IEEE 802 Wireless Systems

Protocols, Multi-Hop Mesh/Relaying, Performance and Spectrum Coexistence

Bernhard H. Walke | Stefan Mangold | Lars Berlemann

Companion Website

WILEY

IEEE 802 Wireless Systems

IEEE 802 Wireless Systems

Protocols, Multi-hop Mesh/Relaying, Performance and Spectrum Coexistence

Bernhard H. Walke

ComNets, RWTH Aachen University, Germany

Stefan Mangold

Swisscom Innovations, Switzerland

Lars Berlemann

ComNets, RWTH Aachen University, Germany

John Wiley & Sons, Ltd

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wiley.com

Other Wiley Editorial Offices

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

ISBN-13 978-0-470-01439-4 (HB)
ISBN-10 0-470-01439-3 (HB)

Typeset by 9.5/11.