404

[2] Sun Q, Ge X, Liu L, Xin X, Zhang Y, Niu R, Zeng Y. Review of smart grid comprehensive assessment systems. Energy Procedia. 2011;**12**:219-229. The Proceedings of International Conference on Smart Grid and Clean Energy Technologies ICSGCE 2011

[3] Usman A, Shami SH. Evolution of communication technologies for smart grid applications. Renewable and Sustainable Energy Reviews. 2013;**19**:191-199

[4] Aurzada F, Lévesque M, Maier M, Reisslein M. Fiwi access networks based on next-generation pon and gigabit-class wlan technologies: A capacity and delay analysis. IEEE/ACM Transactions on Networking. Aug 2014;**22**(4):1176-1189

[5] Bade MG, Toycan M, Walker SD. Cost and energy efficient operation of converged, reconfigurable optical wireless networks. Optical Switching and Networking. 2015;**18**(Part 1):71-80

[6] Coimbra J, Schütz G, Correia N. Energy efficient routing algorithm for fiber-wireless access networks: A network formation game approach. Computer Networks. 2014;**60**:201-216

[7] Ghazisaidi N, Maier M. Fiber-wireless (fiwi) access networks: Challenges and opportunities. IEEE Network. Jan 2011;**25**(1):36-42

[8] Lévesque M, Maier M. Probabilistic availability quantification of pon and wimax based fiwi access networks for future smart grid applications. IEEE Transactions on Communications. Jun 2014;**62**(6):1958-1969

[9] Liu J, Guo H, Nishiyama H, Ujikawa H, Suzuki K, Kato N. New perspectives on future smart fiwi networks: Scalability, reliability, and energy efficiency. IEEE Communications Surveys Tutorials. Secondquarter 2016;**18**(2):1045-1072

[10] Zhou J, Qingyang Hu R, Qian Y. Scalable distributed communication architectures to support advanced metering infrastructure in smart grid. IEEE Transactions on Parallel and Distributed Systems. Sep 2012;**23**(9):1632-1642

[11] Niyato D, Wang P. Cooperative transmission for meter data collection in smart grid. IEEE Communications Magazine. Apr 2012;**50**(4):90-97

[12] Gungor VC, Sahin D, Kocak T, Ergut S, Buccella C, Cecati C, Hancke GP. Smart grid technologies: Communication technologies and standards. IEEE Transactions on Industrial Informatics. Nov 2011;**7**(4):529-539

[13] Fan Z, Kulkarni P, Gormus S, Efthymiou C, Kalogridis G, Sooriyabandara M, Zhu Z, Lambotharan S, Chin WH. Smart grid communications: Overview of research challenges, solutions, and standardization activities. IEEE Communications Surveys Tutorials. First 2013;**15**(1):21-38

[14] Galli S, Scaglione A, Wang Z. Power line communications and the smart grid. In: 2010 First IEEE International Conference on Smart Grid Communications. Oct 2010. pp. 303-308

[15] Feng Z, Yuexia Z. Study on smart grid communications system based on new generation wireless technology. In: 2011 International Conference on Electronics, Communications and Control (ICECC). Sep 2011. pp. 1673-1678

[16] Overman TM, Sackman RW. Smart grid control systems communications architecture. In: 2010 First IEEE International Conference on Smart Grid Communications. Oct 2010. pp. 19-24

[17] Bu S, Yu FR, Liu PX, Zhang P. Distributed scheduling in smart grid communications with dynamic power demands and intermittent renewable energy resources. In: 2011 IEEE International Conference on Communications Workshops (ICC). Jun 2011. pp. 1-5

[18] Xiao Y. Communication and Networking in Smart Grids, volume 1 of Taylor & Francis Group. 1st ed. CRC Press; 2012. https://www.taylorfrancis.com/books/e/9781439878729

[19] Tuttle DP, Baldick R. The evolution of plug-in electric vehicle-grid interactions. IEEE Transactions on Smart Grid. Mar 2012;**3**(1):500-505

[20] Kaebisch S, Schmitt A, Winter M, Heuer J. Interconnections and communications of electric vehicles and smart grids. In: 2010 First IEEE International Conference on Smart Grid Communications. Oct 2010. pp. 161-166

[21] Bouhafs F, Mackay M, Merabti M. Links to the future: Communication requirements and challenges in the smart grid. IEEE Power and Energy Magazine. Jan 2012;**10**(1):24-32

[22] Inga E, Cespedes S, Hincapie R, Cardenas CA. Scalable route map for advanced metering infrastructure based on optimal routing of wireless heterogeneous networks. IEEE Wireless Communications. Apr 2017;**24**(2):26-33

[23] Bhatt UR, Chhabra A, Fiber-wireless RU. (fi-wi) architectural technologies: Asurvey. In: 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT). Mar 2016. pp. 519-524

[24] La Oliva AD, Perez XC, Azcorra A, Giglio AD, Cavaliere F, Tiegelbekkers D, Lessmann J, Haustein T, Mourad A, Iovanna P. Xhaul: Toward an integrated fronthaul/backhaularchitecture in 5g networks. IEEE Wireless Communications. Oct 2015;**22**(5):32-40

[25] Huang X, Wang S. Aggregation points planning in smart grid communication system. IEEE Communications Letters. Aug 2015;**19**(8):1315-1318

[26] Erol-Kantarci M, Mouftah HT. Energy-efficient information and communication infrastructures in the smart grid: A survey on interactions and open issues. IEEE Communications Surveys Tutorials. Firstquarter 2015;**17**(1):179-197

[27] Inga E, Hincapie R. Matched channel allocation for advanced metering infrastructure based on cognitive mobile virtual network operator. IEEE Latin America Transactions. April 2016;**14**(4):1780-1785

[28] Li P, Guo S, Cheng Z. Joint optimization of electricity and communication cost for meter datacollection in smart grid. IEEE Transactions on Emerging Topics in Computing. Dec 2013;**1**(2):297-306

[29] Yu R, Zhang Y, Gjessing S, Yuen C, Xie S, Guizani M. Cognitive radio based hierarchical communications infrastructure for smart grid. IEEE Network. September 2011;**25**(5):6-14

[30] Yu R, Zhang Y, Chen Y. Hybrid spectrum access in cognitive neighborhood area networks in the smart grid. In: 2012 IEEE Wireless Communications and Networking Conference (WCNC). April 2012. pp. 1478-1483

[31] Huang J, Subramanian VG, Agrawal R, Berry R. Joint scheduling and resource allocation in uplink ofdm systems for broadband wireless access networks. IEEE Journal on Selected Areas in Communications. February 2009;**27**(2):226-234

[32] Challita U, Al-Kanj L, Dawy Z. On lte cellular network planning under demand uncertainty. In: 2014 IEEE Wireless Communications and Networking Conference (WCNC). April 2014. pp. 2079-2084

[33] Milling C, Subramanian S, Shakkottai S, Berry R. Routing over multi-hop wireless networks with non-ergodic mobility. In IEEE INFOCOM. April 2009;**2009**:2079-2087

[34] Lin B, Tapolcai J, Ho PH. Dimensioning and site planning of integrated pon and wireless cooperative networks for fixed mobile convergence. IEEE Transactions on Vehicular Technology. Nov 2011;**60**(9):4528-4538

[35] Aalamifar F, Shirazi GN, Noori M, Lampe L. Cost-efficient data aggregation point placement for advanced metering infrastructure. In: 2014 IEEE International Conference on Smart Grid Communications (SmartGridComm). Nov 2014. pp. 344-349

[36] Maier M, Levesque M, Ivanescu L. Ng-pons 1 & 2 and beyond: The dawn of the uber-fiwi network. IEEE Network. Mar 2012;**26**(2):15-21

[37] Karimi B, Namboodiri V. On the capacity of a wireless backhaul for the distribution level of the smart grid. IEEE Systems Journal. June 2014;**8**(2):521-532

[38] Peralta Sevilla A, Inga Ortega E, Hincapie R. Fiwi network planning for smart metering based on multistage stochastic programming. IEEE Latin America Transactions. Dec 2015;**13**(12):3838-3843

[39] Peralta-Sevilla A, Inga E, Cumbal R, Hincapié R. Optimum deployment of fiwi networks using wireless sensors based on universal data aggregation points. In: IEEE Colombian Conference on Communication and Computing (IEEE COLCOM 2015). May 2015. pp. 1-6

[40] Peralta-Sevilla AG, Amaya-Fernández F, Hincapie R. Multiservice hybrid wdm/tdm-pon dimensioning using a heuristic method. In: 2014 IEEE Colombian Conference on Communications and Computing (COLCOM). Jun 2014. pp. 1-5

[41] Trigeorgis L. Real Options: Managerial Flexibility and Strategy in Resource Allocation. The MIT Press; 1996. https://mitpress.mit.edu/books/real-options

[42] Ramirez W, Harmantzis FC, Tanguturi VP. Valuing wireless data services solutions for corporate clients using real options. International Journal of Mobile Communications. Feb 2007;**5**:259-280

[43] Beltran-Royo C, Escudero LF, Monge JF, Rodriguez-Ravines RE. An effective heuristic for multistage linear programming with a stochastic right-hand side. Computers & Operations Research. 2014;**51**:237-250

[44] Wang T, De Neufville R. Building real options into physical systems with stochastic mixed-integer programming. In: Proceedings of the 8th Real Options Annual. 2004. pp. 1-35

[45] Shapiro A, Dentcheva D, Ruszczynski A. Lectures on Stochastic Programming: Modeling and Theory. 2nd ed. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics; 2014. http://epubs.siam.org/doi/book/10.1137/1.9780898718751

[46] Eisenblätter A, Schweiger J. Multistage stochastic programming in strategic telecommunication network planning. Computational Management Science. 2012;**9**(3):303-321

[47] Dixit AK, Pindyck RS. Investment under Uncertainty. Princeton University Press; 1994. https://press.princeton.edu/titles/5474.html

[48] Shreve SE. Stochastic Calculus for Finance II. Springer; 2004. http://www.springer.com/la/book/9780387401010#aboutAuthors

[49] Growe-Kuska N, Heitsch H, and Romisch W. Scenario reduction and scenario tree construction for power management problems. In: 2003 IEEE Bologna Power Tech Conference Proceedings, Vol. 3; Jun 2003. pp. 1-7

[50] Peters L. Real Options Illustrated. SpringerBriefs in Finance. Cham: Springer International Publishing; 2016

Chapter 18

Economic Interests and Social Problems in Realization of Broadband Network

Milan Ivanović

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72037

Abstract

Investments in broadband access are very useful for local community, especially for the underdeveloped and developing countries. The emphasis is on importance of broadband infrastructure and the use of Internet in the world, the EU and the Republic of Croatia as one of its member state. Implementation analysis of the "Slavonian Network" project in Slavonia, (Croatia region) for the period 2012 to 2017, points to a number of problems that were recorded in development of the broadband and to technical, economic, legal, and social issues as well. The density of broadband connections and users in Croatia is significantly lower than the average of the EU, and in five counties in the Slavonia region, this average is in turn lower than in Croatia. This state of affairs prevents social and economic development, effective functioning of the public administration, and inclusion of the region in modern communication within the country and within EU. The construction of broadband infrastructure is a significant economic and technical activity for all countries, especially for the countries lagging behind in economic and technological development. This paper points to the model of regional approach to building broadband infrastructure that can be a good model for all developing countries.

Keywords: broadband, economic interests, infrastructure, local development, society

1. Broadband network

Scientific knowledge and new civilization's knowledge grow exponentially in time; at the beginning of the twenty-first century (2005–2010), thus, the amount of knowledge in scientific disciplines has doubled every 5 years and in some areas in less than 2 years. Expansion and accumulation of knowledge depend today on establishing communication network for fast and efficient data transmission [1, 2]. Broadband is a civilization tool that provides the fastest data transfer today. The chapter deals with broadband access conditions for existing networks in the world, particularly in Europe, and economic issues related to the construction of broadband networks. The example of one Croatian region refers to social problems that are characteristic of the construction of infrastructure installations in the transition of underdeveloped countries.

1.1. Importance of broadband network

Development of fast access networks expresses the same revolutionary effect in society today as it has been during the development of railway or power grid 150 years ago. Modern information and telecommunication technology (ITT) has significantly altered the lifestyle of people over the last 20 years by accelerated data transfer, increase of their quality and reliability, by reduced operating costs, by accelerated business transactions and by providing fast access to the global market. New ITT is the foundation of knowledge economy. Instead of capital, information and knowledge have become the basis of individual and social growth and development. Digital content and applications are expected to be covered almost completely by Internet delivery after 2020 [3, 4]. Concepts and approaches to building smart cities have already been set up [5]. However, development of high-quality, fast, reliable, and inexpensive public services is important for the whole country as well. This is important for public sector state administration and local self-government, healthcare, education, as well as for business of the entire economy, and encouraging development of rural and underdeveloped areas in each country. Development and implementation of new ITT are particularly important for developing countries because increased coverage of ITT infrastructure that enables high-speed Internet reduces migration of rural population to cities or developed countries.

1.2. Categorizing the availability of broadband access

The accessibility to broadband is categorized on maps with three shadings:

a. **White areas** include areas where broadband access is not available, or where there is no adequate broadband infrastructure.

b. **Gray areas** are those where only one operator offers broadband services or multiple operators offer with insufficient level of competition, resulting in an inadequate supply of broadband services for end users in terms of quality and price of services.

c. **Black areas** are those where at least two operators offer broadband services with a satisfactory level of competition, i.e., the quality and price of services for users.

This categorization of broadband services is carried out in two levels:

a. Basic or traditional broadband access based on copper wires that allow data transmission up to 10 Mbit/s, or with some additional procedures up to 30 Mbit/s.

b. NGA (broadband access network of new generations) based on fibers that enable speeds faster than 30 Mbit/s.

1.3. Developmental effects of broadband access

Developmental effects of broadband infrastructure are socially positive and can be classified into four main sectors: (a) education, (b) health and social care, (c) employment and economic development, and (d) energy and transport—as is already discussed in Refs. [6–10].

http://dx.doi.org/10.5772/intechopen.72037

1.4. Development models of broadband infrastructure in EU

Building a fiber optic network is infrastructure project that cannot be funded only as a private venture investment. EU has set out Digital Agenda [4] as a strategic document for equal access to broadband Internet for entire EU population. The viability of such investments does not motivate operators or private companies to invest in sparsely populated or rural areas. For these reasons, a number of EU funds have been offered and state aid grants approved for the construction of broadband infrastructure, all under special conditions in order not to undermine the rules of free market competition.

The development of broadband infrastructure in EU countries has been intensive for a decade and several models have been developed: (a) business, (b) markets, and (c) investment, determined by: (1) competitive services, (2) authorities and operators (private companies), (3) investment shares and responsibilities for infrastructure design and management, and (4) acquisition and retention of ownership over the built infrastructure. This complex issue of building broadband infrastructure further aggravates the number of models of possible use of telecommunication technologies, as each of them has its advantages and disadvantages. It is therefore important to get acquainted with all the essential elements in the process of broadband access planning.

1.5. Financial sources for broadband infrastructure construction

Construction of broadband infrastructure in large cities and urban agglomerations ensures quick return of investment in several years so that telecommunication companies (private entrepreneurship) fund this construction and will continue to finance it—according to market criteria [11, 12]. For other areas—smaller cities and rural areas—models of stimulated broadband infrastructure construction have been developed and their sources can be divided into three basic groups:

a. **Public funds**—include all budget funds at the national, regional, and county levels and local level (cities and municipalities), as well as all funds invested by publicly owned companies. Funds from the EU Structural Funds-the European Regional Development Fund, the European Social Fund, and the EU Cohesion Fund-are also considered as public funds. From EU funds, it is possible to cofinance infrastructure projects up to 85%, while other funds are secured from national budget sources.

b. **Private funds**—include funds from private operators in the electronic communication market and, eventually, end user funds that may be involved in cofinancing the implementation of broadband infrastructure, commonly the end-segments of the access network.

c. **The funds of institutional investors**—institutional investors are considered as banks and investment funds, including pension funds. Since their primary interest is the realization of economic gains, institutional investors appear to be the investors of broadband infrastructure projects only in the most densely populated areas (mostly black areas) where they become sustainable business models.

State subsidies are justified in white and mostly in gray areas, while in black areas, they are not justified. The share of grants in project financing is rising to less populated areas (generally white areas) and can reach (in special cases) up to 100%. In contrast, the share of private fund operators increased to more densely populated areas (gray and black areas), while decreasing the share of public funds in the financing of projects. Public funds can be invested in black areas only under market conditions together with the funds of private operators and institutional investors [12, 13].

2. Investment in broadband network infrastructure

Broadband network investments are complex, because of fast-growing broadband technology, the issues of monopoly and network exploitation, and different uses of broadband and spectrum of users. It is important for each investment project to study a variety of details ranging from area mapping choice of investment model construction costs to the cost of exploitation in order to determine the term of the investment return.

2.1. Mapping areas

Mapping area is a spatial view of the area with data on availability status of basic and NGA broadband access of all operators. The purpose of mapping is to determine areas where it is justified to carry out individual projects based on the rules and guidelines of the Framework National Program (FNP). The shading of area is determined as follows [12–16]:

- White areas: areas where there is no adequate broadband infrastructure and no operator plans to build broadband infrastructure in the next 3 years,
- Gray areas: areas where there is an operator's broadband network and no other operator plans to build an additional network in the next 3 years,
- Black areas: areas where there are at least two broadband networks belonging to two different operators, or according to expressed interest of operator, there are at least two networks planned to be built in the next 3 years.

2.2. Cost of broadband network building

The cost of building a broadband network depends on the above density of access, and here are three zones: (a) black = highest density, large urban agglomeration, (b) gray = smaller density, smaller cities, and (c) white = poorly populated area, villages. According to this criterion, fixed cost per single connection is also shared.

In the example of the UK and similarly in other EU countries, 67% of the connections have lowest cost of building broadband connection in each regional area. It is twice as expensive to build a connection in gray zone amounting to next 23% of connections, and cost of building a connection in white zone is three times higher than in first zone, **Figure 1**. For this reason, the European Commission (EC) cofinances investments in white and gray zone according to certain criteria [12, 15–18].

The share of land works in costs of communal infrastructure construction such as plumbing, sewage, public lighting, water supply, and public infrastructure such as electric underground

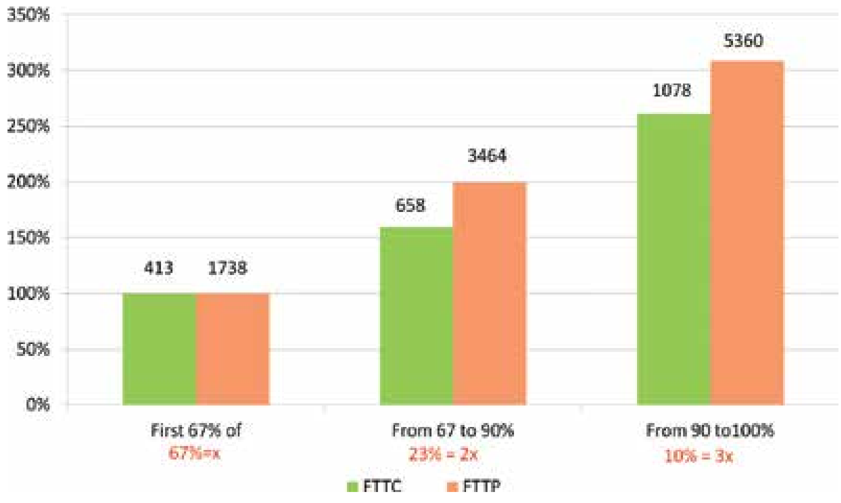

Figure 1. Deployment cost per premises connected (£) [17].

network and gas pipeline in Croatia ranges from 20 to 60% of total investment depending on the type of terrain. The share of land works in the construction of fiber optic infrastructure (FOI) is about 70%. This fact suggests that integrated approach can achieve a significant reduction in investment costs in telecommunication infrastructure construction. More on this is given in Ref. [14–21]. For this reason, EC has adopted Directive 2014/61/EU on measures to reduce the cost of setting up high-speed electronic communication networks [22] and requests from Member States to jointly use FOI and other types of electricity, gas, pipeline, water, drainage, and hydroelectric infrastructure, commonly referred in the Directive as "physical infrastructure."

2.3. Operating costs and return on investment

Costs of exploitation and profit making in broadband networks that is realized as return on investment depend on several criteria: number and density of connections and type and scope of service use. Thus there exist different exploitation models in economic and technology segment. The return on investment in large urban agglomerations with high density of connectors and large consumption of different services is 2–4 years. The return on investment in smaller cities with lower density and volume of spending is about 8 years, while in poorly populated areas, this amounts to 20 years [18].

2.4. Investing models in broadband access

Broadband infrastructure implementation projects can be carried out using several investment models that are defined by the relationship between public authorities and operators as private entrepreneurs in the project. These relationships include investment shares, responsibility for network building, management, and acquisition and retention of ownership over the built infrastructure [17, 18, 22]. The following investment models are most frequently used in practice:

- **Bottom up** model (local community model) involves a group of end users in the local community that are jointly owned and a democratically controlled group capable of overseeing a local network construction contract. The public sector here is limited to granting aid—as a guarantor of loans and/or facilitating access to the public property of infrastructure such as FOI.
- **Public DBO** model encompasses all cases in which the implementation of broadband infrastructure construction is performed under supervision of public authorities without any private sector assistance and where the ownership of built infrastructure remains permanently in public ownership. Public DBO model requires significant engagement of administrative and technical capacities of local government bodies, but it allows for long-term preservation of public interest. A public-sector operating company can operate on the entire network or it can only provide wholesale services and allow private operators to sell retail services.
- **Private DBO** model includes cases where private operators are beneficiaries of granted grant right to build and manage infrastructure with permanent retention of ownership over the built infrastructure. This model does not require significant involvement of public authorities in project implementation. In this context, the protection of public interest is limited as the infrastructure built with incentives remains the property of private operator.
- **Joint Venture** model is contract based whereby ownership of the network is shared between public and private sectors. Construction and operational functions are likely to be undertaken by the private sector. The model implies a joint investment venture of local authorities and private operators possibly with financial hunt of institutional investors. It is possible in this way to balance public interest by coverage of broadband infrastructure and interests of private investors who can achieve economic gain.
- **Public model** of external service; this model is similar to private DBO model, with a difference that infrastructure built up by public incentives remains in public ownership after expiration of external service contract. Under this model, one contract is awarded for all aspects of construction and operation in network. The main feature of this model is that the network is set up by private sector, but public sector retains ownership.

2.5. Investment models in Croatia

The construction and expansion of broadband network are complex processes and it is necessary to point to the specificity of investment in this sector as this infrastructure by its nature cannot be a monopoly of private entrepreneurs. Thus, three investment models were defined in Croatia, defined by the ratio of investment stakeholders, public bodies, and private companies. Models define the responsibility for building and managing the network and the acquisition and retention of ownership of the built infrastructure [15, 16, 19, 23].

- Model "A": private DBO model; in this model, responsibility for design, construction, and operation of the network is left to private operator, and built-up network remains owned by that operator. The design of network means here the process of making detailed technical specifications of network construction project according to relevant regulations and based on general specifications is made by the local community project manager (PM).

- Model "B": public DBO model in which responsibility for design and construction and network management lies on local government bodies and the built-up network remains in permanent public ownership. Local authorities, as project manager (PM), are fully responsible for implementing the project on model "B." In some activities, PM may engage specialized private companies due to lack of administrative capacity and/or expertise in public authorities. Also, in the case of network maintenance and management in Model "B," private companies can be engaged, whereby it is essential that, in the management, private companies do not get right to collect fees from end users of the network because such access would have characteristics of concession.
- Model "C": public-private partnership (PPP); this investment model combines models "A" and "B." In the context of broadband infrastructure construction projects, private partner in PPP model can take responsibility for design, construction, management, and maintenance of the network and also partly cofinances network building by remaining part of network financing that provides a public partner through state aid. The constructed network returns to the public property after the expiry of duration of PPP contract, but not later than 40 years.

3. Development of broadband network

3.1. Contemporary broadband access in the world

Over the past 20 years, technological advances have been particularly perceived in the development of computing and telecommunications; a whole range of new devices and product services have been developed, developed and built infrastructure, service and product cost are reduced, a large part of population has mastered the use of information technology, and billions of people around the world are interconnected and communicating; the world is moving faster toward the information society. Industrial and information developed countries are at forefront in these processes; **Figures 2** and **3** and **Table 1** illustrate the above.

3.2. European development framework

The Europe 2020 Strategy brings vision of the European social market economy for the twenty-first century and proposes three complementary priorities: (1) smart growth: developing a knowledge-based and innovation-based economy; (2) sustainable growth: promoting a more efficient use of economies that are "greener" and more competitive; and (3) inclusive growth: maintaining an economy with a high employment rate that brings social and territorial connectivity [27]. The Digital Agenda for Europe has adopted concrete measures and targets and recommended deadlines for meeting these goals in the area of broadband access at European Union level in order to achieve the greatest benefits from such development for the economy and the EU population. The Digital Agendas for Europe's pillars are: (1) a single digital market; (2) improving interoperability and standards; (3) strengthening confidence in online and security; (4) promoting fast and ultrafast internet access for all: (a) basic access coverage: 100% of EU population by 2013, (b) fast access (30 Mbit/s or more) coverage: 100%

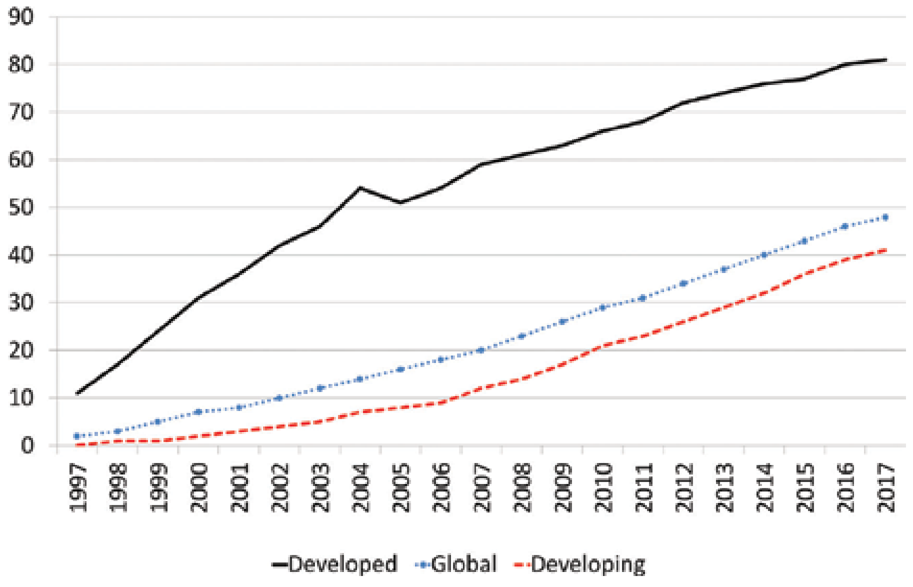

Figure 2. Internet users per 100 inhabitants [24].

of EU population by 2020, (c) ultrafast access (100 Mbit/s or more) use: 50% of EU households by 2020; (5) investments in research and development; (6) promoting digital literacy, skills, and digital inclusion; and (7) benefits for European society that enables ICT.

Investments in the development of broadband access are very useful for the community as stated in a series of studies prepared for the European Commission (EC). The increase of the number of broadband access users has an impact on GDP growth, and the impact is more significant as the country is more developed. Estimates of possible GDP growth are 0.47% in countries with less developed broadband, 0.63% in countries with faster growth of broadband, 0.7% in large industrial countries, and 0.89% in the most developed countries—which fully utilize all possibilities of the knowledge society. EC study [8] further explores the above assumptions

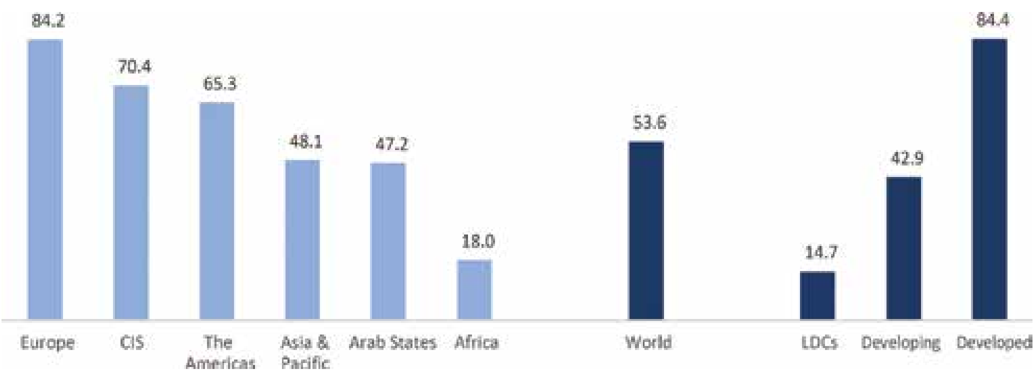

Figure 3. Broadband household penetration by region, 2016 (ITU estimates, %) [24].

World regions	Population		Internet users 30 June 2017	Penetration rate % population
	2017 (estim.)	% World		
Africa	1247	16.6	388	31.2
Asia	4148	55.2	1938	46.7
Europe	823	10.9	660	80.2
Latin America/Caribbean	648	8.6	404	62.4
Middle East	250	3.3	147	58.7
North America	363	4.8	320	88.1
Australia/Oceania	41	0.5	28	69.6
World (total)	7519	100	3886	51.7

Source: [25, 26].

Table 1. World internet usage and population statistics—June 30, 2017.

and directly relates four indicators to the benefits of broadband access: average income, number of computer users, number of smart phone users, and network coverage. Based on the estimates of the direct and indirect benefits of the development of broadband access to the analysis, EU could have direct benefits worth between 2.2 and 3.2 bil. € in the period from 2010 to 2019. It also states that in general terms a 10% increase in the number of users of broadband enables increase of GDP growth by 1.38%, which is reflected in the increase in the number of jobs in network development and maintenance activities and the growth of economic activity due to increased use of electronic services available through broadband access [28, 29].

EU Member States, alongside the efforts of the EC to promote broadband access development, independently adopt national plans and strategies for broadband access development. National plans differ from member states, whereby the following common trends can be noticed: plans relate to 3–5 years for a basic broadband approach and 7 years for fast and ultrafast broadband access. Goals are set to cover a certain percentage of households, i.e., households with broadband access at a certain minimum speed. Funds are provided to achieve the set goals. **Figures 4–7** show dynamics of broadband development in EU.

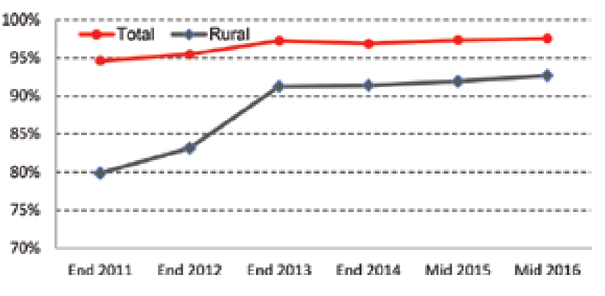

Figure 4. Fixed broadband coverage in EU, 2001–2016 [29].

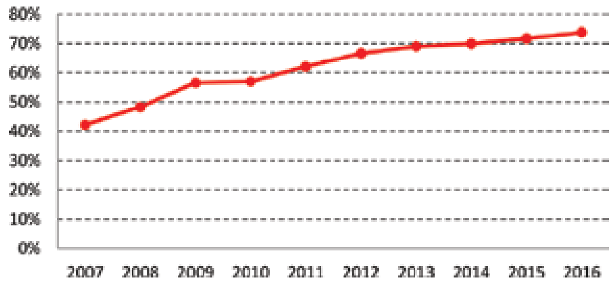

Figure 5. Household with a fixed broadband subscription in EU, 2007–2016 [30].

3.3. Development frameworks in Croatia

The Government of the Republic of Croatia adopted the 2011 Strategy for Development of Broadband Access in Croatia 2012–2015 [32] and Strategy for the period 2016–2020 [33]. Planning and construction of broadband networks NGA is financially very demanding depending on geographical characteristics and geo-demographic state. According to characteristics of settlements, Croatia (4.3 mil. inhabitants) is predominantly rural with an average population density of 75.7 inhabitants per km^2; a third of population is concentrated in 10 largest cities. Majority of Croatia possesses a spatially dispersed population, with a large number of smaller settlements; approximately one third of population lives in 211 settlements between 2000 and 30,000 inhabitants, while the last third of population lives in remaining 6384 settlements with less than 2000 inhabitants.

Croatia possess a 23.02% coverage in fixed public communications network, while EU's average is 31.6%. Broadband access map (2 Mbit/s) in Croatia (**Figure 9**) at city/municipality level shows large digital gap between urban and rural areas. Average availability of LTE technology in the EU Member States was 85.9% (from 100% in Norway and the Netherlands to 48.1% in Bulgaria) in June 2015. The availability of LTE technology in Croatia is 68.9% by June 2015. The progress made in implementing the NGA broadband infrastructure is unsatisfactory. According to Eurostat in June 2015, Croatia was at the mercy of EU members. Infrastructure availability in Croatia was 52%, and in the EU, 70.9% [30].

Construction costs of NGA networks throughout Croatia are about 1.65 bil €. The state of broadband access is shown in **Figures 8–10**. Croatia has shown a slow growth in density of

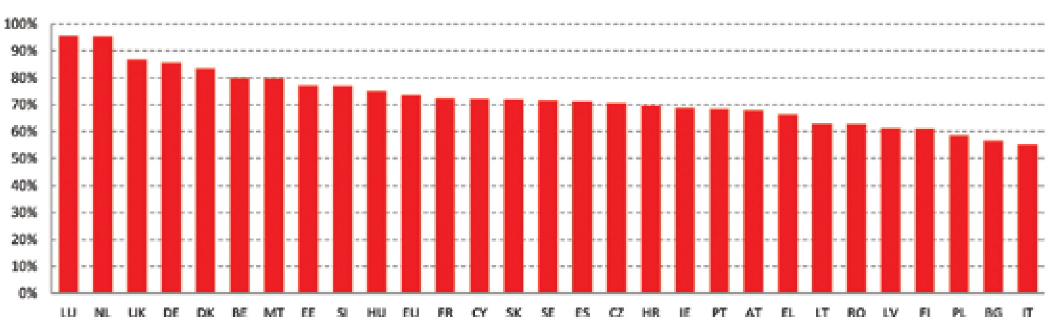

Figure 6. Household with a fixed broadband subscription in EU by countries, 2016 [30].

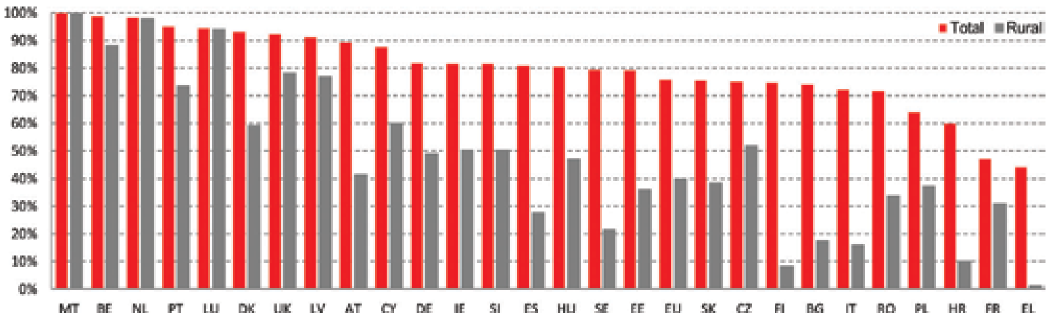

Figure 7. Broadband Internet access in EU countries—fixed network NGA (FTTP, VDSL, etc.,) [31].

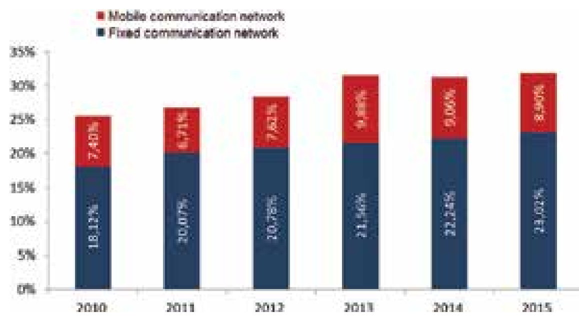

Figure 8. Density of broadband access in Croatia [33].

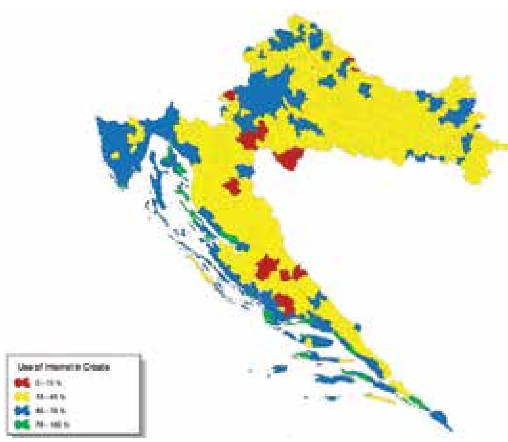

Figure 9. Map of Internet use in Croatia (speeds greater than 2 mbps) [33].

broadband connections from 2010 to 2015, **Figure 8**, but this is still insufficient, as Croatia lags behind average of EU, especially in fiber optic access networks.

The Slavonia region, which makes for quarter of Croatia, has been lagging behind the state average for the last 30 years. The reason for this was the 1991–1995 war destruction, mostly agrarian

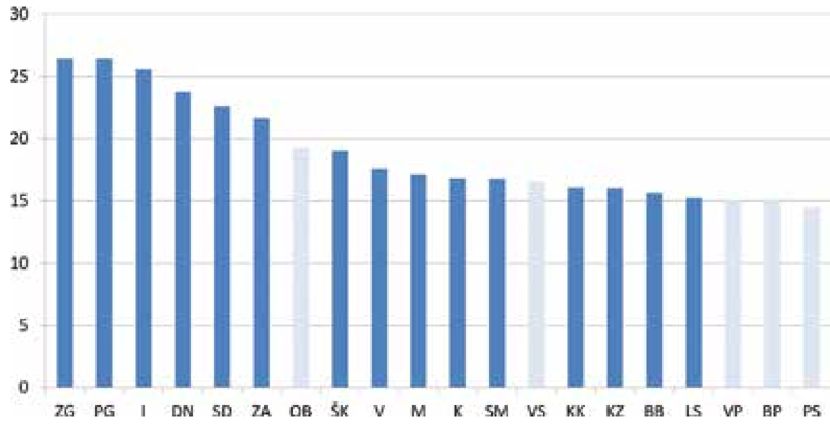

Figure 10. Densities of broadband access by Croatia counties—fixed network, 2011 [32].

production, and weak central government policy. The unemployment rate in Slavonia with once-developed industrial centers in the region is twice as high as the average of Croatia, so that significant inhabitant's migration to other Croatian regions or to countries of Western Europe has started in recent years. That is why, it was significant to launch the project "Slavonian Network." According to the density of broadband connections from the five Slavonian counties, three were at the end of the list (Požega, Virovitica, and Brod), one county (Vukovar) in the middle, and one (Osijek) in the first third of the county list in Croatia (gray columns).

4. The "Slavonian network" project

After adoption of the Broadband Access Development Strategy for Croatia from 2012 to 2015 (October 2011), in February 2012, Faculty of Electrical Engineering Osijek (FEE), in cooperation with Croatian Regulatory Agency for Networking, organized the conference "Development telecommunication infrastructure - strengthening competitiveness and efficient investment of local self-government," and it was attended by leaders of many municipalities, cities, and five Slavonian counties. At the end of that year, FEE initiated project "Broadband Access Development" with several engineer employees.

4.1. Launching the project

At the initial stage of the project, FEE colleagues called new researchers to team and create an interdisciplinary expert team (IET) consisting of PhDs, masters, and graduates of ICT, geodesy, economics, sociology, and law—from FEE and private companies "Geoprem" Ltd. Osijek (geodesy jobs), "Sokol" Ltd. Vinkovci (telecom_infrastructure construction company), and Panon—think tank for strategic studies Osijek. At that time, a new project concept was developed that covered the entire region (five counties of Slavonia), and the project was named "Slavonian Network."

The research team has set more provisions on future project activities that—apart from research on the state of telecommunication network and potential for future broadband use—need to launch a series of activities that the project needs to do interdisciplinary. The first provision

Figure 11. Five counties of eastern Croatia—Slavonia region [34].

(decision) of IET was (1) that the project will be structured modularly and (2) that results of research by individual modules will be published at scientific and professional conferences—in order to test hypotheses and at the same time promote the project, informing and mobilizing experts and local public about launching broadband in the region by gathering new experts and creating new teams for project work. Second important decision is that the IET will accept and follow the overall processing and rules of European funds to support broadband construction. Basic elements of this project were set up after introductory research, starting from goals of the National Broadband Strategy and local opportunities as:

- **Project implementation area**: five counties in Slavonia with 22 cities, 105 municipalities, and 998 settlements with more than 800 thousand inhabitants (**Figure 11**).
- **Project goal**: enable broadband and Internet access for 75% of population in five Slavonian counties by 2015.
- **The final beneficiaries of the project results are**:

1. Population in the five counties of the region,
2. State and public services in five counties of the region (health, education, social welfare, and public administration), and
3. Economy in the area of five counties of the region.

- **Evaluation of economic benefit** from the realization of the project:

a. More efficient functioning of public administration,

b. Better business results of economic entities,

c. Better and higher standard of living of the population, and

d. Development of new business-based broadband activities.

Ultimately, project realization can only contribute to GDP growth of 0.7% in the region—starting with project implementation.

- **Project tasks**:

a. Information and mobilization of local units for:

1. Determining of the telecommunication state,
2. Arranging cadastre of the electronic communication infrastructure (ECI) lines,
3. Arranging spatial plans for ECI,
4. Making a decision on collection of rentals for the use of existing ECIs,
5. Arranging relationships with ECI users,
6. Unifying funds from the road right to ECI,
7. Constructing and developing broadband,
8. Study of the state of local telecom infrastructure,
9. Developing a broadband funding study.

b. Establishment of a consortium for activities coordination of local units,

c. Professional assistance to local units—cities and municipalities—in technical, legal, and economic areas,

d. Expert assistance to local units in the design of bidding projects,

e. Regional project application for funds in Croatia and EU,

f. Launch of the macroproject "Development of Broadband Services in the Slavonia Region."

In the new concept of the project, it is emphasized that it is about:

a. Important issue of Croatia's technological connection to European communication flows,

b. Complex technological process for the development of broadband services,

c. Significant investment engagement,;

d. Demanding task of determining the status in local units, and

e. Important elements of spatial plans of local units.

Following internal discussions, IET concluding that it would be wrong to focus the whole project only on (a) activities of broadband construction, the team decided to start simultaneously with a wide spectrum of development activities. These activities should be initiated: (b) for future distribution of network maintenance and development of local broadband services, (c) for education of institutions, companies, and inhabitants for the use of broadband services; and (d) to motivate and direct students FEE to create applications for broadband services. In order to realize the project, it is necessary to unite all social, professional, and financial potentials in the region. With regard to the efficient implementation of the project, establishment of the "Slavonian Network" consortium was suggested, which would harmonize procedures and coordinate implementation of key stages of the project. Members of the consortium would be representatives of five Slavonian counties, Faculty of Electrical Engineering, Osijek, Panon think tank, Osijek, and telecomm operators, which express a business interest in this field. In addition to team interdisciplinary work, two basic stages of launching broadband in the region are structured, **Figures 12** and **13**.

4.2. Project development

In March 2013, IET created the study "Slavonian Network: Broadband Network Development in the Five Counties of the Slavonia Region" [34], which was submitted to the Ministry of Regional Development; between 500 projects received, "Slavonian network" has entered a shortlist of 50 projects and ranked 11th. Project "Slavonian network" was the only project

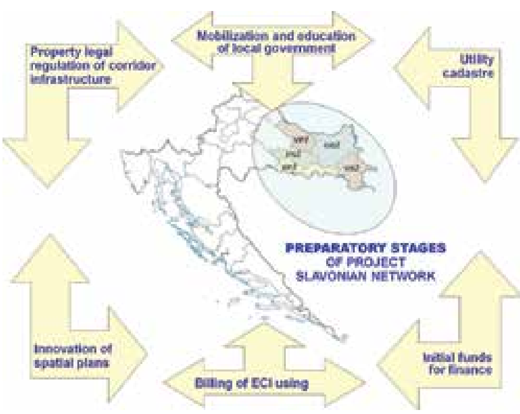

Figure 12. Structure of preparatory phases of "Slavonian network" project [35].

Figure 13. Structure of the implementation phases of "Slavonian network" project [35].

that united activities of more than one county and was the first project in the area of broadband access; this study involved five pilot projects of construction of broadband access in five Slavonian counties with a total value of 21.5 mil. €.

During the first half of 2014, project modules of the project were reviewed by expert teams of Ministry of Regional Development; in all three phase, reports of experts for assessment of preparation of the project received positive ratings. The project is foreseen to start immediately in implementation of the preparatory stages of "Slavonian network" (**Figure 12**), and when they are finished preparatory stages and provide financial resources for the implementation will be started second stages of the project (**Figure 13**).

4.3. Project implementation

Unfortunately, after the project got all the positive criticisms, FEE left that project and the project was taken over by the University of Osijek. The University has set up a new team of researchers who did not have broadband network references, and at the same time excluded all IET members and researchers who made the project concept, wrote, and developed this project from the project work. New team and project management did not realize the project as is accepted, but already started toward realization of the project according to its ideas. So they lost the next 4 years in an attempt to set up the company "Slavonian network" Ltd. who would carry out the project [36]. This company was registered only in late 2016. The founders of this company were University of Osijek and five counties of Slavonia. The Company was established as the executive pole for the implementation of the project "Slavonian network" in the area of five Slavonian counties.

In coming years—2014 and 2015 and 2016—new project management has periodically sought to choose a project management model and coordination of activities in municipalities, cities, and counties, i.e., IET proposal on founding of consortium has not been realized. During these years, new project team performed sporadically and partially the activities of the proposed modules of the "Slavonian network" (2013) project [19]. In this way, structured in the modules, the project "Slavonian network" which was highly ranked on the list of the Ministry of Regional Development and EU Funds and which received positive reviews, stopped halfway. Nearly 4 years have been lost in which all preparatory stages of the project and a large part of implementation stages could have been realized. Members of the new team did not make any new analysis for all 4 years, nor did they publish any professional or scientific paper on the broadband. Although the new project management and the established company "Slavonian network" Ltd. has taken over all the activities for the implementation of the project, "Slavonian network" IET has continued informal work—considering strategic issues of project—elaborated the project modules "Slavonian network" and published over 30 professional and scientific papers on domestic and international conferences and journals about this [37].

4.4. Analysis of results

During (informal) monitoring of preparation activities of Slavonian local units for development of broadband access, IET has noted:

a. A number of local units did not regulate fees from the right to service (renting), i.e., those municipalities and cities that regulated funds received from rent used the funds for other purposes instead of financing necessary business on developing broadband network.

b. In past years, more than 90 municipal projects on infrastructure in the Slavonia region were detected; with realization of integrated construction principle, significant savings that would have a positive (financial and temporal) impact on realization of broadband access can be achieved; namely, building or reconstructing on hundreds of kilometers of physical infrastructure (plumbing, sewage, gas, and public lighting)—that opportunity was not used for laying plastic pipes to which fiber optics could be later drawn—which would reduce construction costs by up to 70%. This recommendation was issued by the IET at the beginning of 2013, and there is also a commitment under the EC Directive on Integrated Construction (2014).

c. Leaders and expert services of local units are not educated about importance of broadband access, development potential of this project, and complexity and structure of these investments.

d. In Croatia, local units have started to develop Broadband Infrastructure Development Programs (BIDP) since 2015; local units from Slavonia were delayed in this activity; only a few BIDPs were created in mid-2017.

This statement is true of BIDPs in Slavonia, and the actual case analysis was published in the paper [23]. Insufficiently informed and unprepared local units in Slavonia started the development of local plans for BIDPs only at the end of 2016 and in early 2017. Only six BIDPs are prepared. **Tables 2** and **3** show the state of BIDP in local units in the region of Slavonia; only 40 cities and municipalities (out of 127) have prepared broadband network development programs within 6 BDIPs, and 5 of them have selected investment model "A," **Tables 2** and **3**.

The basic findings are:

- In case of five BIDPs, the investment model "A" (private DBO model) was chosen whereby responsibility for design, construction, and operation of network is left to private operator, whereby network is owned by that operator. This means that local units who receive EU grant for broadband construction lend their investment rights to investors or to private operators and at the same time they miss a number of development opportunities for their local community: such as rental income from broadband infrastructure built-up, network management, and development of their own local communications and establishment of a local network maintenance service, all of which can raise local potentials, among others, employment of local experts as the best means to reduce emigration of young people abroad.
- After review of all these BIDPs, it can be concluded that in these documents, there are no essential elements from original projects of IET (**Figures 12** and **13**) which was submitted to Ministry of Regional Development (2013).

BIDP	№ local units	% local units	% settlements	% inhabitants	% households
Yes	40	31.5	27.2	24.8	24.6
No	87	68.5	72.8	75.2	75.4
Slavonia	127	100	100	100	100

Table 2. State of BIDPs in Slavonia region—October 2017.

Invest_model	№ local units	% local units	% inhabitants	% households
A	39	30.7	24.1	24.1
B	1	0.8	0.7	0.7
C	0	0	0	0

Source: https://nop.hakom.hr./

Table 3. Proposed investment models in Slavonia region—October 2017.

Thus BIDPs didn't elaborate on:

1. Management of (future local) fiber optic network and its maintenance;
2. Integrated construction (in the construction of other physical infrastructure), which can reduce construction costs by as much as 70%;
3. Billing issues (rent) and directing these funds to the broadband development fund;
4. Mobilization of local expert and wider public regarding broadband access;
5. Education of (future) users and their preparation for the use of broadband services;
6. Organization of service distribution;
7. Development of new applications and sensors;
8. Regional control of broadband functioning in accordance with rules on consumer protection; and
9. Developmental effects of broadband access implementation at local level.

The untreated mentioned elements (from 1 to 9) in BIDPs indicate that whole process of broadband access will be based on unprepared users, which will result in a long period of time from construction of the broadband network to its full use. Likewise, opportunities for development will not be used, i.e., not will be to take advantage of local human, material, and financial potentials from broadband construction and that series of procedures will be left to the mercy (and misery) of market approach that brings profitable benefits only to telecom service providers and more lost benefits to the unprepared local community as it has been so far.

5. Business issues of broadband development

Failure and unsatisfactory result of a well-conceived, elaborated, and reported project to the Ministry of Regional Development of Croatia, which was approved, but poorly conducted and unfinished, require interdisciplinary analysis. The analysis should start with the unfavorable position of Croatia in terms of the density and speed of broadband access compared to EU average, as well as to the very unfavorable state of broadband in Slavonian counties compared to the average of Croatia. These facts provide a framework and direction for acting to any professional expert or serious politician and other public persons. It should also be recalled that the development of broadband services is also an opportunity for the economic and technological development of local companies and employment of the local population to restrain the wave of emigration from Slavonia.

When Croatia started implementing the Digital Agenda of EU, the region of Slavonia was the first in the following [23, 34, 35, 37, 38]:

a. The first conference on broadband access in Croatia was held in Osijek (2012).

b. The first broadband access development project in Croatia for the area of more than one local community was launched in Osijek (2012).

c. The "Slavonian network" project was the best-ranked project for the development of broadband network by the Ministry of Regional Development of Croatia (2013) and was approved after the three expert analysis from expert teams of the Ministry (2014).

d. A stage for the mobilization and education of leaders and expert services for local units was planned in the "Slavonian network" project and a series of real measures to finance preparatory stages in the development of broadband network in the region (2014).

e. The interdisciplinary expert team (IET), which has conceived and developed the "Slavonian network" project, has a large number of expert references in national and international frameworks (2012–2017).

Pointing to the facts that:

1. FEE abandoned "Slavonian network" project and submitted it to the University of Osijek.
2. The University has appointed a new team without expert references for the implementation of the project.
3. The new project team showed an inadequate work, and finding only one culprit for the failure of the project does not make a great discovery. It does not contribute to finding a complete response to this problem and does not contribute to finding the direction for the necessary social action.

Namely, during the process of launching the "Slavonian network" project and during its development, five counties as well as dozens of cities and municipalities representatives were informed about the project by prefectures and expert services. Public debates and adoption of series of broadband documents in Croatia and articles in the media and TV provided enough information on the importance of broadband for development of local units. However, the municipalities' and cities' leaders in Croatia do not have technically educated experts in their work environment that would emphasize to them the importance of development projects. Other experts with social orientation don't understand such specialized professional materials or are not using opportunities for further education (lifelong learning). Thus, 4 years were lost for the development of broadband network in the area of eastern Croatia. These 4 years can cause further extensions and bring even more damage due to failure to implement the already adopted project. The towns and municipalities in the Slavonia areas were unprepared and started late in the process of making BIDPs. From ignorance (or perhaps due to conformism) of local representatives, telecom companies have been given the option of investing in broadband from EU funds. This financial support can only be obtained by local units. In this way, these local units have reduced their development chances. Constructing broadband and rental and managing the local broadband network as well as its maintenance were great opportunities to raise technical level and economic development. It was a chance for awakened hope for the youth and to be one kind of horizon of better times, which would deter many young specialists from leaving Croatia and migrate to Ireland, Germany, Austria, Sweden, Norway, and other EU countries.

5.1. Structure of business problems

Immediately after taking the project from the University and forming a new team for the project "Slavonian network," the IET considered problems of further development of this project and published (2014) a paper, which hypothesizes the structure of business problems in implementation of infrastructure projects in Croatia, as outlined in **Figure 14**.

Technical issues evaluate for 10% in the structure of business problems in project implementation of launching the broadband in Slavonia. Legal problems in unfair real property relationships, regarding lands through which corridors of broadband are passed, are estimated to account for 25% of total business issue. Unsuitable financial resources and other economic issues were assessed as 30% of effort representation. However, the largest share of the difficulties in implementing the Slavonian network project, as well as in other infrastructure projects, are sociological (social) problems—which reach 35% of total disadvantage. The issue of social relations is also a crucial brake for the successful implementation of the "Slavonian network" project. Later analyzes and published papers assert this hypothesis [37, 38].

The reasons for this great and decisive influence of sociological problems in the realization of this project should be sought in the underdeveloped society, underdeveloped social relations, i.e., in unfinished transition processes toward civil society with the market economy. The same happens with economic initiatives as well as many infrastructure projects in Croatia. This situation is similar in other transition countries.

Figure 14. Structure of business problems in implementing the "Slavonian network" [35].

5.2. Economic interests and processes of regulation

Companies in the market economy are driven by economic interest—profit—and some of the authors here also include their derivatives such as social power or social reputation. In the realization of economic interests, one should distinguish individual, partnership, shareholder, and state interest. Business problems are solved in the market economy, and the results are obtained expost, only after the realization of the market. Therefore, for company business and project implementation in developed market economies, crucial knowledge about processes (technological, market, legal, etc.) and important stakeholders who have experts with references and results in practice are a must.

There are large, medium, and small businesses in each specific market of goods and services. In transition countries, in the social relationships in economic arena, as this part of society calls Lintz and Stepan [39], do not allow economic efficiency as in developed industrial countries. Namely, an educated civil society accepts known economic (and other) interests and roles of these large, medium, and small players; written and unregistered rules of conduct are known; and the participants adhere to them. This will mean that the market is regulated and close to market equilibrium, and within the framework of the national economy, the best results are achieved. In such an economic arena, big players will not endanger medium and small entrepreneurs; they will even cooperate with them, or small business owners will join forces in technology or marketing because of the cost reduction and because they are all aware of the division of the role and the need for all efficient people to participate in the market.

In the economic arena of transition economies, many interests are not transparent, written and unregistered rules are insufficiently respected, and the roles are hazy. In transitional countries, few of the participants in infrastructure projects take care of the whole spectrum of interests; personal and subjective interest is in the first plan solely. All this indicates that the market is not transparent, clear, and far from equilibrium; it can be said to be in a kind of chaos. Croatia, among other unsuccessful transitional countries, is a good illustration of our demands.

The desires for great profit in private companies have no boundaries, and therefore in the regulated countries, there exist regulatory mechanisms that do not allow a monopoly position in the market. This is particularly true of the telecomm sector. In transition countries, these regulatory mechanisms are only formal so that monopoly companies, lobbing in different even in nonpermissible ways in professional circles and government bodies, are able to skip the boundaries in earning profits for present and future times. Such are broadband services in Croatia, with the slowest speed in EU, being the most expensive between EU members at the same time. It should be noted first that employee salaries in Croatia are several times lower than the EU average, and second is that the unemployment rate in Croatia is among the largest in EU [37, 38, 40].

5.2.1. Conflicts of interest and systems of social regulation

People are less dependent on nature with the advancement of civilization and more and more dependent on society, on social relationships (and their technical systems). Written norms, laws and regulations, cannot cover or regulate everything, so that in regulated countries, social norms and social relationships are of increasing importance. That is why processes of transition to civil society are important [41, 42]. Unfortunately, although Croatia has become a member of EU, the transition processes have not yet been completed.

Civil society is pluralistic and there are legitimate interests of all individuals and groups. These interests are almost always in opposition and everyone is fighting for their own interest. From this, there are a number of conflicting relationships, e.g., worker—entrepreneur, officer—manager, rich—poor, old—young, employed—unemployed, modernist—traditionalist, etc. Many and different instruments of social regulation exist: legal regimes, peace-making, and reconciliation methods to a series of rules in social relations. One of these methods is the motivation to engage in affairs as a way of resolving conflicts of interest, which is not well known in the public and even in professional circles.

Economist Nobel Prize winner John Kenneth Galbraith wrote in his book "The New Industrial State" (1960s) about the business organization and (indirect) resolution of conflicts of interest. Describing a motivation system whose elements have been developed in human history, Galbraith distinguishes four motivational models for work engagement: (a) physical coercion, (b) financial reward, (c) identification, and (d) adaptation. Galbraith emphasizes that in modern terms physical compulsion in a business organization no longer exists and that monetary compensation as an individual's interest can be balanced, if employees are motivated by identification with a common goal, the benefit to the community, or motivation through the adaptation of the purpose function to use the business effects regarding their own affirmation and proving this in the environment (reputation in society) [43].

In this sense, the project "Slavonian network" is conceived not only as an infrastructure but also as a sociological project. Regardless of the regional name, the project has the ambition to give an example of community problem solving, and broadband access is certainly a must. The speed of implementation will also depend on the economic impact for all project stakeholders; these effects will be dispersed throughout the community, including business, education, service, culture, health, management, and any other activity, and the realization of the inclusion idea of every citizen in the e-citizen society.

5.2.2. Knowledge society

While the postsocialist countries are not dealing with rebuilding their unstructured society although some are successful and move on to development and some return to patriarchal and clerical ideas of the eighteenth century, the developed civil societies are transformed into a knowledge society and build a business based on knowledge.

The paradigm of contemporary civilization is a knowledge society that has four main pillars: knowledge economy, information society, networked society, and lifelong learning, **Figure 15**. The basic characteristics of the knowledge society can be described by analogy with the learning organization model: a continuous process in which society by producing and capitalizing on new knowledge is rapidly evolving in constant adaptation to new, increasingly brisk challenges in the environment. In doing so, the way of life, work, and acquiring knowledge quickly changes. In these processes, individuals are constantly learning and adapting rapidly to changes, but also the attitude of institutions to the environment changes and attempts to manage events (designing the future). At the same time, new structures and forms of production, transfer, and application of knowledge are emerging, including a large number of participants with a corresponding increase in the internationalized network-initiated context.

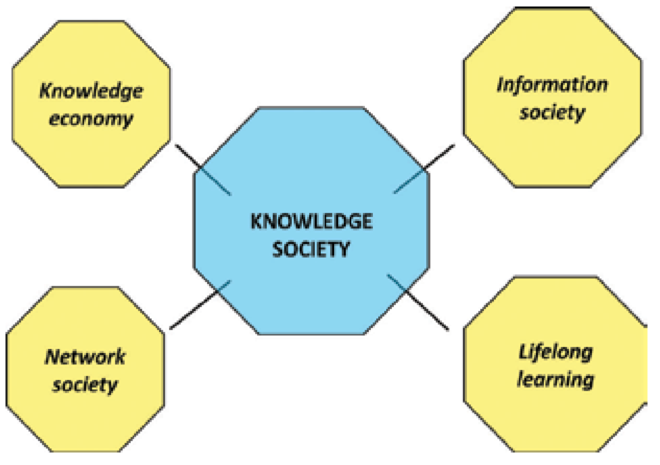

Figure 15. Main pillars of knowledge society [2].

Science and technology (in knowledge society) are in a continuous and reciprocal relationship, i.e., theory and practice are in dynamic interdependence that is rapidly anticipated in technological and organizational innovations; knowledge is capitalized pragmatically, and people are trying to learn and act more effectively in the team [1, 2].

Lifelong learning implies organized active education of people throughout the life span with a view to improving knowledge, skills, and abilities within a personal career and a civic, social, and/or business perspective. This learning implies acquisition and modernization of all kinds of knowledge and qualifications, from preschool until the end of the working age and even after retirement. In this concept, the development of knowledge and abilities allows people to adapt to life in a society of knowledge and active participation in all areas of social and economic life. The lifelong learning system is a basic prerequisite for the growth and development of society and the economy of knowledge.

Broadband infrastructure is a prerequisite for continuous learning. Some countries have already recognized this infrastructure that has become an integral part of formal education. We expect this to happen in the foreseeable future in Croatia as well. Some of the regions in Croatia in part of year have cut off from formal education, and through the realization of broadband projects, the inclusion of human resources will be enabled throughout the year and throughout the country.

5.3. Social problems in the implementation of infrastructure projects

Launching infrastructure projects—with a large number of participants, their various interests, with a small (but continuous) profit for the investor, with a great impact on the functioning of the community and good influence on GDP—is often subject to unpredictable influences on the mutual trust of project participants, i.e., social cohesion. Important elements of social cohesion are:

a. Trust in people and social institutions; this manifests itself in willingness to cooperate not only with family members or acquaintances.

b. Respect for social norms and legal regulations is shown by a smaller scope of political manipulation and less corruption.

c. Association and collective social activity: this is manifested by the experience of cooperation in the pursuit of interests outside the scope of an individual's ability.

These three elements are interrelated and affect one another, **Figure 16** [41, 42].

Contribution of a large number of sociologists points to the complexity of gaining mutual trust and the mentioned problem. The results of the research of the trust of individuals in the groups of people and institutions and of the hierarchy of trust in Croatia are decisively deviating from the results of similar research in well-organized EU societies. Realization of social cohesion in society is a process that is measured with decades [39, 41, 42, 44–47]. We point to these processes—because we recognized another problem the "Slavonian network" is facing—either in the light of project implementation priorities or in ensuring the size of the budget in the area of a local self-government unit. Disruption of social cohesion in this way could be reflected in the mistrust between political elites and the mistrust of citizens toward their political representatives. This could undermine the ongoing implementation of the project, and the continuity of project implementation is one of the conditions for the use of grants from EU funds.

If there is no social cohesion, then society is not using objective criteria for creating teams in project and business implementation. Instead of this, the criteria of loyalty to a political party, national, religious, or racial criteria are used. This is often in use by the mafias that are masked by patriotism and the interests of the nation or religion as well. Thus, in the case of the implementation of the "Slavonian network" project, many of the criteria listed in the selection of the new project implementation team, but only when the project was drafted and approved, were used only on the basis of criterion of expertise and proven references.

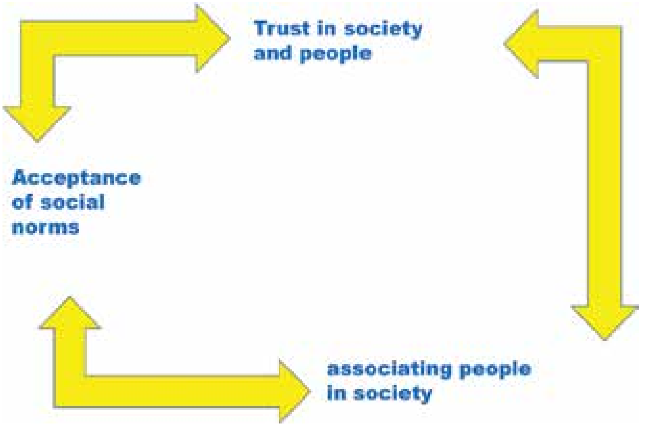

Figure 16. Complex structure of social cohesion [42].

http://dx.doi.org/10.5772/intechopen.72037

Figure 17. Map of nöGIG broadband project in the Lower Austria region [48].

As an illustration of our ratings and as an example of the good practice, we have listed the Austrian example of the introduction of broadband in the Lower Austria region, nöGIG GmbH project, and later company, was winner of the 2016 European Broadband Awards competition in category no 4, openness and competition (**Figure 17**). "This 3 layer open project has public entities in the driving seat and is set on establishing an open, public infrastructure in underserved rural areas. nöGIG GmbH, a public authority coordinates planning and construction. The Broadband Coordination of Lower Austria (BBK) is in charge of the broadband strategy and allocation of public financial resources. Municipalities, private companies and service providers have cooperated to make this a successful open and public model." [48].

This project was started later than the Croatian project "Slavonian network." But it is already in high level of realization because local governments have recognized the initiative and supported it, and citizens have adopted the model and the project is being realized.

6. Model of broadband development for transition countries

During its efforts to realize the EU Digital Agenda and national broadband strategy, Croatia faced significant problems that make it difficult and in some cases delay the implementation of this project. Practice has shown that Croatia is significantly lagging behind in the implementation of the Digital Agenda, and the example of Slavonia region shows a number of important reasons. Since the beginning of the broadband application in Croatia, some telecomm companies have a strong influence on many local units and impose only issues of the construction of a network and an investment model, which is favorable only for their profit.

Since the beginning of the concept of broadband application in Slavonia, IET members have started to create a project with a spectrum of related themes, not just technological and construction goals. The project was conceived by several modules, the project modules were analyzed, the possibilities of their efficient realization were explored, and all the findings and solutions were published in professional circles and the general public. After a large number of practical examples and inductive conclusions regarding the broadband implementation in the Slavonia region (and Croatia), it is time for the synthesis. The basic conclusion at the paradigm level reads:

- Citizens (residents, their public services, and private companies) do not exist for the favor of telecomm companies and broadband, but the opposite is true: telecomm companies and broadband exist for the needs of citizens and should be at their service.

Therefore, broadband activities should involve citizens, i.e., their representatives and institutions. Of course, we are thinking of educated citizenship (their representatives), not of formalized involvement. Namely, according to the existing model, very unfavorable and dangerous distances in the opportunities for personal development of people and the development of local communities in Croatia are transferred from the past and the present to the future. These are (a) nontransparent business interests and aims of prominent individuals, interest groups, and individual companies; (b) an unhealthy imbalance in social relations that is at the limit of exclusion of a large number of people from all developmental processes; and (c) inequality of opportunity for access to modern communications. All this is transmitted to the future, but concealed with the goal of high communication technology.

In their works, Lintz and Stepan [39] pointed out at the beginning of postsocialist transition processes the possibility that the processes of democratization of society and the creation of an open market economy in some countries need not be completed. These countries will not build civil society, i.e., their societies will be in a kind of chaos with unhealthy imbalances. The economic situation in Croatia, as well as in a number of transition countries, and the very pronounced emigration of young people from these countries illustrate sufficiently anticipated processes.

The introduction of broadband is an opportunity for democratization of social relations, strengthening of transparency in the policy of technological and economic development. Broadband is through the process of preparatory stages an opportunity to simultaneously make a step in building social cohesion, steps toward a civic society, and a step toward the knowledge society. In these processes, it is very important to educate citizens and implement the concept of lifelong learning. And really broadband is one of the essential tools for that. This is not the question of what is older—chicken or egg (whether education first and broadband, or vice versa) but the matter of opportunity for simultaneous activities to make a leap into a higher quantum trait of social relations in the path of modern civilization.

It has already been said that practice has shown that local units have no skilled staff for broadband applications. This is true, but in the their universe only. And their universe is on the lower quantum trajectory, in the social relationships of the nineteenth century. In modern civil society, developed democracies with developed economies for decades have used think tanks in many areas of human life and work. Think tanks exist both at the international level and at the national and local levels [49]. However, the number and impact of think tanks on development policies in transition countries are minimal, unlike developed societies and states. Democratic and developed societies use this model of nonpartisan, nonprofit, and independent teamwork by scientists and top experts [50].

Panon Institute from Osijek has manifested itself in the area of industrial strategy, broadband strategy, and renewable energy sources. Unfortunately, it is not recognized in the local communities of the Slavonia region.

7. Closing remarks

There is no single detailed recipe for planning and implementing investments in infrastructure installations, especially in the fast-growing technology sector. However, the basic professional framework must be taken into account, whereby the essential elements for the development and dynamics of the process will be determined according to local circumstances and the existing references.

Here are the experiences of the 5-year work of an expert interdisciplinary team that has so far successfully realized lot of national and regional projects of millions of values. Although the project has not been implemented (to the fullest extent) as conceived, and as approved by the Ministry, the experiences mentioned above are still precious. The established original development models, the simultaneous implementation of the module, the interdisciplinary problem-solving strategiesm and the transparency in the work that are outlined here can be used well in the development of broadband in other countries. And to young professionals, this can be good for a professional reading.

Acknowledgements

I would like to thank Professor Franjo Jović for his expert advice and encouragement throughout this difficult project.

Author details

Milan Ivanović

Address all correspondence to: panon.institut@gmail.com

PANON – Think Tank for Strategic Studies, Osijek, Croatia

References

[1] Ivanović M. Three Essays on Science. 2nd revised edition. Faculty of Electrical Engineering Osijek, Croatian, Tri eseja o znanosti. 2008. ISBN: 978-953-6032-55-6226

[2] Ivanović M, Keser T, Blažević D. A capitalization of knowledge—Innovation processes in transition countries. Technical Gazette. 2010;**18**(1):15-22. ISSN 1330-3651

[3] Ivanović M, Širić M. The Knowledge Economy and Development of New Technologies for Knowledge Society. XXVII Science in Practice; Osijek; 5 May 2008. Faculty of Electrical Engineering; pp. 31-36. ISBN: 978-953-6032-62-4

[4] EC. A Digital Agenda for Europe COM (2010) 245. Brussels; May 19, 2010 http://ec.europa.eu/information_society/digital-agenda/ [Accessed: 2017-7-7]

[5] Abdelfatteh H, Sanae EH, Abdelhak A, Asmaa El H. In: Ivan NDS, editor. The Role of Communication Technologies in Building Future Smart Cities, Smart Cities Technologies. London, UK: InTech; 2016. DOI: 10.5772/64732

[6] Mesaric D, Ambroš F, Ivanović M. Broadband infrastructure in the region of Slavonia and Baranja (Croatia). IJECES – International Journal of Electrical and Computer Engineering Systems. 2014;**4**(2):21-28. ISSN 1847-6996

[7] Ambroš F, Ivanović M, Mesarić D. Construction of municipal infrastructure and development of optical networks in Slavonia. In: Proceedings of the GAS 2013 Conference; Osijek; October 26-28, 2013. pp. 1-12. ISSN 1849-0638

[8] Fornefeld M, Delaunay G, Elixmann D. The Impact of Broadband on Growth and Productivity. Düsseldorf, Germany: MICUS Consulting GmbH. 2008. https://www.breitbandinitiative.de/wp/wp…/2008_micus-studie-broadbandeu_long.pdf [Accessed: 2013-3-13]

[9] Analysis Mason, Guide to Broadband Investment, Final Report. http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=7630 [Accessed: 2013-3-13]

[10] Ivanović M, Ambroš F. The economic impacts of broadband access investment in Croatia. In: Proceedings of the 5th International Conference on Economy of Eastern Croatia; Osijek; May 23-25, 2016. pp. 54-64. ISSN 1848-959

[11] Frontier Economics. The Impact of Broadband in Eastern and Southeast Europe. 2010. https://www.cdn1.telekomaustria.com/.../broadband-study.pdf [Accessed: 2013-3-13]

[12] EC. Public Consultation—Revision of the Guidelines on Public Funding to Broadband Networks. http://ec.europa.eu/competition/consultations/2012_broadband_guidelines/ [Accessed: 2012-12-12]

[13] EC. Digital Agenda Scoreboard—Electronic Communications Market Indicators. http://ec.europa.eu/information_society/digital-agenda/scoreboard/2012.pdf [Accessed: 2016-6-16]

[14] Medved J, Tumir P, Ivanović M. Važnost amortizacije infrastrukturnih instalacija. In: Proceedings of the 23rd Znanstveni skup OTO ′2014; Požega. April 24-25. pp. 183-192 ISBN 978-953-6032-84-6

[15] Lator, d.o.o. Studija o odabiru najpovoljnijih modela financiranja i poticajnih mjera za ulaganja u infrastrukturu širokopojasnog pristupa, (www.mppi.hr/) [Accessed: 2016-6-16]

[16] Government of the Republic of Croatia. Okvirni nacionalni program za razvoj infrastrukture široko-pojasnog pristup u područjima u kojima ne postoji dostatan komercijalni interes za ulaganja, NN 68/2016

[17] Analysis Mason. Guide to broadband investment. Final Report. http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=7630 [Accessed: 2016-6-16]

[18] Ivanović M, Ambroš F, Mesarić D. Models of investment in fiber optic networks and implementation of Slavonia net project. Proceedings of the 3rd Conference Economy of Easter Croatia; Osijek; May 23-24, 2014. pp. 15-24. ISSN 1848-959

[19] Ambroš F et al. Razvoj i održavanje optičke mreže nove generacije na području Slavonije. OTO 2013; Osijek; 25-26.4.2013. pp. 77-86. ISBN 978-953-6032-79-2

[20] Ambroš F, Ivanović M. Izgradnja komunalne infrastrukture i razvoj optičke mreže na području Slavonije. Plin 2013; Osijek; 27-30.9.2013. pp. 134-142. ISSN 1849-0638

[21] WIK-Consult. The Economics of Next Generation Access [Accessed: 2012-12-12] http://wik.org/uploads/media/ECTA_NGA_masterfile_2008_09_15_V1.pdf

[22] EC. EU DIRECTIVE 2014/61/EU on measures to reduce the cost of deploying high-speed electronic communications networks

[23] Ivanović M, Ambroš F, Mesarić D. Osnivanje službi za upravljanje lokalnim optičkim mrežama i njihovo održavanje na području Slavonije. 26. Proceedings OTO ′2017; Osijek; 26.5.2017. pp. 141-150. ISBN 978-953-6032-92-1

[24] International Telecommunication Union (ITU). http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf [Accessed: 2017-9-19]

[25] Broadband Commission UN. State of Broadband: Catalyzing Sustainable Development www.broadbandcommission.org/Documents/reports/bb-annualreport2016.pdf

[26] http://www.internetworldstats.com/stats.htm [Accessed: 2017-10-10]

[27] EC. Europe 2020: A European strategy for smart, sustainable, and inclusive growth. COM (2010) 2020; Bruxelles; 3.3.2010

[28] Frontier Economics. The impact of broadband in Eastern and SE Europe. 2010. https://www.cdn1.telekomaustria.com/.../broadband-study.pdf [Accessed: 2013-3-13]

[29] IHS Markit Ltd. Broadband Coverage in Europe 2016—Mapping Progress Towards the Coverage Objectives of the Digital Agenda. London: IHS Markit Ltd. http://ihs.technology.com. Point Topic. [Accessed: 2017-7-17]

[30] Eurostat. http://ec.europa.eu/eurostat/data/database [Accessed: 2017-10-10]

[31] Analysis Mason Ltd. Digital-Progress-Report-2017. 2017. https://ec.europa.eu/digital-single-market/en/news/europes-digital-progress-report-2017 [Accessed: 2017-10-10]

[32] Government of the Republic of Croatia. Strategy for Broadband Development in the Republic of Croatia for 2012-2015.,www.mppi.hr/ [Accessed: 2013-3-13]

[33] Government of the Republic of Croatia. Strategy for Broadband Development in the Republic of Croatia for 2016-2020, www.mppi.hr/ [Accessed: 2017-10-10]

[34] Mesarić D, Ambroš F, Ivanović M. Slavonska mreža, razvoj širokopojasne mreže regije Slavonija. Idejni projekt. Zagreb: Ministarstvo regionalnog razvoja; 15.3.2013

[35] Ivanović M, Ambroš L, Mesaric V. Establishing a consortium-way for successful implementation of investment projects. In: Proceedings of the 3rd International Conference Economy of Eastern Croatia; Osijek; 23 May 2014. pp. 34-45

[36] Osijek County. Osnivanje trgovačkog društva "Slavonska mreža". http://www.obz.hr/hr/pdf/2015/14_sjednica/05_Prijedlog_za_osnivanje_trgovackog_drustva_slavonska_mreza_doo.pdf [Accessed: 2017-5-5]

[37] Ivanović M. Ekonomski interesi i društveni problemi realizacije velikih infrastrukturnih projekata. In: Elektroničke komunikacijske mreže velikih brzina. Informator, Zagreb; 2017. pp. 51-71. ISBN 978-953-819-15-16

[38] Stojnović V, Ivanović M. Optičke mreže u Baranji, 6. simpozij „Kopački rit—jučer, danas, sutra" 2017. Proceeding of Abstracts; Tikveš; 28-29.9.2017. p. 243

[39] Linz J, Stepan A. Problems of Democratic Transition and Consolidation. Baltimore, Maryland: Johns Hopkins University Press; 1996. ISBN 08018-5157-0

[40] EC. Broadband Internet Access Cost (BIAC) Autumn 2015—Prices as of September–October 2015, Brussels, Belgium. 2016. https://ec.europa.eu/digital-single-market/en/news/broadband-internet-access-cost-biac-study [Accessed: 2017-7-17]

[41] Ivanović M. Izgubljene šanse—procesi postkomunističke tranzicije. Osijek: AlbertE; 2005 ISBN 953-6980-02-9

[42] Ivanović M, Lauc A, Bognar L, Kovač J, Pecnik J. Kamo idemo—društvene vrijednosti u procesu tranzicije u Hrvatskoj. Albert^{E}, Osijek, 2006. ISBN: 953-6980-01

[43] Galbraith JK. The New Industrial State. Boston, USA: Houghton Mifflin Company. 1967. ISBN 13; 978-0-691-13141-2

[44] Fukuyama F. Trust—The Social Virtues and the Creation of Prosperity. New York: The Free Press; 1995. ISBN: 0-02-910976-0

[45] Kulić S. Rekonstrukcija društva kao preduvjet restrukturiranja gospodarstva i uključivanja u europske integracije. Zagreb: Ekonomski institut Zagreb; 1999

[46] Rimac I, Štulhofer A. Sociokulturne vrijednosti, gospodarska razvijenost politička stabilnost kao čimbenici povjerenja u Europsku Uniju. Zagreb: Institut "Ivo Pilar"; 2002

[47] Županov J. O toleranciji. u "Tolerancijom protiv mržnje". Zagreb: HHO; 1995

[48] nöGIG-Fibre infrastructure of Lower Austria https://ec.europa.eu/digital-single-market/en/news/winners-european-broadband-awards-2016 [Accessed: 2017-8-8]

[49] Ivanović M. Think Tank—An Analytical Centre for Public Policy Thinking and Advocacy. http://www.panon.eu/tekst/hr/42/6%20STO%20JE%20THINK.pdf

[50] Ivanović M. What is Think Tank—The Core Functions of the Analytical Centre, PPP. https://docs.google.com/viewerng/viewer?url=www.panon.eu/tekst/hr/55/FZ%201%20Ivanovic.pdf

Chapter 19

Techniques for Reducing Redundant Unicast Traffic in HSR Networks

Jong Myung Rhee and Nguyen Xuan Tien

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71783

Abstract

High-availability seamless redundancy (HSR) is a seamless redundancy protocol for Ethernet networks. HSR provides seamless communication with fault tolerance based on the duplication of every unicast frame sent in a ring topology. HSR is very useful for mission- and time-critical systems such as substation automation systems (SASs). However, the main drawback of HSR is to generate excessively redundant network traffic in HSR networks. This drawback would unnecessarily waste network bandwidth and hence could degrade network performance in HSR networks. Several traffic reduction techniques for HSR networks have been proposed to improve the network performance in the networks. These techniques can be classified into two main groups: traffic filtering-based and dual paths-based techniques. In this chapter, we provide a description and comparison of these HSR traffic reduction techniques. This chapter describes these traffic reduction techniques and compares their network performance. The operations, advantages, and disadvantages of these techniques are investigated and summarized.

Keywords: high-availability seamless redundancy (HSR) protocol, traffic reduction techniques, fault-tolerant Ethernet networks, seamless communications, smart grids

1. Introduction

Seamless communication with fault tolerance is one of the key requirements for Ethernet-based, mission critical, and real-time applications such as substation automation systems (SASs), automation control networks, and other industrial Ethernet networks. The Ethernet standardized by the IEEE in IEEE 802.3 [1] does not support fault-tolerant capability. Various protocols have been developed and standardized to provide high availability and fault tolerance for Ethernet networks, such as rapid spanning tree protocol (RSTP) [2], media redundancy protocol (MRP) [3], shortest path bridging (SPB) [4], redundancy protocol for Ethernet (RPE) [5], time-sensitive network (TSN) [6, 7], parallel redundancy protocol (PRP) [8], and high-availability seamless redundancy (HSR) [8]. RSTP can be applied in arbitrary mesh topologies, whereas MRP is restricted to ring topology. Both these protocols provide redundancy in networks and suffer a switchover time delay [9, 10]. SPB, specified in the IEEE 802.1aq standard [4], is a computer networking technology intended to simplify the creation and configuration of networks while enabling multipath routing. It is the replacement for the RSTP. RPE is a redundancy protocol for Ethernet networks that not only provides seamless communications with zero switchover time in case of failure but also supports any topologies [5]. TSN is a set of IEEE 802 Ethernet substandards that are defined by the IEEE TSN task group (IEEE 802.1) [6]. These standards enable deterministic real-time communication over Ethernet. TSN is the second generation of audio and video bridging (AVB) standards [11]. It achieves determinism over Ethernet by using time synchronization and a schedule, which is shared between network components. It offers a way to send time-critical traffic over a standard Ethernet infrastructure. PRP and HSR provide seamless redundancy with zero recovery time. Both the HSR and the PRP are based on the principle of providing duplicated frames for separate physical paths with zero recovery time [8, 12]. PRP and HSR are seamless redundancy protocols that provide seamless communication with fault tolerance for Ethernet-based applications. However, unlike the PRP that requires dual redundant independent networks, the HSR can be applied to a single network and still retain its property of zero recovery time. HSR is one of the redundancy protocols selected for substation automation in the IEC 61850 standard [13].

1.1. HSR overview

HSR, a redundancy protocol for Ethernet-based networks, was standardized in IEC 62439-3 [8]. It provides seamless redundancy with fault tolerance for Ethernet networks by duplicating all frames sent in a ring topology. In other words, the HSR protocol provides two frame copies for the destination node, one from each side, enabling zero-fault recovery time, in case one of the frame copies is lost. This means that even in the case of a node or link failure, there is no stoppage of network operations. If both sent copies reach the destination, it takes the faster copy and discards the duplicate. This feature of the HSR protocol makes it very useful for time-critical and mission-critical systems, such as SAS.

The HSR protocol defines several types of nodes [8]: the doubly attached node with HSR protocol (DANH), the redundancy box (RedBox), and the quadruple port device (QuadBox). DANHs are HSR terminal nodes that have two ports operated in parallel. RedBoxes are used to connect legacy devices, such as maintenance laptops and printers, to HSR rings. QuadBoxes are used to connect HSR rings. The standard HSR protocol is mainly applied for ring topologies, including single-ring and connected ring topologies.

1.1.1. Single-ring topology

A single-ring HSR network consists of DANHs, each having two ring ports, interconnected by full-duplex links, as shown in the example of **Figure 1**. A source DANH receives a frame

http://dx.doi.org/10.5772/intechopen.71783

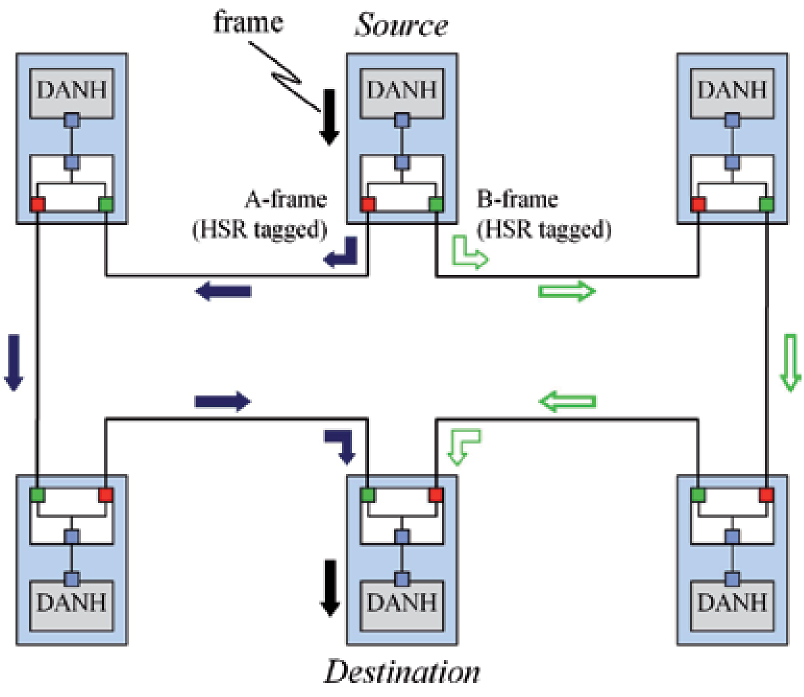

Figure 1. An example of HSR single-ring network.

passed from its upper layers, prefixes the frame by an HSR tag to identify frame duplicates, and sends the frame over each port. When a DANH receives the frame, the DANH forwards the frame to its other port, except if it already sent the same frame in that same direction. A destination node of a unicast frame does not forward the frame for which it is the only destination. In the fault-free case, the destination node receives two identical frames from each port, removes the HSR tag of the first frame before passing it to upper layers, and discards the duplicate.

1.1.2. Connected ring topology

To allow more complex network topologies, HSR rings can be connected through the use of QuadBoxes. A pair of QuadBoxes is used to connect two rings to prevent a single point of failure. A QuadBox forwards frames over each ring and passes the frames to the other ring without changes. **Figure 2** shows an example of a connected ring network that consists of eight DANH rings, each DANH ring includes four DANHs.

We consider a scenario in which DANH 1 in DANH ring 1 sends unicast frames to DANH 10 in DANH ring 3. When the source sends a unicast frame to the destination under the standard HSR protocol, the frame is duplicated and circulated in all rings except the destination DANH ring. **Figure 3** shows the process of forwarding an HSR frame from the source DANH to the destination DANH in the sample HSR network under standard HSR protocol.

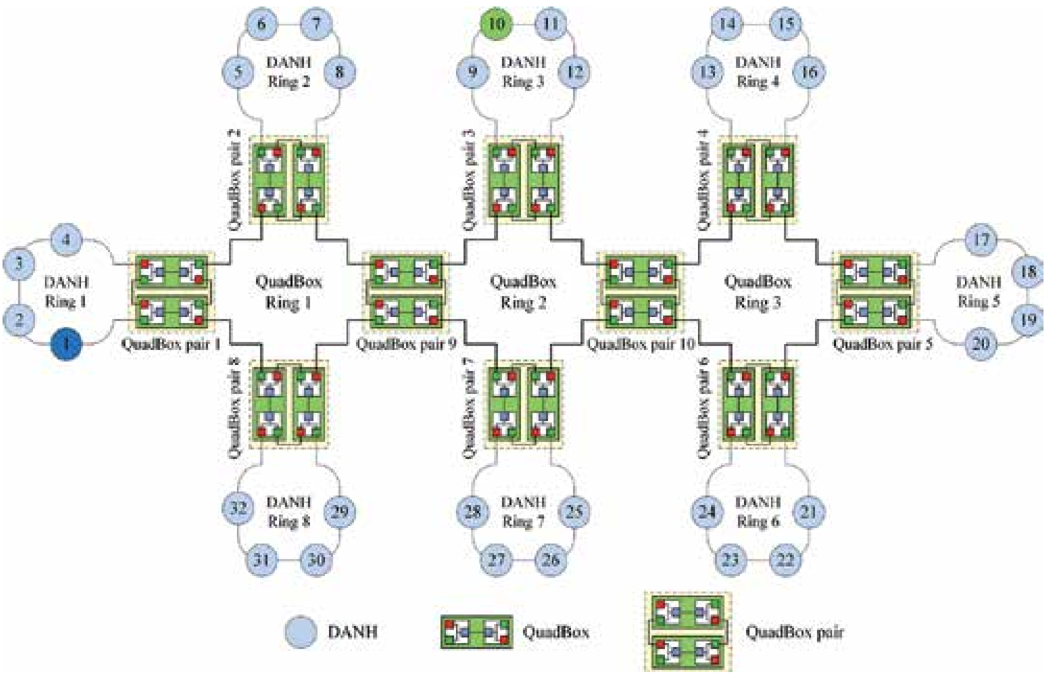

Figure 2. An example of HSR connected ring network with eight DANH rings.

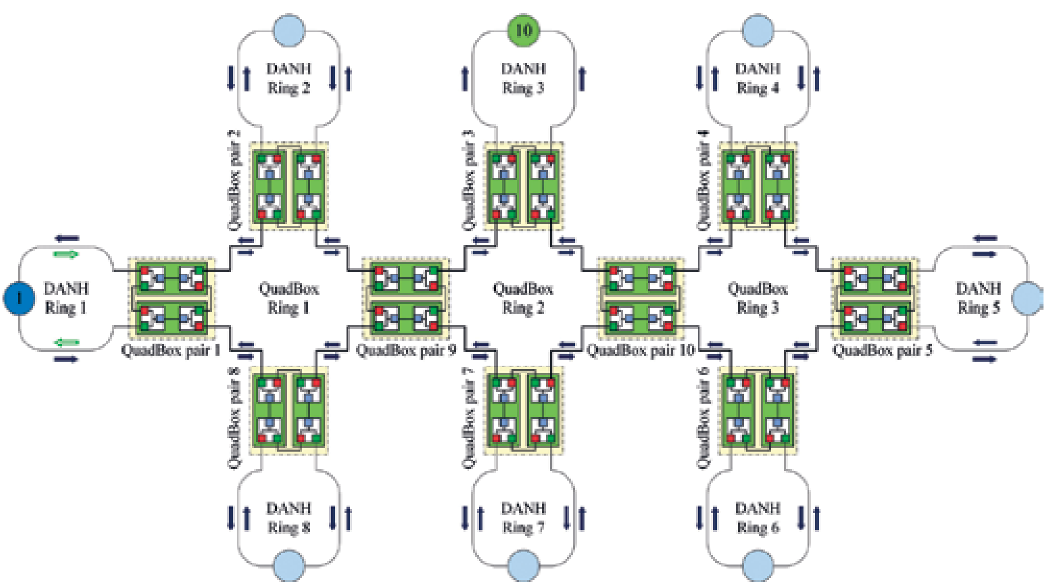

Figure 3. The process of forwarding a unicast frame under standard HSR protocol.

1.2. HSR drawbacks

HSR protocol has no issues with unicast frames inside single-ring networks. However, in connected ring networks, unicast frames are duplicated and circulated in all rings, except

http://dx.doi.org/10.5772/intechopen.71783

the destination DANH ring, as shown in **Figure 3**. The standard HSR protocol thus generates excessively redundant unicast traffic in the networks. This is the main drawback of the standard HSR protocol. The drawback is caused by the following issues:

1. Duplicating and circulating unicast frames in all the rings, except the destination ring;
2. Forwarding unicast frames into all DANH rings;
3. Forwarding unicast frames into all QuadBox rings.

This drawback of HSR degrades the network performance and may cause network congestion and delay in HSR networks. Several techniques have been proposed to reduce the redundant unicast traffic in HSR networks.

2. Traffic reduction techniques

2.1. Traffic filtering-based techniques

Traffic filtering-based techniques reduce redundant unicast traffic by solving some or all of the HSR issues mentioned in Section 1. There are several traffic filtering techniques, including quick removing (QR) [14], port locking (PL) [15], and filtering HSR traffic (FHT) [16].

2.1.1. Quick removing (QR)

QR is the simplest technique for reducing redundant traffic in HSR networks. It reduces the redundant traffic by preventing traffic frames from being duplicated and circulated in rings. It can be used in any topology and applied for any traffic, including unicast, multicast, and broadcast traffic.

2.1.1.1. Operations

The key idea of QR is that each HSR node forwards a unicast frame once, at most. When an HSR node receives an HSR frame, the node checks if the unicast frame has previously been received and forwarded. If not, the node sends the HSR frame over all its ports except the received port. If so, it discards the duplicated frame.

Figure 4 shows the process of forwarding a unicast frame between the source and the destination in the sample HSR network shown in **Figure 2** under the QR technique.

2.1.1.2. Advantages and disadvantages

QR is the simplest technique to reduce redundant traffic in HSR networks and is easy to implement. The QR approach can be applied in any network topology for any kind of traffic, such as unicast, multicast, or broadcast traffic.

However, QR forwards unicast traffic frames into all rings in HSR networks, even rings that are not used to deliver the frames from the source DANH to the destination DANH. In other words, QR does not filter unicast traffic for unused rings in the HSR networks.

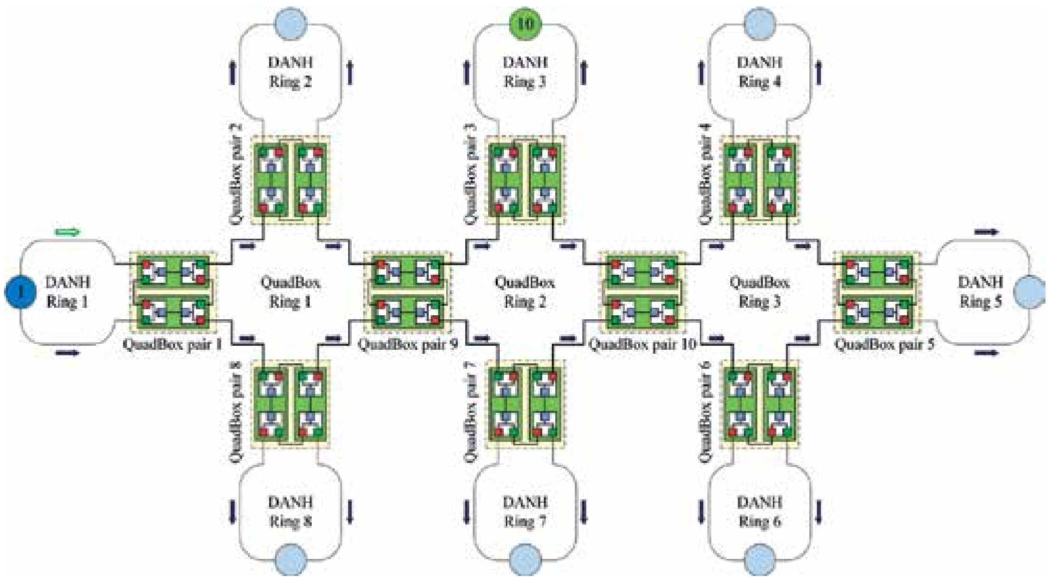

Figure 4. The process of forwarding a unicast frame under the QR technique.

2.1.2. Port locking (PL)

PL reduces redundant unicast traffic in HSR networks by filtering unicast traffic for unused DANH rings. The PL technique does not forward HSR unicast frames into DANH rings that do not contain the destination DANH.

2.1.2.1. Operations

PL is only applied to QuadBoxes that connect to DANH rings called access QuadBoxes. QuadBoxes that do not connect to any DANH ring work as standard QuadBoxes. PL divides an access QuadBox into two sides: DANH side and QuadBox side. The DANH side is connected to a DANH ring, whereas the QuadBox side is connected to a QuadBox ring.

When a source DANH sends a unicast frame to a destination DANH in an HSR network under the standard HSR protocol, the frame is forwarded into all DANH rings of the network. The frame is not circulated in the destination DANH ring that contains the destination DANH because the destination DANH does not forward the frame. Otherwise, the frame is circulated in all nondestination DANH rings that do not contain the destination DANH. The PL technique uses the concept to check and lock nondestination DANH rings. When an access QuadBox sends a unicast frame into its DANH ring over one port and receives the sent frame through the other port connected to the same DANH ring, it means that the destination DANH does not exist in that DANH ring. The access QuadBox then locks its DANH side to prevent it from sending unicast frames into its DANH ring.

The PL consists of two phases: the learning phase and the working phase.

http://dx.doi.org/10.5772/intechopen.71783

a. *Learning phase*: When the source DANH sends the first frame to the destination DANH, copies of the frame are flooded into the entire network, as under the standard HSR protocol. Each access QuadBox will check if its DANH rings contain the destination node. If not, the access QuadBox will lock its DANH site to prevent unicast frames from being forwarded to the DANH ring.

b. *Working phase*: After the learning phase, all DANH sites of access QuadBoxes of which DANH rings do not contain the destination DANH are locked. Unicast frames are not forwarded into DANH rings that do not contain the destination.

Figure 5 shows the process of forwarding a unicast frame between the source and the destination under the PL technique.

2.1.2.2. Advantages and disadvantages

The PL technique filters HSR unicast frames for nondestination DANH rings in HSR networks. In addition, PL does not generate additional control overhead in HSR networks because it does not use any control messages.

The main drawbacks of PL are that it still forwards HSR unicast frames into unused QuadBox rings and does not prevent the frames from being duplicated and circulated in all QuadBox rings.

2.1.3. Filtering HSR traffic (FHT)

FHT is one of the techniques that solves all the issues caused the main drawback of the standard HSR protocol. The FHT filters HSR unicast frames for all unused rings, including

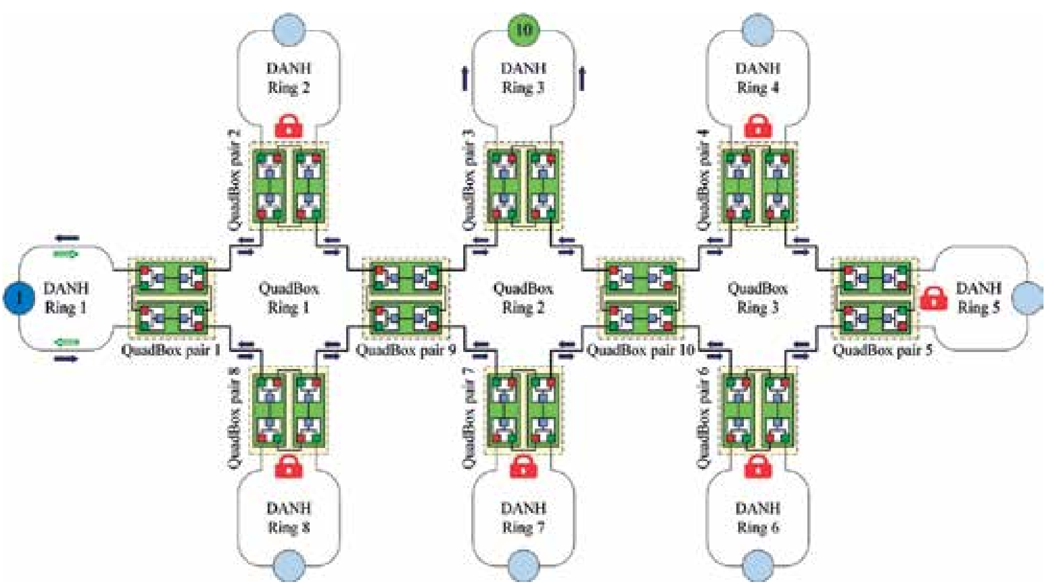

Figure 5. The process of forwarding a unicast frame under the PL technique.

nondestination DANH rings and unused QuadBox rings. In addition, the unicast frames are not circulated in HSR rings under the FHT technique.

2.1.3.1. Operations

FHT learns media access control (MAC) addresses of DANHs and builds MAC tables to filter unicast frames for unused rings in HSR networks. To prevent HSR unicast frames from being forwarded into nondestination DANH rings, each access QuadBox builds and maintains a MAC1 table that contains MAC addresses of DANHs that connect to its DANH ring. Each trunk QuadBox builds a MAC2 table that is a collection of MAC1 tables of access QuadBoxes that connect to its QuadBox rings. The MAC2 table is then used to filter unicast frames for unused QuadBox rings.

Two new traffic filtering rules are defined in order to filter unicast traffic for unused DANH and QuadBox rings:

- Filtering rule 1: An access QuadBox node forwards a unicast frame into its DANH ring, if and only if its MAC1 table contains the destination MAC address of the frame.
- Filtering rule 2: A trunk QuadBox node forwards a unicast frame from its first QuadBox ring to its second QuadBox ring, if and only if the first QuadBox ring's MAC2 table does not contain the destination MAC address.

Additionally, a new traffic forwarding rule is also defined to prevent unicast traffic frames from being duplicated and circulated in rings:

- Forwarding rule: HSR nodes forward an HSR unicast frame once, at most. In other words, a node forwards an HSR unicast frame when the node receives the frame for the first time and discards duplicates.

The FHT approach has two phases: the learning phase and the forwarding phase.

a. Learning phase

In this phase, the FHT learns MAC addresses of DANHs and builds MAC tables. The learning phase consists of two steps: building MAC1 table and building MAC2 table.

Step 1. Building MAC1 table. The table is built at access QuadBoxes. To learn MAC addresses of DANHs, QuadBoxes periodically flood a Hello message. Upon receiving the Hello message, DANHs reply by sending an ACK message back to the sending QuadBox. Each access QuadBox learns MAC addresses of DANHs connecting to its DANH ring and builds its MAC1 table.

Step 2. Building MAC2 table. The table is built at trunk QuadBoxes. Each access QuadBox sends a MAC message that contains its MAC1 table into its QuadBox ring once it has completed building the MAC1 table. Upon receiving MAC messages, trunk QuadBoxes build MAC2 tables.

b. Forwarding phase

FHT uses MAC tables to forward HSR unicast frames in an HSR network. An access QuadBox forwards an HSR unicast frame into its DANH rings if its MAC1 table contains the destination

MAC address of the frame. In other words, the MAC1 table is used to filter unicast frames for nondestination DANH rings. FHT filters unicast frames for unused QuadBox rings based on MAC tables at trunk QuadBoxes. In addition, the new traffic forwarding rule of FHT prevents unicast frames from being duplicated and circulated in rings.

Figure 6 shows the process of forwarding a unicast frame from the source to the destination under the FHT approach.

2.1.3.2. Advantages and disadvantages

The FHT approach filters unicast traffic frames not only for the unused DANH rings but also for the unused QuadBox rings. Additionally, it prevents the unicast frames from being duplicated and circulated in active rings. In other words, the FHT approach solves all HSR issues abovementioned in Section 1. These features make the FHT approach become one of the most efficient HSR traffic reduction techniques.

The main drawback of the FHT approach is that it uses control messages to learn MAC addresses and build MAC tables. This results in additional control overhead in HSR networks.

2.1.4. Simulations and comparisons

As described above, different traffic filtering techniques have different advantages and disadvantages. The QR technique removes duplicated and circulated unicast frames from rings, but it still forwards the frames into all rings, even the nondestination DANH rings and the unused QuadBox rings. The PL technique filters unicast traffic for nondestination DANH rings, but unused QuadBox rings. In addition, it does not prevent unicast frames from being duplicated and circulated unicast traffic in QuadBox rings. The FHT technique filters unicast

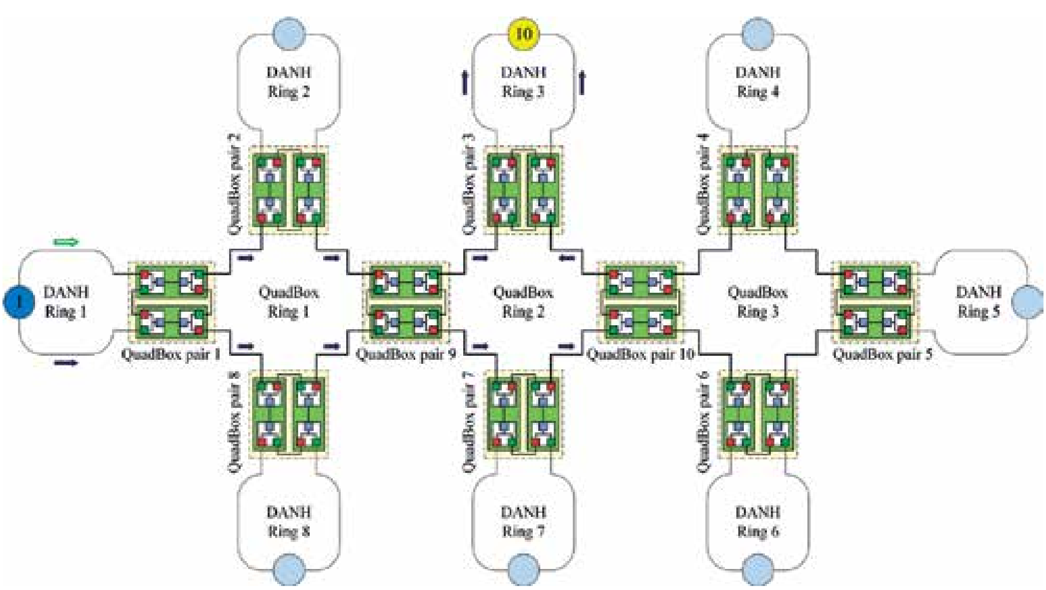

Figure 6. The process of forwarding a unicast frame under the FHT technique.

traffic for nondestination DANH rings and unused QuadBox rings. FHT also prevents unicast frames from being duplicated and circulated in rings. However, the FHT generates an additional control overhead in HSR networks. The characteristics of these filtering techniques are summarized in **Table 1**.

To compare traffic performance of these techniques, several simulations were conducted using the OMNeT++ simulation tool [17]. We consider the sample HSR network consisting of eight DANH rings; each DANH ring has four DANHs, as shown in **Figure 2**. In these simulations, source DANH 1 in DANH ring 1 sends N (N = 10, 20, …, 100) unicast frames to destination DANH 10 in DANH ring 3. **Figure 7** shows the comparison of traffic performance of these traffic filtering-based techniques. The simulation results show that, in the sample network, the QR technique reduces network unicast traffic by 40% compared to the standard

Features	QR	PL	FHT
Preventing unicast traffic from being duplicated and circulated in rings	Yes	No	Yes
Filtering unicast traffic for unused DANH rings	No	Yes	Yes
Filtering unicast traffic for unused QuadBox rings	No	No	Yes
Generating additional control overhead	No	No	Yes

Table 1. Characteristics of traffic filtering-based techniques.

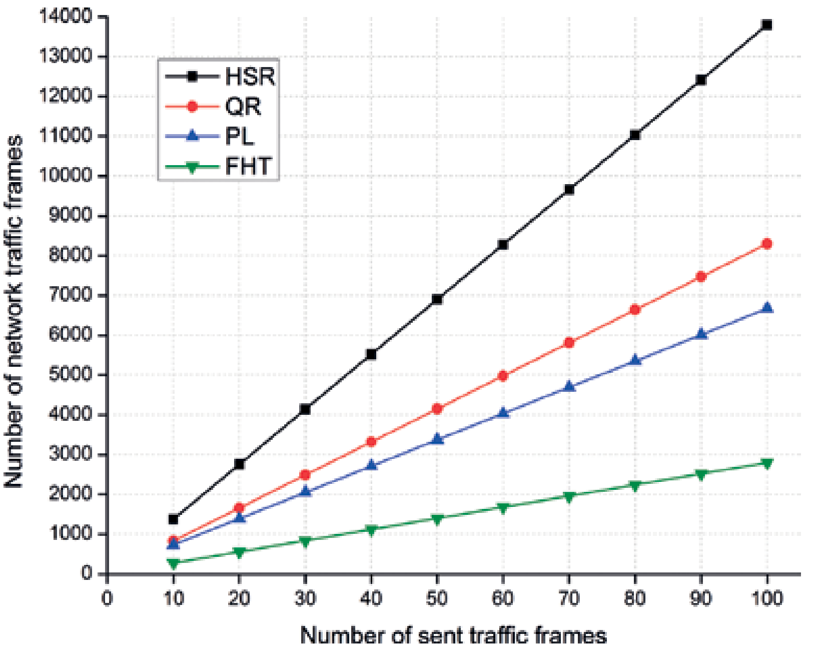

Figure 7. A comparison of the traffic performance of traffic filtering-based techniques.

HSR protocol. The PL technique has better traffic performance than the QR technique; PL reduces network unicast traffic by 51% compared to standard HSR protocol. The FHT technique provides the best traffic performance and reduces network traffic by 80% compared to standard HSR protocol.

2.2. Dual paths-based techniques

Dual paths-based techniques reduce the redundant unicast traffic in an HSR network by forwarding unicast frames from a source to a destination in the network through two predefined separate paths instead of duplicating and flooding the frames into all rings as in the standard HSR protocol. There are several dual paths-based techniques, including dual virtual paths (DVP) [18], ring-based dual paths (RDP) [19], and dual separate paths (DSP) [20].

2.2.1. Dual virtual paths (DVP)

DVP discovers and establishes two separate paths for each connection pair of DANHs in HSR networks. These two paths are DANH-based dual paths. The DVP technique discovers the DANH-based dual paths by sending and receiving control messages, including path selection (PaS) and path confirmation (PaC) messages.

2.2.1.1. Operations

The operation of DVP includes the following phases:

a. Announcement phase

In the announcement phase, each HSR node learns MAC addresses of DANHs by broadcasting an announcement (Ann) message. The MAC addresses are then stored in a table called neighbor (Ne) table.

b. Path establishment phase

In this phase, all the DANHs establish DANH-based dual paths with each other DANHs. To discover and establish DANH-based dual paths, each DANH sends PaS messages to other DANHs listed in its Ne table. When a node, such as QuadBox or DANH that is not the destination DANH, receives a PaS message, it forwards the message to other nodes. The destination DANH receives two identical copies of the PaS message. Once the destination DANH receives a PaS message, it replies by sending a PaC message back to the source DANH. There are two PaC messages sent back to the source DANH from the destination DANH; each travels through one of the established dual paths. Upon receiving PaS and PaC messages, QuadBoxes in between the dual paths build a final path (FP) table. The FP table is then used to forward HSR frames from the source to the destination.

c. Final phase

When the source DANH sends unicast frames to the destination DANH, the frames are delivered through the pre-established dual paths between the source and the destination.

2.2.1.2. Advantages and disadvantages

As a dual paths-based technique, DVP significantly reduces redundant unicast traffic compared with the standard HSR protocol by forwarding unicast traffic frames through pre-established dual paths instead of duplicating and flooding in the overall network.

However, since the DVP technique establishes dual paths for each connection pair of DANH nodes, the number of connection pairs is very large. This results in high control overhead for discovering and establishing dual paths in HSR networks.

2.2.2. Ring-based dual paths (RDP)

RDP discovers and establishes two separate paths for each connection pair of DANH rings in HSR networks. These dual paths are ring-based dual paths. Like DVP, the RDP uses control messages to discover and establish the ring-based dual paths.

2.2.2.1. Operations

The functions of discovering and establishing dual paths between rings are implemented at QuadBox nodes in HSR networks. The dual paths establishment of RDP consists of three phases, namely building the MAC table, establishing ring-based dual paths, and building the forwarding table.

a. Building the MAC table

Each DANH periodically sends an HSR_Supervision frame over both its ports. Each access QuadBox learns MAC addresses of DANH nodes connected to its DANH ring based on the HSR_Supervision frames sent by the DANH nodes. By learning the MAC addresses of DANH nodes, each access QuadBox builds its MAC table that contains MAC addresses of DANHs that connect to its DANH ring. By looking up the MAC table, access QuadBoxes do not forward unicast frames into nondestination DANH rings.

b. Establishing ring-based dual paths

In the phase of establishing dual paths, the following two-step process is performed:

Step 1. Path request. In this step, each access QuadBox sends a path request (PREQ) message to discover ring-based dual paths. Upon receiving the PREQ message, QuadBoxes process information fields of the message, add their node ID into the received message, and then forward the updated message to other QuadBoxes. Based on information of the received PREQ message, trunk QuadBoxes update their ring table that contains MAC addresses of DANHs and ring IDs of DANH rings, whereas access QuadBoxes reply by sending a path reply (PREP) message back to the sending access QuadBox.

Step 2. Path reply. When an access QuadBox has received a PREQ message sent by another access QuadBox for the first time, the access QuadBox builds a path between the sending and receiving access QuadBoxes. The path is the first ring-based path between the source DANH ring and the destination DANH ring. The QuadBox then sends a PREP message back to the sending access QuadBox.

c. Building the forwarding table

Based on received PREQ and PREP messages, trunk QuadBoxes in between ring-based paths build their forwarding table. Each entry of the forwarding table consists of a source ring, a destination ring, and an output port.

2.2.2.2. Advantages and disadvantages

Like other dual paths-based techniques, the RDP technique significantly reduces redundant unicast traffic in HSR networks compared with the standard HSR protocol by forwarding the unicast traffic through two predefined paths. Like DVP, the RDP technique uses control messages to discover and establish dual paths. However, unlike the DVP technique that establishes DANH-based dual paths, the RDP technique establishes ring-based dual paths; therefore, the RDP dramatically reduces the number of established dual paths, which in turn decreases additional control overhead compared with the DVP technique.

However, because of using control messages to discover and establish dual paths, the RDP technique still generates additional control overhead to setup dual paths in HSR networks.

2.2.3. Dual separate paths (DSP)

DSP establishes two separate paths for each connection pair of QuadBoxes in HSR networks. These dual paths are QuadBox-based dual paths. Unlike DVP and RDP that use control messages to discover and establish dual paths, the DSP technique finds the QuadBox-based dual paths based on network topology information.

2.2.3.1. Operations

The establishment of dual paths is implemented at QuadBoxes. The DSP technique first builds MAC and link tables for each QuadBox, then discovers and establishes dual paths for each connection pair of access QuadBox, and finally, sets up the dual paths and builds a forwarding table for each trunk QuadBox.

a. Building the MAC table

Both access and trunk QuadBoxes learn and build their MAC table; however, the MAC table of access QuadBoxes is different from that of trunk QuadBoxes. Access QuadBoxes learn MAC addresses of DANH nodes connected to its DANH ring, whereas trunk QuadBoxes learn MAC addresses of all DANH nodes in the HSR network. Like in the RDP technique, each access QuadBox learns the MAC addresses of DANH nodes connected to its DANH ring based on HSR_Supervision frames sent by the DANH nodes. The process of learning MAC addresses for trunk QuadBoxes is similar to that of Ethernet switches. Trunk QuadBoxes learn MAC addresses of all DANH nodes in the network.

b. Building the link table

RDP establishes ring-based dual paths in an HSR network based on the network's link information. Each QuadBox builds and maintains a link table that contains information of all links in the network. QuadBoxes first build a neighbor list by exchanging Hello messages. Two access QuadBoxes of a QuadBox pair have the same node ID and the same neighbor list. QuadBoxes then send a Link message that contains their neighbor list. Upon receiving Link messages, QuadBoxes build their link table.

c. Finding dual paths

Each access QuadBox pair finds QuadBox-based dual paths to all other access QuadBox pairs. To find the dual paths for each connection pair of QuadBox pairs, each access QuadBox pair applies the DSP algorithm to the network link table. The DSP algorithm consists of the following steps: searching step, sorting step, and selecting step.

Step 1. Searching step: In this step, the DSP algorithm searches all available paths between the source and destination QuadBox pairs in the network. The searched paths associated with their path distance are then added to a path list.

Step 2. Sorting step: The path list found in the searching step is sorted in ascending order of path distance.

Step 3. Selecting step: In this final step, based on the sorted path list, the DSP algorithm selects two node-disjoint paths that have the best path distances.

d. Establishing dual paths

After finding dual paths for each connection pair of access QuadBox pairs, the QuadBox pair with a lower node ID sends a path setup (PSET) message through each path of the dual path to the corresponding QuadBox pair. The corresponding QuadBox pair replies by sending a path acknowledgment (PACK) message once it receives the PSET message. Based on the received PSET and PACK messages for each connection pair of QuadBox pairs, trunk QuadBoxes in between the dual paths build their forwarding table. Each entry of the forwarding table consists of the source QuadBox pair ID, the destination QuadBox pair ID, and the corresponding output port.

2.2.3.2. Advantages and disadvantages

Like DVP and RDP, the DSP technique significantly reduces redundant unicast traffic in HSR networks compared with the standard HSR protocol. In addition, unlike the DVP technique that discovers and establishes DANH-based dual paths for each connection pair of DANH nodes, the DSP technique finds QuadBox-based dual paths for each connection pair of QuadBox nodes; therefore, the DSP technique significantly reduces the number of connection pairs, which in turn reduces additional control overhead for discovering and establishing dual paths compared with the DVP technique.

However, because of using control messages to build the link tables and establish the dual paths, the DSP technique still generates additional control overhead in HSR networks.

2.2.4. Simulations and comparisons

These dual paths-based techniques discover and establish dual paths for each connection pair of nodes in HSR networks. These dual paths are then used to forward unicast frames from

http://dx.doi.org/10.5772/intechopen.71783

the source to the destination. **Figure 8** shows the process of forwarding a unicast frame from source DANH 1 to destination DANH 10 through dual paths under the dual paths-based technique.

As described above, these dual paths-based techniques have advantages and disadvantages. While the DVP and RDP techniques discover and establish dual paths by sending and receiving control messages, the DSP technique establishes dual paths based on the network topology information. The DVP technique setups DANH-based dual paths for each connection pair of DANHs, while the RDP and DSP techniques establish ring-based and QuadBox-based dual paths for each connection pair of DANH rings and for each connection pair of QuadBox nodes, respectively. Therefore, the RDP and DSP techniques significantly reduce the number of connection pairs required to discover and establish dual paths. The characteristics of these dual paths-based techniques are summarized in **Table 2**.

Several simulations were conducted using the OMNeT++ simulator to compare the traffic performance of these dual paths-based techniques. We consider the sample HSR network as shown in **Figure 2**. In these simulations, source DANH 1 in DANH ring 1 sends unicast frames to destination DANH 10 in DANH ring 3. The line graph in **Figure 9** shows the comparison of traffic performance of these dual paths-based techniques. Overall, it can be seen from the graph that these traffic reduction techniques have the same network traffic performance and significantly reduce network unicast traffic compared with standard HSR protocol.

2.3. HSR SwitchBox technique

Unlike the mentioned techniques, which propose algorithms implemented in existing HSR components to improve the network traffic performance in HSR networks, the HSR SwitchBox

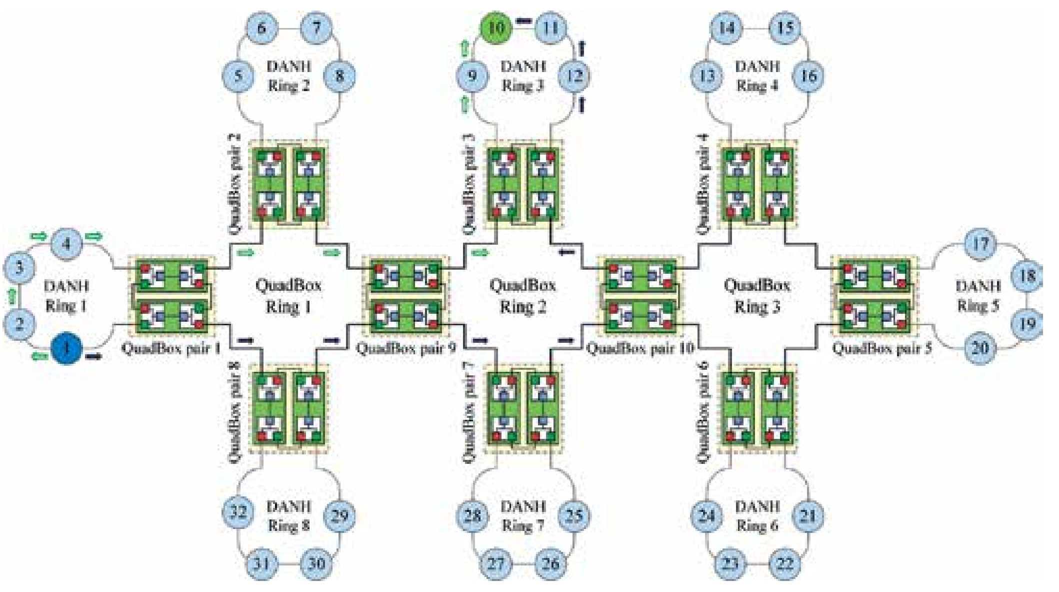

Figure 8. The process of forwarding a unicast frame under the dual paths-based techniques.

Features	DVP	RDP	DSP
Type of dual paths	DANH-based	Ring-based	QuadBox-based
Method of establishing dual paths	Control messages	Control messages	Topology information
Number of dual paths	Large	Small	Small
Generated control overhead	High	Medium	Medium

Table 2. Characteristics of dual paths-based techniques.

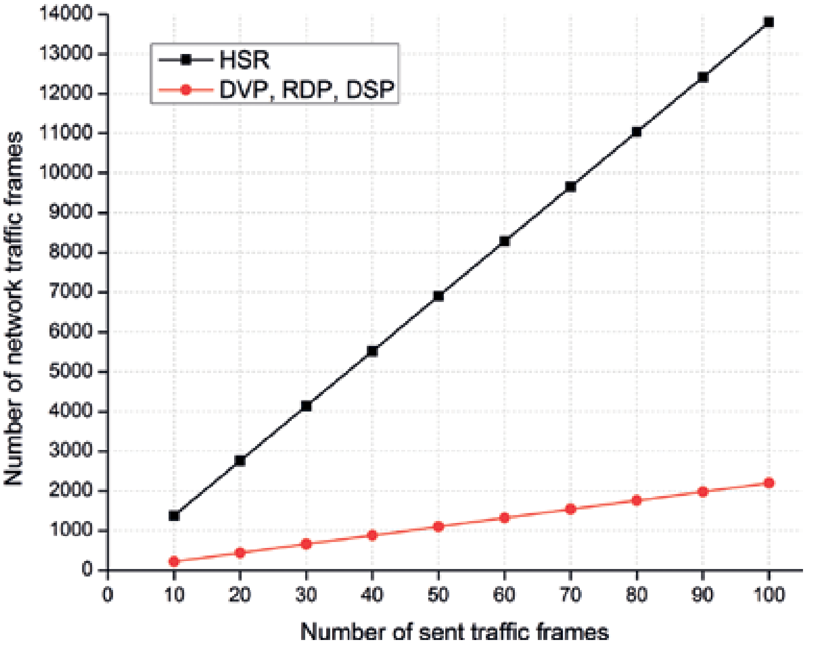

Figure 9. A comparison of the traffic performance of dual paths-based techniques.

technique [21] defines a new HSR switching node (SwitchBox) for HSR networks. By using the SwitchBoxes, HSR can be applied to any topology, such as a ring, mesh, or star topology. In addition, SwitchBox-based HSR significantly reduces redundant network traffic compared to the standard HSR.

2.3.1. Operations

An HSR SwitchBox is a switching node with many ports that performs the switching functionality in HSR networks. SwitchBoxes forward a received HSR frame by looking up the MAC table instead of flooding and circulating the HSR frame in the HSR network as current HSR nodes.

The HSR SwitchBox technique defines the following new forwarding rules:

- Do not forward HSR unicast frames into nondestination DANH rings.
- Forward HSR unicast frames on network trunk links.
- Prevent HSR frames from being circulated in HSR networks.

The operation of HSR SwitchBox technique consists of three phases, including setting port type, building MAC table, and forwarding frame.

a. Setting port type

There are two types of port defined for HSR SwitchBoxes: access and trunk ports. An access port is connected to a DANH node, whereas a trunk port is connected to another SwitchBox. SwitchBox ports are set to the access type, by default. The type of SwitchBox ports can be configured automatically or manually. A Hello message is used to automatically configure the port type. Whenever a port of a SwitchBox changes its status to up, the SwitchBox sends a Hello message over the port. When a SwitchBox receives a Hello message on a port, it sets the type of the port to the trunk type.

b. Building MAC table

Each SwitchBox learns the MAC addresses of DANHs that connect to it based on receiving the HSR_Supervision frames sent by the DANH nodes. Each SwitchBox builds its MAC table, which contains MAC addresses of the DANH nodes that connect to the SwitchBox.

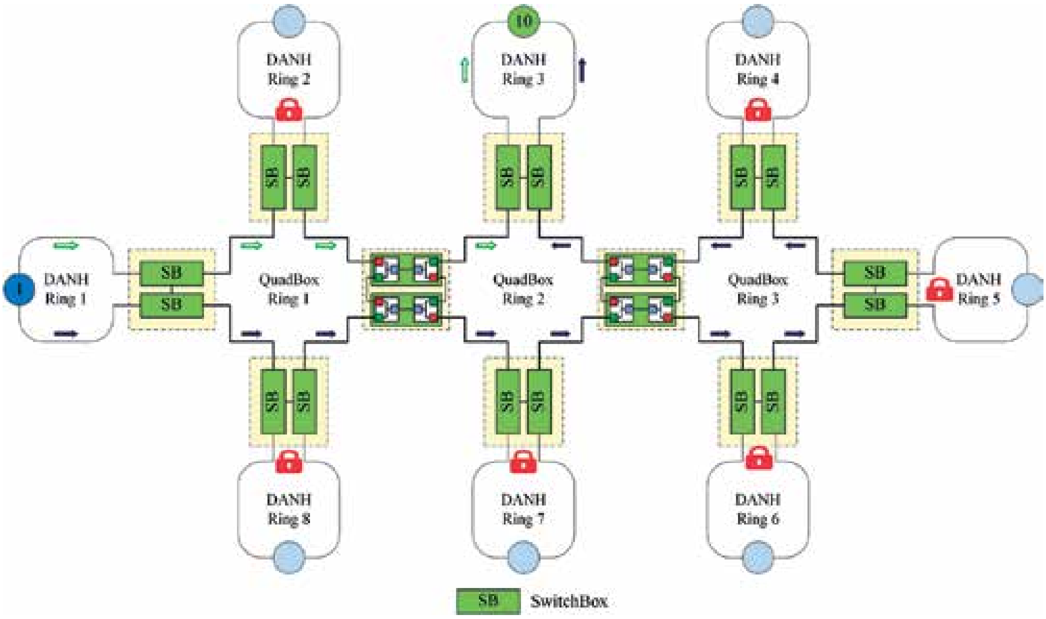

Figure 10. The process of forwarding a unicast frame under the SwitchBox-based HSR.

c. Forwarding frame

Instead of flooding and circulating HSR unicast frames, SwitchBoxes forward the frames based on looking up the MAC table. When a SwitchBox receives a unicast frame, it checks if its MAC table contains the destination MAC address of the frame. If not, it sends the received frame over its trunk ports, except the received port. If so, it forwards the frame to the port that connects to the destination node. The SwitchBox discards duplicated copies of the frame.

When source DANH 1 in DANH ring 1 sends unicast frames to destination DANH 10 in DANH ring 3, SwitchBox-based HSR filters the unicast traffic for DANH rings that do not contain the destination node, as shown in **Figure 10**.

2.3.2. Simulations and comparisons

Several simulations were carried out using the OMNeT++ simulator to evaluate and compare the traffic performance of the SwitchBox-based HSR with that of the standard HSR. In these simulations, we considered the sample HSR network as shown in **Figure 2**. **Figure 11** shows the comparison of the traffic performance between the SwitchBox-based HSR and the standard HSR. Unlike standard HSR, which floods unicast frames to all rings, SwitchBox-based HSR does not forward the unicast frames to nondestination DANH rings. In addition, SwitchBox-based HSR prevents the frames from being duplicated and circulated in rings. Therefore, SwitchBox-based HSR significantly reduces network traffic compared with

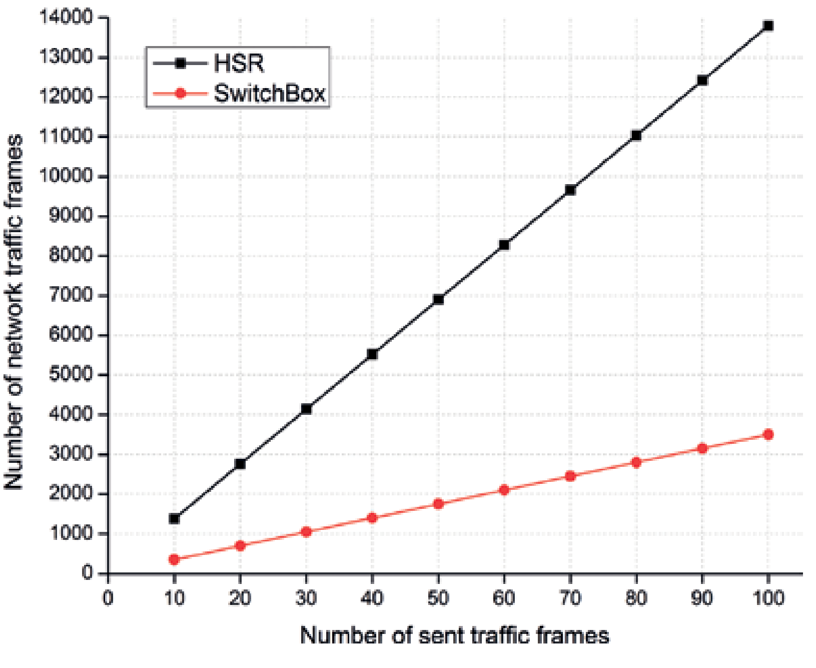

Figure 11. A comparison of the traffic performance between standard HSR and SwitchBox-based HSR.

http://dx.doi.org/10.5772/intechopen.71783

the standard HSR. Numerically, the simulation results show that the SwitchBox-based HSR reduced network unicast traffic by 75% compared to the standard HSR.

3. Conclusions

The standard HSR protocol generates excessively unnecessary redundant unicast traffic in HSR networks, resulting in the degradation of network performance. Several traffic reduction techniques have been proposed to solve this problem. In this chapter, we present a review of HSR traffic reduction techniques. These traffic reduction techniques are divided into two categories: traffic filtering-based and dual paths-based techniques. Each HSR traffic reduction technique has its own advantages and disadvantages. The selection of which technique to implement depends on the particular application and trade-offs. With this chapter, researchers can acquire what has been investigated, and network designers can identify which technique to use and what are the trade-offs.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (grant number: 2017R1A2B4003964).

Author details

Jong Myung Rhee* and Nguyen Xuan Tien

*Address all correspondence to: jmr77@mju.ac.kr

Information and Communications Engineering Department, Myongji University, Republic of Korea

References

[1] IEEE. IEEE 802.3: IEEE Standard for Ethernet [Internet]. [Updated: 2015]. Available from: http://standards.ieee.org/about/get/802/802.3.html [Accessed: Aug 10, 2017]

[2] IEEE. 802.1D-2004–IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges [Internet]. [Updated: 2004]. Available from: http://standards.ieee.org/getieee802/download/802.1D-2004.pdf [Accessed: Aug 10, 2017]

[3] IEC. Industrial Communications Networks: High-Availability Automation Networks, Part 2: Media Redundancy Protocol (MRP). IEC Standard 62439-2. 2010

[4] IEEE. IEEE 801.1aq Standard for Shortest Path Bridging (SPB) [Internet]. Available from: http://www.ieee802.org/1/pages/802.1aq.html [Accessed: July 30, 2017]

[5] Tien NX, Rhee JM. RPE: A seamless redundancy protocol for Ethernet networks. IEICE Transactions on Communications. 2017;**E100-B**(5):711-727. DOI: 10.1587/transcom.2015EBP3486

[6] IEEE. IEEE Time-Sensitive Networking Task Group [Internet]. Available from: http://www.ieee802.org/1/pages/tsn.html [Accessed: July 30, 2017]

[7] Kehrer S, Kleineberg O, Heffernan D. A comparison of fault-tolerance concepts for IEEE 802.1 time sensitive networks (TSN). In: The 2014 IEEE Emerging Technology and Factory Automation (ETFA); 16-19 September 2014; Barcelona, Spain. 2014. pp. 1-8

[8] IEC. Industrial Communications Networks: High-Availability Automation Networks, Part 3: Parallel Redundancy Protocol (PRP) and High-Availability Seamless Redundancy (HSR). IEC Standard 62439-3. 2012

[9] Prytz G. Network recovery time measurements of RSTP in an ethernet ring topology. In: The 2007 IEEE Conference on Emerging Technologies and Factory Automation (ETFA2007); 25-28 September 2007; Patras, Greece. 2007. pp. 1247-1253

[10] Giorgetti A, Cugini F, Paolucci F, Valcarenghi L, Pistone A, Castoldi P. Performance analysis of media redundancy protocol (MRP). IEEE Transactions on Industrial Informatics. 2013;**9**(1):218-227. DOI: 10.1109/TII.2012.2186584

[11] IEEE. IEEE Audio and Video Bridging Task Group [Internet]. Available from: http://www.ieee802.org/1/pages/avbridges.html [Accessed: July 30, 2017]

[12] Kirrmann H., Weber K., Kleineberg O., Weibel H.. HSR: Zero recovery time and low-cost redundancy for industrial ethernet (high availability seamless redundancy, IEC 62439-3). In: The 2009 IEEE Conference on Emerging Technologies & Factory Automation (ETFA2009); 22-25 September 2009; Mallorca, Spain. 2009. pp. 1-4

[13] IEC. Network Engineering Guidelines for Communication Networks and Systems in Substations. IEC Standard 61850-90-4. 2013

[14] Nsaif SA, Rhee JM. Improvement of high-availability seamless redundancy HSR traffic performance for smart grid communications. Journal of Communications and Networks. 2012;**14**(6):653-661. DOI: 10.1109/JCN.2012.00031

[15] Abdulsalam IR, Rhee JM. Improvement of high-availability seamless redundancy (HSR) unicast traffic performance using port locking. In: The Fourth World Congress on Software Engineering (WCSE); 3-4 December 2013; Hong Kong, China. 2013. DOI: 10.1109/WCSE.2013.45

[16] Tien NX, Rhee JM. FHT: A novel approach for filtering high-availability seamless redundancy (HSR) traffic. Energies. 2015;**8**(7):6249-6274. DOI: 10.3390/en8076249

[17] OMNeT++. Discrete Event Simulator [Internet]. Available from: https://omnetpp.org/ [Accessed: August 30, 2017]

[18] Nsaif SA, Rhee JM. DVP: A novel high-availability seamless redundancy (HSR) protocol traffic-reduction algorithm for a substation automation system network. Energies. 2014;**7**(3):1792-1810. DOI: 10.3390/en7031792

[19] Tien NX, Kim S, Rhee JM. A novel ring-based dual paths approach for reducing redundant traffic in HSR networks. Computer Networks. 2016;**110**:338-350. DOI: 10.1016/j.comnet.2016.10.009

[20] Tien NX, Kim S, Rhee JM, Park SY. A novel dual separate paths (DSP) algorithm providing fault-tolerant communication for wireless sensor networks. Sensors. 2017;**17**(8):1699. DOI: 10.3390/s17081699

[21] Tien NX, Rhee JM. Developing a new HSR switching node (SwitchBox) for improving traffic performance in HSR networks. Energies. 2016;**9**(1):36. DOI: 10.3390/en9010036

Chapter 20

Mobile Broadband Scaling and Enhancement for Fast Moving Trains

Vipin Balyan, Mario Ligwa and Ben Groenewald

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.71782

Abstract

Internet is an important part of our life, whether traveling or at home. The broadband services available at home are reliable and are usually at constant speed. The people traveling especially in fast moving trains are at higher mobility and may be moving in areas of less connectivity, and providing a reliable service to them is a challenging task. One possible solution to this is to provide communication through an on-board Wi-Fi, which takes services from a central Wi-Fi situated in the middle of the train, which is connected to cellular radio service long-term evolution for railways. The network consists of LTE-R which is dedicated for railway communication only, a public mobile network, which supports LTE-R in the areas of no coverage and high traffic conditions and a public safety network in emergency conditions. The work is verified with the help of simulations on MATLAB, considering different traffic scenarios. The BSs placed at a distance of 2.5 Km and antenna height used is 45 m are equipped with 3G and 4G interfaces, a universal mobile telecommunications services (UMTS) and radio access network (RAN). The UMTS interface is used for voice services and handover when spectrum available in the next cell is less.

Keywords: RAN, LTE-R, GSM-R, UMTS, access schemes, high-speed trains

1. Introduction

The internet is an inseparable technology from our day to day life. Certain countries are working on deploying LTE-R specially dedicated for railway communications. The Republic of Korea has brought into place the national disaster safety network since 2015, which costs over $1.6 billion and is established in 700 MHz frequency band [1]. The LTE-R network also works in the same frequency band. However, the global system for mobile-railway (GSM-R) is the most widely used communication standard, especially in Europe. The GSM-R is employed for more than 10 years as a stable integrated wireless communication technique. With the emerging technology, the rate requirement of the calls arrive in the network also increased, GSM-R is not able to meet the service requirements both in terms of capacity and speed. The LTE-R emerged as a solution to meet the demand generated, it provides benefit related to connectivity and also increases performance [2]. The public safety network mainly used by fire-fighters, police, rescue team, medical emergency, etc., and sometimes railway network also uses it for communications related to controlling of the train and its crew. For the reliable and secure functionality of railway communication, a dedicated and fast communication network is required and LTE-R provides a dedicated and reliable frequency band. The on-board passengers also require high rate data services, which is also a challenging task. For the efficient utilization of this available frequency band, research needs to be done to lay down assignment schemes which can use it optimally.

2. Related work

The work in the literature on the LTE-R communication technology is related to modeling of channels with a layout of LTE-R network [3–5]. In [6], coexistence of railway and public safety network is considered; it poses challenges like co-channel interference and priority of services. In the literature, techniques have been proposed to reduce the co-channel interference together with interference alignment and channel diagonalization. One proposed in [7] uses coordinated multipoint (CoMP) by utilizing a two-step precoder in presence of a multi-user CoMP. The paper in [8, 9] proposes schemes related to power control and interference management in 3rd Generation partnership project (3GPP). For all the schemes proposed [6–9], in order to achieve benefits in form of better quality of service (QoS), fairness in assignment and load balancing a complex feedback mechanism is required to provide channel state information (CSI) additionally. The work in [6] employed enhanced inter-cell interference cancelation (eICIC) and further eICIC (FeICIC) in presence of coordinated scheduling (CS) CoMP under the RAN sharing case for offloading more public safety users to the railway network. However, CS CoMP is utilized for the LTE-R eNodeBs. In [10], a dynamic ICIC along with CS CoMP is employed in order to perform interference management for both public safety and railway networks existing together. The paper considers a CS CoMP between public safety (PS) LTE and LTE-R eNodeBs, public safety LTE eNodeBs, and LTE-R eNodeBs. The radio resource assignment management is investigated like a resource sharing scheme which is ware of interference in [11], a joint scheduling mechanism in [12], and a game-based resource allocation in [13]. These schemes optimized system efficiency and throughput by using resource assignment independently. The assignment schemes algorithm did not provide priority to any type of calls and no consideration for mission-critical services (MCS) of a user.

The literature also contains research papers which work on LTE-MIMO performance improvement when using antenna arrays; the work is very limited for their employment for railway communication in [14]. The high-speed railway's unique property [15] is the presence of the line of sight (LOS) component. Due to the unique characteristics of high-speed railways [16], for instance, the existence of the LOS component and the deficiency of scattering in a series of bridges seriously influence the MIMO performance [17]. The work needs to be done on enhancing characteristics of antenna arrays in order to enhance the LTE-R efficiency [15–17]. The LTE-R in [18], which are LTE specifications for railway communications, they are proposed in order to meet the high-speed train requirements of broadband communication. The handover of calls is a critical issue, which becomes more critical for real-time calls as the probability of handover failure are more in high-speed railway due to high speed. The problem of handover is enhanced due to the existence of only hard handover supports in LTE. The hard handover needs to be taken properly and in time for non-disruption and drop of calls during handover.

The handover decision taken too early and too late both will lead to disruption in calls. In this chapter, the handover is done with the help of a device mounted on the train boggy, and passengers on-board get seamless handover. Also, the paper uses three types of LTE network: the public mobile LTE network, public safety LTE network, and LTE-R network. The LTE-R is used for railway communication services and passengers' on-board services, the public mobile LTE network is used for providing carrier aggregation (CA) [19] and access in areas with no LTE-R services. The public safety LTE-R is used in case of emergency services in areas with no LTE-R coverage.

The rest of the chapter is organized as follows. Section 3 discusses the problem statement and the network parameters used. In Section 4, the proposed work is explained. Simulation and results are given in Section 5. The chapter is concluded in Section 6.

3. Problem statement and network parameters

The LTE-R is a solution of railway communication which is employed to handle voice and data traffic for high-speed trains. Most of the research work in the literature is focused on providing compatibility of LTE-R with previous GSM-R, using public safety LTE networks which can be used for MCSs. The main problem in LTE-R communication is handover and availability. In this chapter, the LTE-R network is used for railway communication which uses public mobile LTE network for providing better services of railway and public safety LTE network in emergency conditions. The LTE system for any type LTE-R, LTE public mobile network, and LTE public safety all contains remote radio heads (RRH) which are connected to eNodeBs which are connected to each other by X.2 line and in the backbone connected to the wireless core network as shown in **Figure 1**. The RRH is used in all the three types of cellular systems. The railway communication system which uses RRH deploy fiber network to send information along the track. The LTE-R, LTE public mobile network, and LTE public safety deployed are equipped with the *eNodeB of two interfaces* UMTS *and* LTE*. The eNodeB provides access to user equipment's (UE) with different traffic requirements and different mobility. The 4G* LTE network structure and its 3G network are explained in this section. The *LTE* network with its 3G interface is illustrated in **Figure 2**.

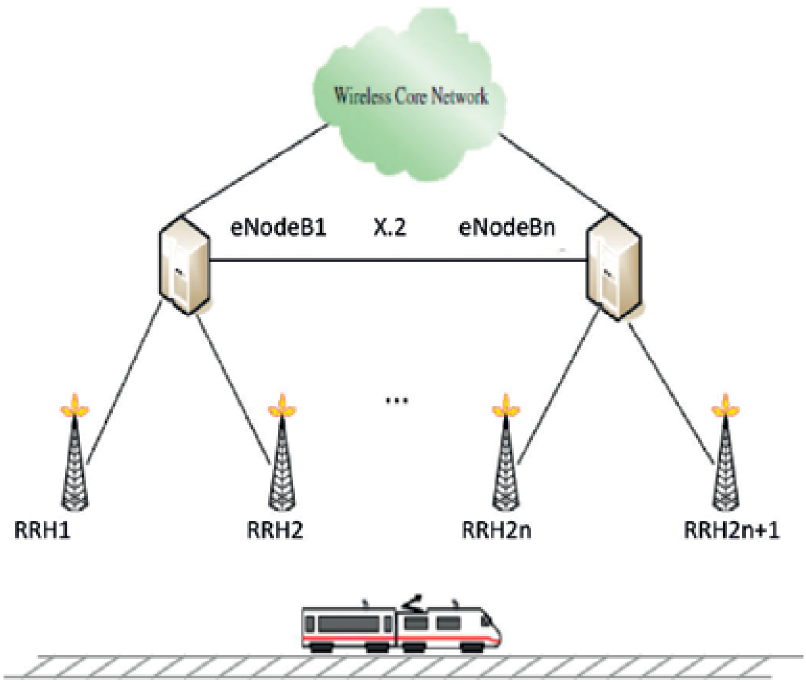

Figure 1. LTE-R system based on RRH.

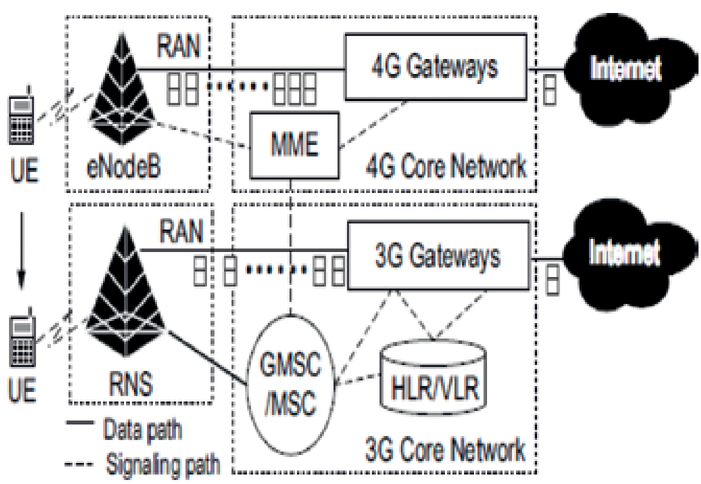

Figure 2. 4G-3G core network architecture.

The data (packet) service is offered by *LTE* network. It consists of a core network radio access network (*RAN*) and mobile stations (MSs). Its *RAN* uses eNodeB, i.e., *LTE* base station (BS) which allows access to MSs. The network core is IP-based and uses mobile management entity (MME) in order to locate MSs movement, e.g., location update and paging information. The 4G gateways are used to route packets between the 4G RAN and the Internet.

In contrast, the 3G network provides support to both data and voice calls or in other words packet switched and circuit switched calls. Its *RAN* uses radio network system (RNS) to allow access to radio resources. Its network consists of (a) Gateway mobile switching center (GMSC/VLR) which stores/updates user location. (b) 3G gateway provides data (packet) service and the route between the RAN and the internet.

The UMTS [20] interface adopts VSF-OFCDM in order allocate OVSF codes of a code tree spread in two dimensions: time and frequency. The LTE/LTE-A uses orthogonal frequency division multiplexing access (OFDMA) which uses a fixed frame for downlink transmission. The size of a radio frame in OFDMA is 10 ms which is divided into 1 ms, 10 sub frames. Each sub frame is further divided into two slots of 0.5 ms. Each slot has seven or six consecutive OFDM symbols. A basic scheduling unit in LTE is resource block (RB) which is composed of a time slot in the time domain and in frequency domain 12 consecutive sub carriers. The RB(s) are allocated to a call(s) when a call arrives or may vary at each transmission time interval (TTI).

3.1. LTE-R services

The LTE-R communication architecture for railways is given in **Figure 3** [21]. The main components of it are *Base station controller (BSC), home subscriber server (HSS), policy and charging rules function (PCRF), mobility management entity (MME), serving general packet radio service (GPRS) support node (SGSN), and packet data network (PDN).*

The LTE-R communications are used to provide services with minimum latency and least failure. The suggestion given in E-Train project [6], LTE-R must provide the services given below with higher level of security, the higher efficiency with better QoS.

1. *Control Systems Information Transmission:* The control information must be transmitted wirelessly in real time with a time delay less than 50 ms. The information related to the

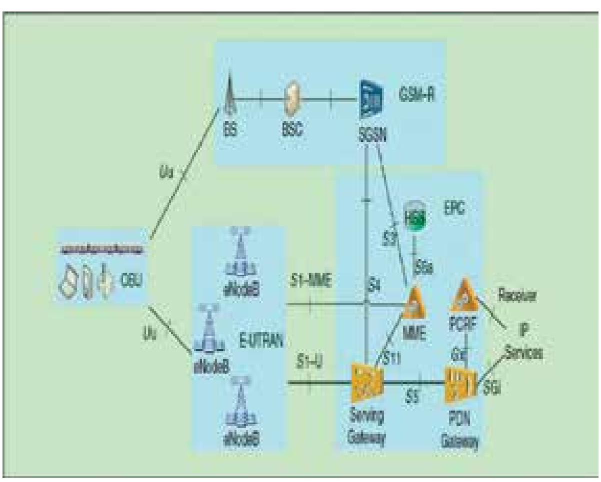

Figure 3. The LTE -R architecture for HSR communication.

location of the train is detected by radio block center (RBC) and radio equipments on the train. This will enhance the accuracy of tracking and dispatch of the trains. LTE-R may also be used in future for transmission of information for automatic driving conditions.

2. *Monitoring in Real-Time:* The LTE-R can provide video monitoring of all the parts involved in railway transport like rail tracks, rail bogies, connector, etc., in real time. The video monitoring of infrastructures where railway tracks are running, e.g., tunnels in order to provide safety in case of natural or man-made disasters. This monitored information needs to be shared at two places at the same time the control center and the train in real time with a delay not greater than 300 ms.

3. *Multimedia Dispatching:* The LTE-R provides information of drivers and yards to the dispatcher and improves dispatching efficiency [22].

4. *Railway Emergencies Information:* In case of emergency information, like two trains running on the same track, failure of engines, track broken, and accidents, the information needs to be sent not only to the railway authorities, but also to the public safety departments like ambulance, fire-fighters, police, etc., for faster rescue operation. The communication needs to be fast, accurate, and may contain images or videos with a delay not greater than 100 ms when containing videos or images.

5. *Internet of Things (IoT) of Railway Communications:* The railway IoT services like trains tracking, real-time queries, mail services, etc.

In addition to these services, LTE-R must have provisions to provide services like e-ticketing in mobility, upgrading of passenger information, seat reservations dynamically.

4. Proposed work

The LTE system consists of train access units (TAU) which are the on-board unit for access. The number of TAUs used on the train depends on the number of train bogies. In this chapter, two TAUs are placed, one in front and another in back, to reduce their mutual interference. A third TAU is placed which helps to communicate in borrowing capacity from LTE public mobile network and LTE public safety network. The TAUs are connected to inboard access points through an optical fiber. The passengers in the train experience a seamless wireless access. The call with their types and priority are defined below:

A. Emergency calls: Highest Priority needs urgent attention and is denoted as EC.

B. Railway control, track monitoring, railway dispatch information's: Medium Priority and are denoted by MP.

C. Data traffic generated by Passengers: Least Priority denoted by LP.

The algorithm works as follow when a call arrives:

i. Generate a call.

ii. Check the call type (A,B,C)?

iii. For HP (Check whether a voice or data call?)

Data call

Assign LTE interface of LTE-R network. If no RBs available in LTE-R network, shift MP and LP calls using LTE-R network to public safety LTE network which is supporting the LTE-R network to handle HP calls.

Voice call

Assign UMTS interface.

iv. For MP (Check whether a voice call or data call)

Data call

Assign LTE interface of LTE-R network. If no RBs available in LTE-R network, shift LP calls using LTE-R network to mobile LTE network which is supporting the LTE-R network to handle MP calls.

Voice call

Assign UMTS interface.

v. For LP (Check whether a voice call or data call)

Data call

Assign LTE interface of LTE-R network. If no RBs available in LTE-R network, use mobile LTE network which is supporting the LTE-R network to handle LP calls.

Voice call

Assign UMTS interface.

vi. End the call and release the resources.

The calls are served by on-board wireless units, which request for capacity from TAUs. The TAUs request for capacity from RRHs which are connected to eNodeB. The algorithm works fine for any type of call, any location when a number of users are limited and it’s in connectivity area of LTE-R.

Let the number of users which LTE-R can serve are U_{LTE-R} and the number of call request are C_r, where $C_r > U_{LTE-R}$. The LTE-R network needs to borrow capacity for remaining calls $C_r - U_{LTE-R}$. The TAU situated in center will facilitate it, all the calls request which will arrive when the LTE-R capacity is fully utilized will be served by central TAU. The TAU will send the request to LTE public mobile network eNodeB, which may be placed at a larger distance as compared to LTE-R eNodeB placed closely. These calls handled will be at higher latency. When the capacity of LTE-R becomes available, some of the calls are shifted to the LTE-R network again. The LTE public mobile network is used for rural areas also, as LTE-R is not deployed in remote areas.

The handover request is also done through these TAUs; the handover request is not between user equipments and eNodeB. The TAU request for handover to RRH (eNodeB) is done well in advance by estimating the speed of the train and time to connect. The total handover is done using TAU which is seamless for user perspective. The algorithm uses the distance by which two eNodeBs are separated and the power TAU receiving from them (which basically depends upon speed). The problem in handover is complex and it becomes more complex when the next serving eNodeB does not have enough capacity to support, the algorithm in this chapter searches for nearby LTE public mobile network RRH and request for connectivity and if LTE public mobile network does not have enough capacity to handle the request. The capacity load of the three TAUs is shared among three types of networks. The algorithm tries to utilize LTE-R capacity first and according to the preference of the calls.

The LTE-R is not totally equipped with carrier aggregation (CA) [19], which is widely employed in LTE-A. In this chapter, the CA is used in simulations to show the benefits of CA. The CA is used for higher data loads and leads to completion fast. With the use of CA in LTE-A, the networks are used again and the throughput increases considerably. With the employment of CA, UE can simultaneously use two or more frequency bands of 20 MHz, the complete description of which will be done in the sequel of this chapter.

5. Results and simulation

In this section, the performance of the proposed work is investigated using MATLAB. The results are compared for various scenarios for coexistence of public safety LTE, public mobile LTE network, and LTE-R. The resources of public mobile LTE network and public safety LTE are used to enhance the services of LTE-R. The LTE-R network runs parallel to railway track at a distance of 2.5 km from the track and with an antenna of height 45 m, a public mobile LTE network is placed at a distance of 4 km from railway track and public safety LTE network when used for simulations. *The call arrival rate is average from 0 to 6. The arrival of calls during high-passenger traffic scenario is LP calls (70%) and (MP + HP) calls 30%. In low-passenger traffic scenario, the call arrival is LP calls (20%) and (MP + HP) calls 80%. The average call duration of all traffic rates is exponentially distributed with normalized mean value 1. The LTE interface capacity used is 345.6 Mb/s and of UMTS interface 256*R (3.4 Mb/s). The same traffic is generated for network systems compared in **Figures 4** and **5**: LTE-R only (LTE-R), LTE-R + LTE public safety (LTE-RP), LTE-R + LTE public safety + LTE public mobile network (LTE-RPM), and reservation of capacity for handover in LTE-R systems. In **Figure 4**, the average throughput of the user calls is compared with average traffic load in Mbps, the throughput increases and after some time there is slight increase in throughput even though traffic load increases due to loss of the packets, for low passenger traffic condition throughput is higher as compared to high passenger traffic condition.

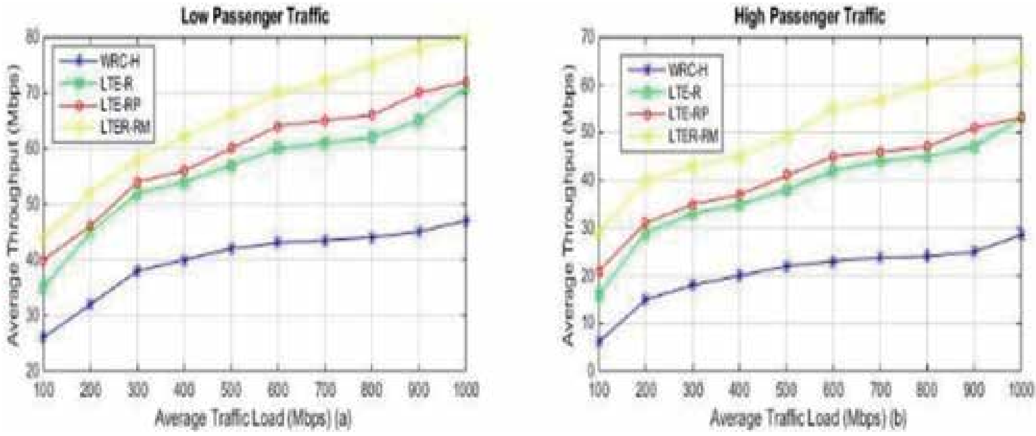

Figure 4. The average throughput comparison of different sharing method in: (a) Low passenger traffic scenario, (b) high passenger traffic scenario.

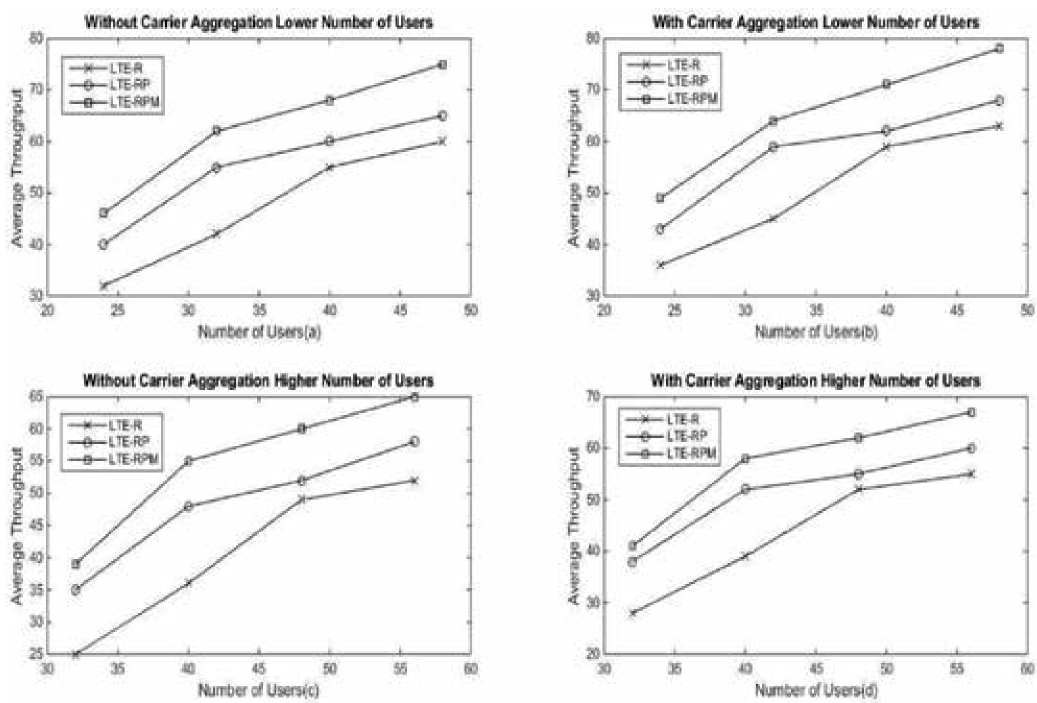

Figure 5. The average throughput comparison of different sharing method in: (a) and (b) lower number of UE with and without CA, (c) and (d) higher number of users with and without CA.

The network is simulated with CA capability means LTE-A is used and is compared with network without CA capability in **Figure 5**, the networks with CA and without CA are compared in lower user arrival **Figure 5(a)** and **(b)** and higher user arrival in **Figure 5(c)** and **(d)**. The network with CA provides better throughput, even though number of users are higher. The CA technique provides higher bandwidth and increases throughput.

6. Conclusion

The algorithm in this chapter provides analysis of LTE-R in coexistence with public safety LTE and LTE public mobile network. The main priority of the work is to handle emergency calls for railways in conjugation with other traffic on a train. The work is using existing network topologies without the requirement of any infrastructure changes. The LTE public mobile network has maximum coverage and its spectrum is not in fully utilized. Instead of deploying new BS, it is better to use the public mobile network for coverage in remote areas and for providing better QoS to on-board passengers. The work uses three networks alone and sharing modes. The throughput of users calls increases. Further, the simulations are done check the performance of the network. Moreover, RAN sharing is applied for the coexistence of two LTE network with CA. The use of CA keeps throughput closer to one achieved in without CA which provides greater benefit to users by achieving higher throughput in better channel condition.

Author details

Vipin Balyan*, Mario Ligwa and Ben Groenewald

*Address all correspondence to: vipin.balyan@rediffmail.com

Department of Electrical, Electronics and Computer Engineering, Cape Peninsula University of Technology, Cape Town, South Africa

References

[1] Choi JK, Cho H, Oh HS, Kim KH, Bhang MJ, Yu IS, et al. Challenges of LTE high-speed railway network to coexist with LTE public safety network. In: IEEE International Conference on Advanced Communication Technology, ICACT; 2015. pp. 543-547

[2] Calle-Sanchez J, Molina-Garcia M, Alonso JI, Fernandez-Duran A. Long term evolution in high speed railway environments: Feasibility and challenges. Bell Labs Technical Journal [Internet]. 2013;**18**(2):237-253. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84883399891&partnerID=tZOtx3y1

[3] Institute ETS. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation. 3GPP TS 36.211 Version 10.0.0 Release 10 [Internet]. 2011.

pp. 1-105. Available from: http://www.etsi.org/deliver/etsi_ts/136200_136299/136211/10.00.00_60/ts_136211v100000p.pdf

[4] Guan K, Zhong Z, Ai B. Assessment of LTE-R using high speed railway channel model. In: IEEE Third International Conference on Communications and Mobile Computing. Vol. 2. 2011. pp. 461-464

[5] Liu L, Tao C, Qiu J, Chen H, Yu L, Dong W, et al. Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz. IEEE Journal on Selected Areas in Communications. 2012;**30**(4):834-845

[6] Chen W, Ahmad I, Chang K. Co-channel interference management using eICIC/FeICIC with coordinated scheduling for the coexistence of PS-LTE and LTE-R networks. EURASIP Journal on Wireless Communications and Networking [Internet]. 2017;**2017**(1): 34 Available from: https://jwcn-eurasipjournals.springeropen.com/articles/10.1186/s13638-017-0822-6

[7] Ren P, Sun L, Xu D, Song H. Precoder-and-receiver design scheme for multi-user coordinated multi-point in LTE-A and fifth generation systems. IET Communications [Internet]. 2016;**10**(3):292-299. Available from: http://digital-library.theiet.org/content/journals/10.1049/iet-com.2015.0229

[8] Saad SA, Ismail M, Nordin R, Ahmed AU. A fractional path-loss compensation based power control technique for interference mitigation in LTE-A femtocell networks. Physical Communication. 2016;**21**:1-9

[9] Sun S, Gao Q, Peng Y, Wang Y, Song L. Interference management through CoMP in 3GPP LTE-advanced networks. IEEE Wireless Communications. 2013;**20**(1):59-66

[10] Ahmad I, Chen W, Chang K. LTE-railway user priority-based cooperative resource allocation schemes for coexisting public safety and railway networks. IEEE Access. 2017;**5**:7985-8000

[11] Zhang R, Cheng X, Yang L, Jiao B. Interference-aware graph based resource sharing for device-to-device communications underlaying cellular networks. In: IEEE Wireless Communications and Networking Conference, WCNC; 2013. pp. 140-145

[12] Wang F, Song L, Han Z, Zhao Q, Wang X. Joint scheduling and resource allocation for device-to-device underlay communication. In: IEEE Wireless Communications and Networking Conference, WCNC; 2013. pp. 134-139

[13] Wei W, Fan X, Song H, Member S, Fan X, Yang J. Imperfect information dynamic Stackelberg game based resource allocation using hidden Markov for cloud computing. IEEE Transactions on Services Computing. 2016;**1374**(c):1-13

[14] Ai B, Cheng X, Kurner T, Zhong ZD, Guan K, He RS, et al. Challenges toward wireless communications for high-speed railway. IEEE Transactions on Intelligent Transportation Systems. 2014;**15**(5):2143-2158

[15] Ai B, Guan K, Rupp M, Kurner T, Cheng X, Yin XF, et al. Future railway services-oriented mobile communications network. IEEE Communications Magazine. 2015;**53**(10):78-85

[16] Chen B, Zhong Z, Ai B, Chen X. Comparison of antenna arrays for MIMO system in high speed mobile scenarios. In: IEEE 73rd Vehicular Technology Conference (VTC Spring), Yokohama; 2011

[17] Bhagavatula R, Oestges C, Heath RW. A new double-directional channel model including antenna patterns, array orientation, and depolarization. IEEE Transactions on Vehicular Technology. 2010;**59**(5):2219-2231

[18] Tingting G, Bin S. A high-speed railway mobile communication system based on LTE. In: 2010 International Conference on Electronics and Information Engineering (ICEIE) [Internet]; Vol. 1; 2010; p. V1-414-V1-417. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5559665

[19] Liao H-S, Chen P-Y, Chen W-T. An efficient downlink radio resource allocation with carrier aggregation in LTE-advanced networks. IEEE Transactions on Mobile Computing [Internet]. 2014;**13**(10):2229-2239. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6701154

[20] Balyan V. Call departure reassignment and performance enhancement of OVSF codes in downlink of CDMA based networks. 2017;**12**(16):6170-6175

[21] 3GPP. TS 136 211 - V13.1.0 - LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (3GPP TS 36.211 Version 13.1.0 Release 13). Etsi. 2016;**0**:0-292

[22] He R, Ai B, Wang G, Guan K, Zhong Z, Molisch AF, et al. High-speed railway communications: From GSM-R to LTE-R. IEEE Vehicular Technology Magazine. 2016;**11**(3):49-58

On the Energy Efficiency of Virtual Machines' Live Migration in Future Cloud Mobile Broadband Networks

Raad S. Alhumaima, Shireen R. Jawad and
H.S. Al-Raweshidy

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72040

Abstract

In this chapter, a live migration of the virtual machine (VM) power consumption (PC) model is introduced. The model proposed an easy and parameterised method to evaluate the power cost of migrating the VMs from one server to another. This work is different from other research works found in the literature. It is not based on software, utilisation ratio or heuristic algorithms. Rather, it is based on converting and generalising the concepts of live migration process and experimental results from other works, which are based on the aforementioned tools. The resulting model eventually converts the power cost of live migration from a function of utilisation ratio to a function of server PC. This means there will be neither a need for additional hardware, a separate software, nor a heuristics-based algorithms to measure the utilisation. The resulting model is simple, on the fly and accurate PC evaluation. Furthermore, the latency cost of live migration process, included the time it take the VM to be completely transferred to the target server, alongside the link distance/delay between the two servers is discussed.

Keywords: live migration, power model, power consumption, virtualisation, C-RAN, power cost

1. Introduction

The relentless growing number of connected mobile devices, along with the abundance of new types of bandwidth-hungry applications, has meant high data rate demands and a huge amount of signalling within the core network (CN). According to Cisco, mobile data traffic is expected to intensify by about 11-fold between 2013 and 2018. Furthermore, mobile device connections will grow to about 10.5 billion by 2018 compared to 7.2 billion in 2013 [1–3]. Additionally, Ericsson reportedly forecasts that in 2021, 150 billion devices will be 5G connected, up from 4.100 billion connections using LTE technology. These rising numbers are alarming and should urge mobile operators to seek out innovative ideas, designs, protocols and advanced digital signal processing (DSP) techniques in order to effectively cope with this explosively high demand for data while simultaneously providing scalable and faster connectivity.

In view of this demand, cloud radio access network (C-RAN) has been suggested by both mobile operators and equipment vendors to introduce cloud computing in 5G cellular networks by pooling the baseband processing units (BBUs) in a shared and centralised data processing centre, known as a BBU pool. In contrast to the legacy eNodeB, the main baseband physical procedures, cooperation and processing of the upper layers in C-RAN are executed in the BBU pool, whereas the simple radio frequency (RF) functions are tackled by the low power consumption (PC) of remote radio heads (RRHs). C-RAN therefore truncates the capital expenditures (CAPEX) and operational expenditures (OPEX) due to lower site leases, reduced maintenance cost and fewer site visits. Other benefits of C-RAN include (i) using advanced DSP, coordination and cognitive radio techniques to process signals through any neighbouring BBU(s) and efficiently utilising the available spectrum; (ii) managing traffic variations by exploiting fewer computing resources and, therefore, not utilising unwanted processors and (iii) reducing cooling PC as well as the total PC.

Recently, the C-RAN has witnessed a rising technology within the pool, which is live migrating from/to the servers. Basically, virtualising the servers has been an efficient way to increase the energy efficiency of the networks by running several BBUs' software using one server [4, 5].

Live migration is an agile concept in nowadays data centres [1]. In virtualisation environment, the interruption-free live migrating of the virtual machines (VMs) form one host server to another is an important issue to sustain running services to the UEs while gaining many benefits. Simply, live migrating of the VM is the movement of one or many VMs for the original host server to another server; this is done when the VM is still running even after it resides in its target server. It is called 'live' since the VM stayed running during the process of migration [4, 6]. Live migration comprises copying memory data on which the VM resides and CPU contents. Practically, an image file is stored in what is called network-attached storage (NAS) rather than the local disc. NAS is accessible by all VMs and operates as HDD drive [7]. This means that physically transferring the local disc is not required. More information about the process of copying a VM can be found in [1]. However, this technique has privileged the date centres with many advantages, as mentioned below:

- Maintenance: this technique represents a solution in case if the source server is required to be decommissioned due to its type promotion. Alongside, it urgently required operating system or hardware maintenance.
- Reachability: the VM usually resides on a host server which is physically located in a certain area. This VM might be serving UEs which are located far away from it. Alternatively, there might be host servers located closer to the UEs. Therefore, such migration will definitely reduce the link delay, channel losses and help improve system administration.
- Load balancing: there might be servers experiencing heavy load due to their position in a dense area or because of the service type they run. In this case, it is beneficial to distribute

http://dx.doi.org/10.5772/intechopen.72040

the load amongst other servers in the network via migrating the VMs. This is while proportionally considering their processing usage without degrading the performance of the participants [8].

- Off-loading: when the traffic in the network is low, some servers can be selected to switch off so as the network EE is increased. In this case, live migrating the VMs from the chosen servers to other active ones is the solution.

On the other hand, the performance of migration process depends on many factors, such as the memory allocated to the VM, the size of work load it serves and the transmission rate at which the migration is occurring. Eventually, these factors affect the latency of migration and network traffic flow [9]. However, there are increasing disadvantages symbolised by two major aspects:

- Time: the time it takes the process of migration degrades the network performance. The transferred VMs increase latency factor as it means more imposed link delay [10]; this delay means degraded coverage and lower network capacity. In real-time services, this factor is essential and crucial, in contrast to the off-line services where the latency is relieved.
- Energy: the overhead cost of live migration is considerable. Up to 10 W is withdrawn from the destination, and this value increases when the server is the source. This is on the basis of more computations within the tagged server which will be performed in one unit of time [11].

In this chapter, we will try to model these costs which are synchronised with this technology, aiming to speculate the cost of such process prior to migration. The structure of this chapter is as follows: in Section 2, the process of live migration is modelled. In Section 3, the results are presented. Finally, the summary is given in Section 4.

2. Live migration power model

The increased PC cost due to migrating a single or group of VMs to the destination server is also counted. It has been noted that the power cost of the source server $\left(P_{cost}^{source}\right)$ is changeable according to the utilisation of the CPU. We have translated this practical value to more understandable data to avoid the need to measurements' server. First, the extracted cost is redrawn as a function of sever utilisation (utl) of [11] instead of a function of downtime (latency) of migration. **Figure 1** shows the utilisation against the source server power cost.

The original data extracted is curve-fitted to a flexible and simple quadratic Eq. (2), with coefficients, cof1 = 0.0011189, cof2 = −0.25916 and cof3 = 16.315:

$$P_{cost}^{source} = cof1 * utl^2 + cof2 * utl + cof3 \tag{1}$$

However, this description is valid for a server with specific characteristics; there might be slight change in this curve when the type of server is changed. To cover this issue, Eq. (2) is

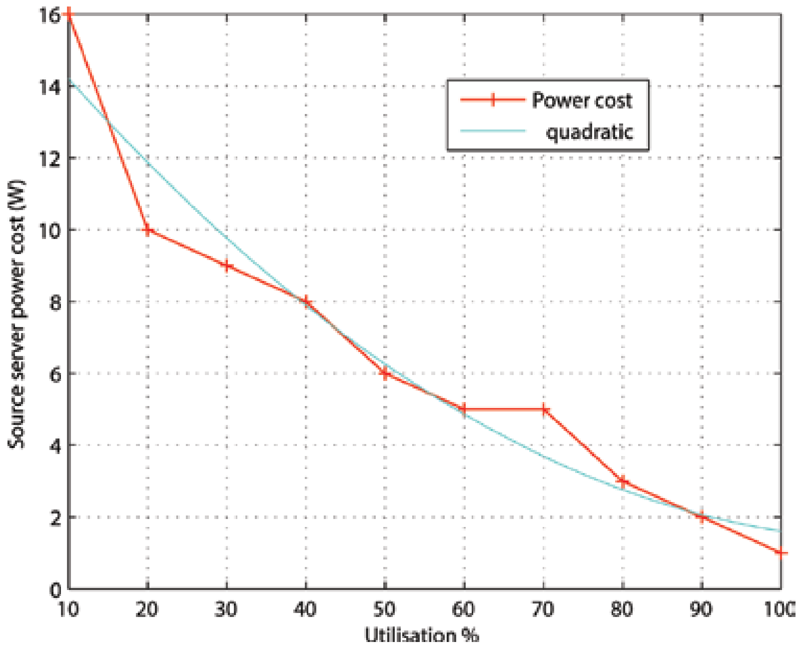

Figure 1. Utilisation against the source server power cost.

generalised by adding a constant called (*scof*), which is a real number. The latter will scale up and down the output of **Figure 1**, so as it matches all possible costs. Therefore, the model in Eq. (2) has been updated, as follows:

$$P^{source}_{cost} = \left(cof1 * utl^2 + cof2 * utl + cof3\right) + scof \tag{2}$$

There has been an objection with measuring the PC using utilisation ratio of a server; the reasons have been mentioned in [5]. Briefly, this method does not offer simplicity for many reasons: it requires real experiment or measurement; it is expensive, because it requires to be physically available at the data centre to observe the utilisation values. On top of that, the service provider (SP) propriety devices are not available to be tested. Therefore, this value is considered, but as a function of maximum and minimum PC of the virtualised server. This means that the real-time measurement power costs which are measured as a function of utilisation ratio are now transferred to a function of PC. To do so, we have used the formula in Eq. (3) to convert the utilisation of a server to a PC. In this case, the power cost of migration can be known directly from the server PC. For example, **Figure 2** shows the utilisation and PC conversions of a server:

$$utl = 100 * \left(P^{BBU}_{server} - min\left(P^{BBU}_{server}\right)\right) / \left(max\left(P^{BBU}_{server}\right) - min\left(P^{BBU}_{server}\right)\right) \tag{3}$$

where P^{BBU}_{server} is the BBU server.

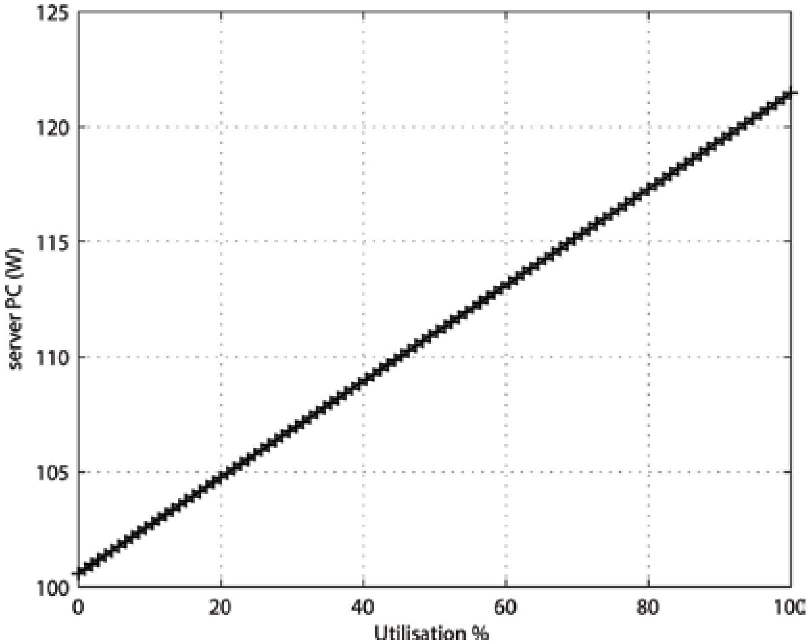

Figure 2. Utilisation ratio and its PC equivalent.

The 100% utilisation means that the server is fully utilised and experiences maximum PC $\left(\max\left(P^{BBU}_{server}\right)\right)$, while 0% utilisation means that the server is load-free $\left(\min\left(P^{BBU}_{server}\right)\right)$ or in idle mode of operation. The resulting PC and its equivalent power cost are shown in **Figure 3**.

In the receiving side, the power cost in the destination server $\left(P^{dest}_{cost}\right)$ during the migration is experimentally measured in [11], which was reported as 10 W, regardless of the percentage of utilisation in the server. This assumption is also generalised by adding another constant $(rcof)$, which is a real number, so as all possibilities of servers' specifications are considered, as follows:

$$P^{dest}_{cost} = 10 + rcof \tag{4}$$

The total number of source servers undergoing live migration is (S_{smig}), the total number of target servers is (S_{rmig}), the number of migrated VMs is (N_{smig}) and the total number of received VMs is (N_{rmig}). The overall PC of live migration-based virtualised C-RAN data centre (P^{mig}_{vCRAN}) is formulated by adding these costs to the total PC of virtualised C-RAN, which is found in [5]. This yields

$$P^{mig}_{vCRAN} = P_{vCRAN} + \sum_{s_{smig}}^{S_{smig}} \sum_{n_{smig}}^{N_{smig}} \left(P^{source}_{cost}\right)_{s_{smig},\, n_{smig}} + \sum_{s_{rmig}}^{S_{rmig}} \sum_{n_{rmig}}^{N_{rmig}} \left(P^{dest}_{cost}\right)_{s_{rmig},\, n_{rmig}} \tag{5}$$

In terms of time, the live-migrated downtime is relatively large; this process can happen within orders of milliseconds [12] or even orders of seconds. In any case, the higher data rate in which

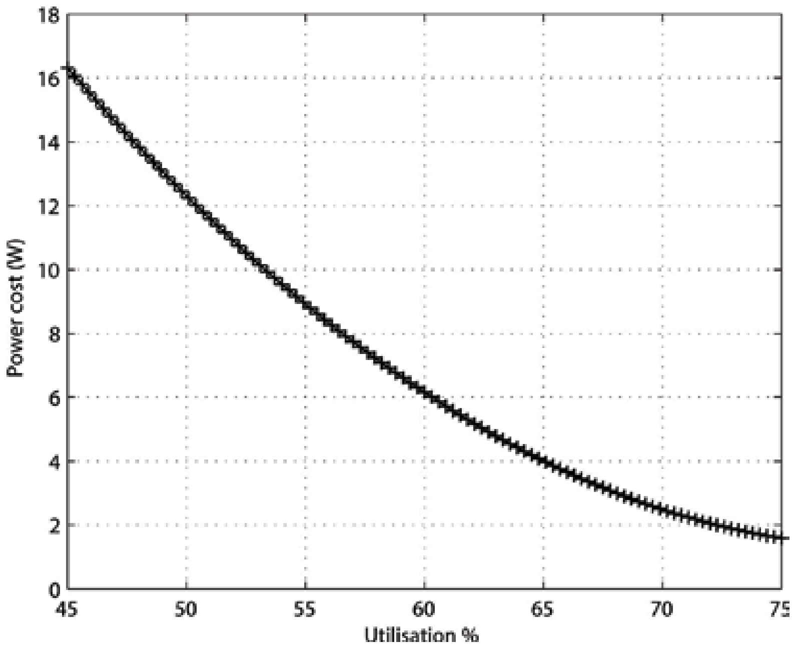

Figure 3. Power cost of live migration and its PC equivalent.

the VM is migrated, the less latency [9, 11]. The latter represents the time it takes the VM to be transferred to other servers seamlessly. As the UEs are still connected while migration, this value does not mean that it is an effective latency to be added to the virtualisation process latency. However, when migrating the VM, the time cost can be established from the channel in which the VM is transferred, which might be wireless or wired [11]. This cost eventually can be added to the modelling of [5].

The time cost due to migrating the VM is (τ_{mig}), which is equivalent to the ($\frac{d_{mig}}{v_{mig}}$), where d_{mig} and v_{mig} denote the distance and speed at which the VM is transferred from the source to target server. v_{mig} is based on the channel type; it can be as fast as the speed light if the channel is wireless, experiencing cable losses if the channel is ethernet via coaxial cable, or can be experiencing a refractive index in case of using an optical fibre channel. However, the effect of τ_{mig} can be major if the VM is moved between distant centres. Therefore, a virtualised data centre experiences a total delay (τ_T) due to virtualisation delay (τ_v) and migration delay τ_{mig}, where

$$\tau_T = \tau_v + \tau_{mig} \tag{6}$$

Approximately, the experimental results in [11] is similar to [9] in terms of power cost and migration time. The latter has measured the cost with respect to the transmission bit rate used to send the VM. While in the former, the power cost is associated with server utilisation. These two measurements are correlated as the highly utilised server means that it is sending using higher bit rate and vice versa. **Figure 4** shows the VM migration bit rate relation with the power cost and migration time.

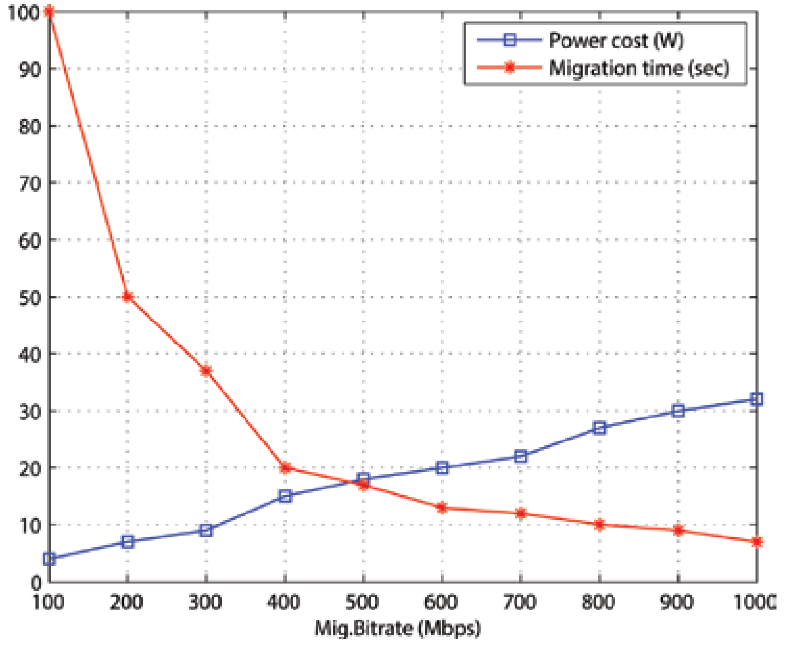

Figure 4. Migration bit rate as a function of power cost and migration time of [9].

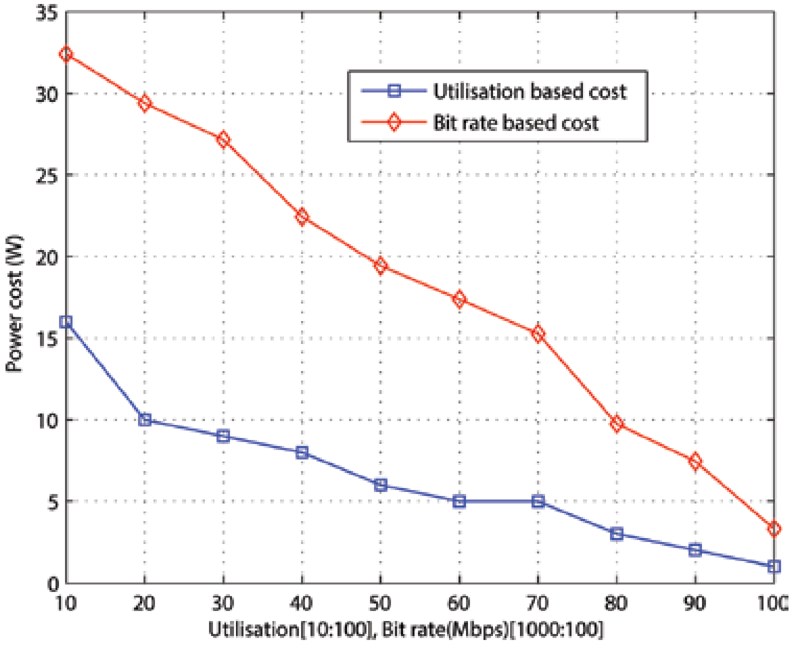

Figure 5. Difference between the outcomes of two experiments in terms of utilisation and bit rate.

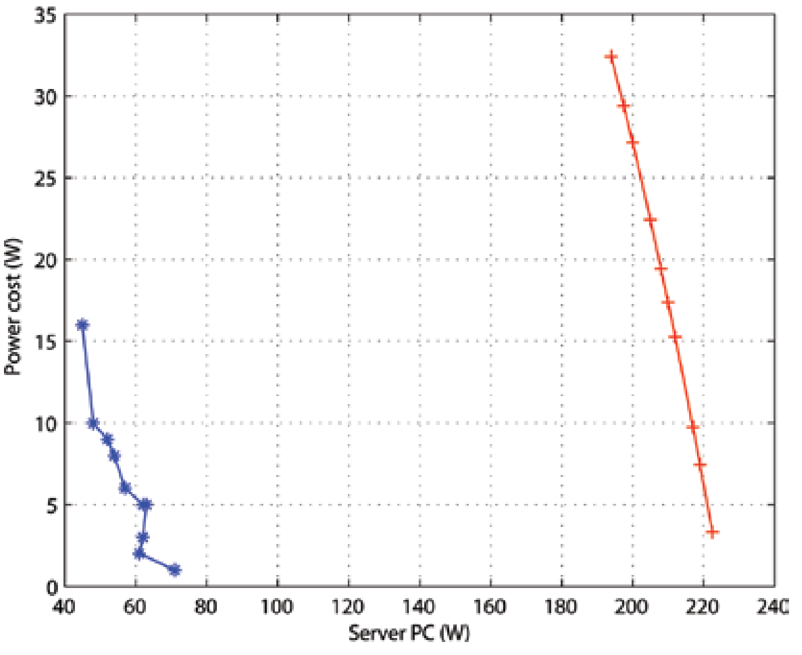

Figure 6. Difference between the outcomes of two experiments in terms of PC.

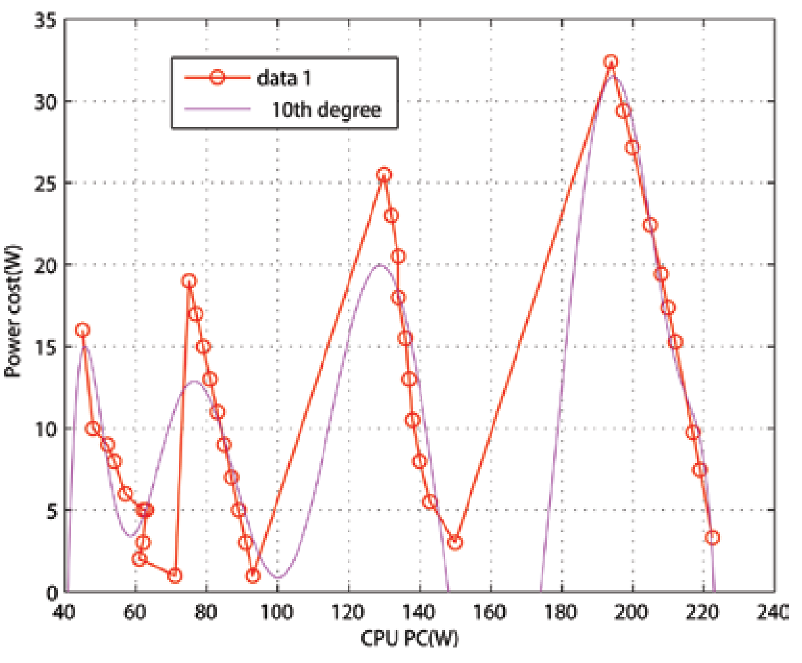

Figure 7. Difference between the outcomes of two experiments.

Figure 5 shows the differences between the outcomes of [9, 11] in terms of bit rate and utilisation, respectively. On the other hand, **Figure 6** shows the difference between the two works in terms of PC. However, this difference is originated from the characteristics of the two operated servers. Nevertheless, the behaviour of the outcomes is identical. To extend these models to a general formulation, another two data sets have been added for servers that are commercially available and consume from (130–150 W) to (75–93 W) [13]. Together, there will be four groups of data sets. These latter produce **Figure 7**, which is then poly-fitted using linear regression to produce the following model.

$$\begin{aligned} P_{cost}^{source} = p1 * P_{server}^{BBU\,10} + p2 * P_{server}^{BBU\ 9} + p3 * P_{server}^{BBU\,8} + p4 * P_{server}^{BBU\,7} \\ + p5 * P_{server}^{BBU\,6} + p6 * P_{server}^{BBU\,5} + p7 * P_{server}^{BBU\,4} + p8 * P_{server}^{BBU\,4} \\ + p9 * P_{server}^{BBU\,2} + p10 * P_{server}^{BBU} + p11 \end{aligned} \tag{7}$$

with coefficients: $p1$ = -8.6139e-17, $p2$ = 1.1482e-13, $p3$ = -6.6938e-11, $p4$ = 2.2424e-08, $p5$ = -4.7684e-06, $p6$ = 0.00067105, $p7$ = -0.063154, $p8$ = 3.9177, $p9$ = -153.09, $p10$ = 3399.7 and $p11$ = -32561.

Subsequently, the value of P_{cost}^{source} is updated.

3. Results

Figure 8 shows the power cost when a virtualised server holding 60 VMs is migrating 10 VMs; it also shows the power cost due to receiving the same number of VMs. Furthermore, it shows the total cost of both cases. This source cost was derived from the virtualised server consumption as mentioned above. At $N = 60$, the virtualised server PC is known. From this consumption, the power cost is obtained (about 4.9195 W) and multiplied by the number of migrated VMs. On the other hand, 10 W times the number of received VMs is the power cost in the receiving side.

Figure 9 shows the increasing cost because of the increasing number of migrated VMs, up to 50 VMs, while the virtualised server already hosted 60 VMs.

As we can see from **Figures 8** and **9**, migrating the VMs is not power-effective process. To migrate and receive 10 VMs, the power cost is about 150 W, which is more than the original server PC. However, this cost is influenced by the period at which these VMs are transferred. Intuitively, the longer period these VMs are migrated, the less power cost and more efficient system. Therefore, algorithms/methods are needed to optimise at what time the VM is required to be moved. This can be based on several parameters, such as the number of UEs connected, load balancing requirement, position of the servers, etc.

Figure 9 exhibits the cost and total PC when the number of both migrated and received VMs is equal. However, sending and receiving a VM by more than one server at the same time represent another facility that can be added and offered by the model. This is when the number of migrated and received VMs is different. **Figure 10** then shows a three-dimensional plot, and

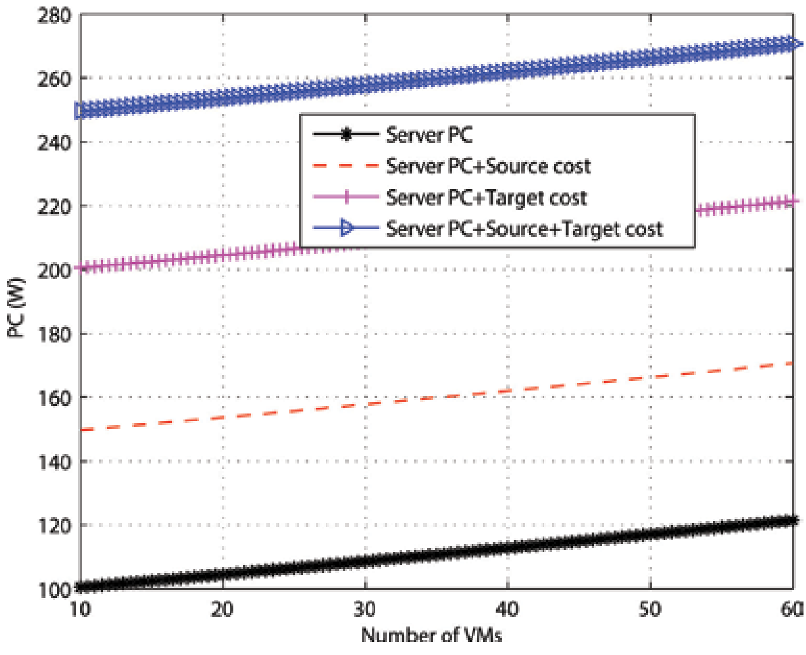

Figure 8. PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.

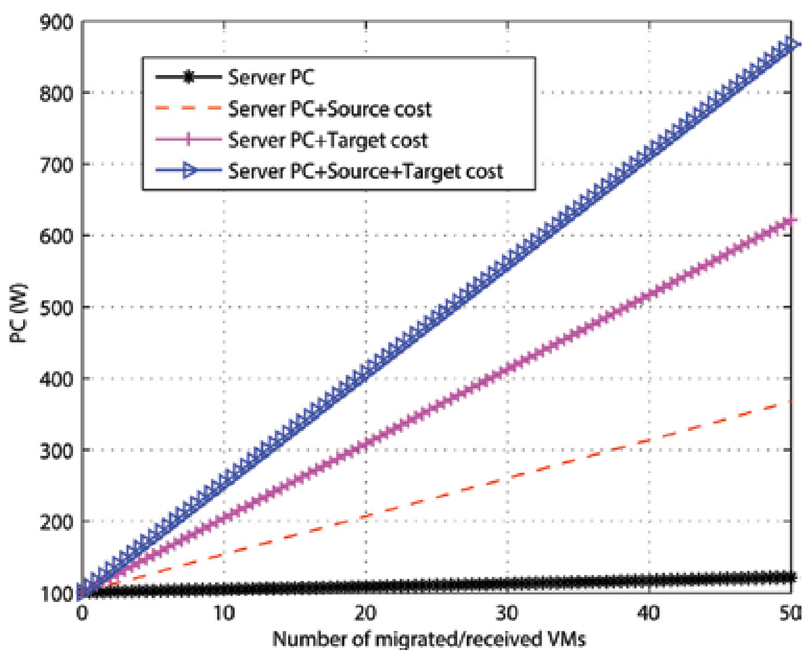

Figure 9. PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.

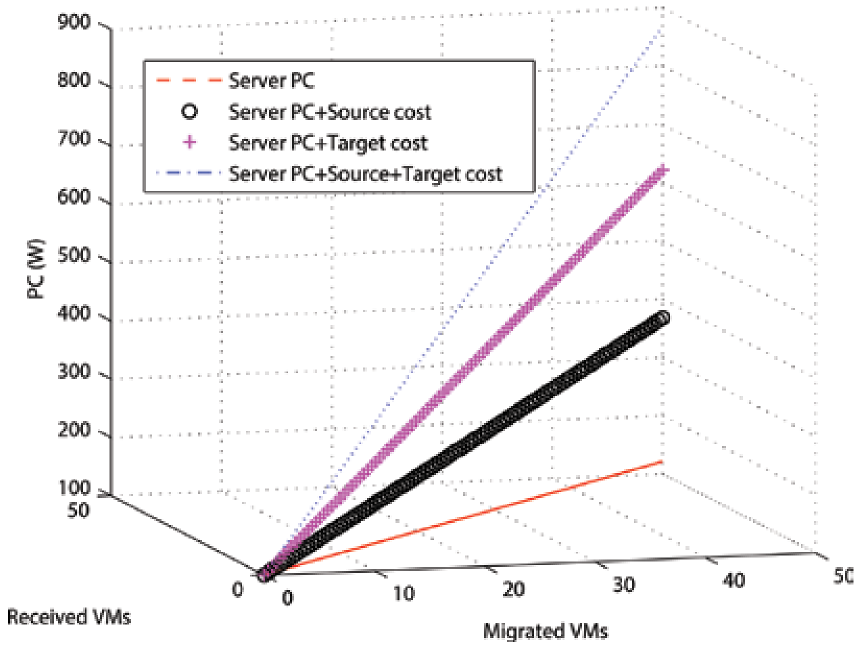

Figure 10. PC of source server hosting 60 VMs; it shows the power cost of migrating 10 VMs.

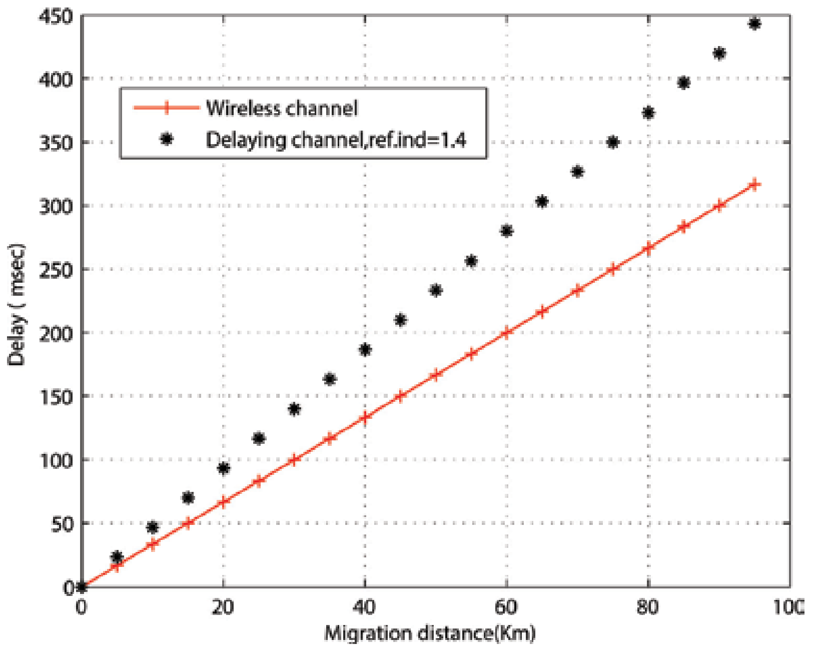

Figure 11. Delay comparison of wireless and delaying channel.

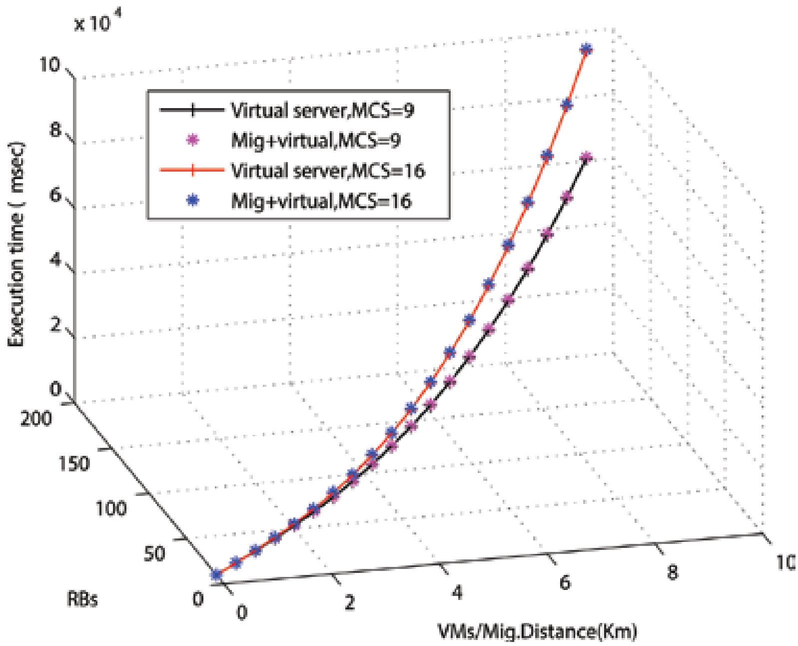

Figure 12. Total delay of virtualised server while migrating a VM to different distances.

x and y axes represent the number of migrated and received VMs, while z axis shows the PC. This figure shows the cost of migrating or receiving 50 VMs; also it shows the cost of both cases. The virtualised server holds 60 VMs.

Figure 11 shows the channel delay experienced due to migrating a VM to a distant data centre. This figure compares wireless with lossy channel with refractive index = 1.4.

However, in both cases, wireless or delaying channel, this amount of delay is neglected compared to the virtualisation process delay. **Figure 12** shows the total delay of virtualisation and migration. This covers the wireless, up to an optical fibre channel with refractive index which is equivalent to 1.4. This is when the virtualised server is hosting 10 VMs, each processing 200 RBs and migrating 1 VM to variable distances (up to 100 km).

4. Summary

A model has been presented to demonstrate the PC cost in a virtualised based data centres, specifically for live migration case. By using the proposed model, the power and time cost can be assessed. Since there is no simple expression to describe such costs, this model has been proposed. It has converted the experimental results which are based on the server utilisation and migration bit rate to a simple-to-adapt mathematical formulation. The model enables the network providers to decide about the EE of such technology in the data centres and virtualised servers. The cost of repeated migration can cause a considerable amount of power lost. Nevertheless, by using a quick and simple calculation of the model, a decision about when and where the migration occurs can be easily made. On the other hand, the expected time cost due to migration has been shown. However, the time cost is negligible compared to the virtualisation. This time was the cost of the distance between the two cooperated servers, while the time of the migration process itself can reach more than 7 s. This was not counted within this modelling since the VM is not terminated from serving the attached UEs and it is still live.

Author details

Raad S. Alhumaima*, Shireen R. Jawad and H.S. Al-Raweshidy

*Address all correspondence to: raad_1990_1990@yahoo.com

WNCC, School of Engineering, Design and Physical Sciences, Brunel University London, Uxbridge, Middlesex, UK

References

[1] Strunk A, Dargie W. Does live migration of virtual machines cost energy?. In: 2013 IEEE 27th International Conference on Advanced Information Networking and Applications (AINA); March 2013. pp. 514-521

[2] Alhumaima RS, Al-Raweshidy HS. Evaluating the energy efficiency of software defined-based cloud radio access networks. IET Communications. 2016;**10**(8):987-994

[3] Alhumaima RS, Khan M, Al-Raweshidy HS. Component and parameterised power model for cloud radio access network. IET Communications. 2016;**10**(7):745-752

[4] Zhang J, Ren F, Lin C. Delay guaranteed live migration of virtual machines. In: INFOCOM, 2014 Proceedings IEEE: IEEE; 2014. pp. 574-582

[5] Alhumaima RS, Al-Raweshidy HS. Modelling the power consumption and trade-offs of virtualised cloud radio access networks. IET Communications. 2017;**11**(7):1158-1164

[6] Zheng J, Ng TSE, Sripanidkulchai K, Liu Z. Pacer: A progress management system for live virtual machine migration in cloud computing. IEEE Transactions on Network and Service Management. December 2013;**10**:369-382

[7] Ruan Y, Cao Z, Cui Z. Pre-filter-copy: Efficient and self-adaptive live migration of virtual machines. IEEE Systems Journal. Dec 2016;**10**:1459-1469

[8] Agarwal A, Shangruff R. Live migration of virtual machines in cloud. International Journal of Scientific and Research Publications. 2012;**2**(6):1-5

[9] Liu H, Jin H, Xu C-Z, Liao X. Performance and energy modeling for live migration of virtual machines. Cluster computing. 2013;**16**(2):249-264

[10] Strunk A. Costs of virtual machine live migration: A survey. In: 2012 IEEE Eighth World Congress on Services. June 2012. pp. 323-329

[11] Huang Q, Gao F, Wang R, Qi Z. Power consumption of virtual machine live migration in clouds. In: Communications and Mobile Computing (CMC), 2011 Third International Conference on; April 2011. pp. 122-125

[12] Clark C, Fraser K, Hand S, Hansen JG, Jul E, Limpach C, Pratt I, Warfield A. Live migration of virtual machines. In: Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation - Volume 2, NSDI'05, (Berkeley, CA, USA), pp. 273-286, USENIX Association; 2005

[13] Buildcomputers, Power Consumption of PC Components in Watts (http://www.buildcomputers.net/power-consumption-of-pc-components.html), 2017 (accessed March 19, 2017)

Edited by Abdelfatteh Haidine and Abdelhak Aqqal

Nowadays, the Internet plays a vital role in our lives. It is currently one of the most effective media that is shifting to reach into all areas in today's society. While we move into the next decade, the future of many emerging technologies (IoT, cloud solutions, automation and AI, big data, 5G and mobile technologies, smart cities, etc.) is highly dependent on Internet connectivity and broadband communications. The demand for mobile and faster Internet connectivity is on the rise as the voice, video, and data continue to converge to speed up business operations and to improve every aspect of human life. As a result, the broadband communication networks that connect everything on the Internet are now considered a complete ecosystem routing all Internet traffic and delivering Internet data faster and more flexibly than ever before.

This book gives an insight into the latest research and practical aspects of the broadband communication networks in support of many emerging paradigms/applications of global Internet from the traditional architecture to the incorporation of smart applications. This book includes a preface and introduction by the editors, followed by 20 chapters written by leading international researchers, arranged in three parts. This book is recommended for researchers and professionals in the field and may be used as a reference book on broadband communication networks as well as on practical uses of wired/wireless broadband communications. It is also a concise guide for students and readers interested in studying Internet connectivity, mobile/optical broadband networks and concepts/applications of telecommunications engineering.

ISBN 978-1-83881-399-4

9 781838 813994

Published in London, UK

IntechOpen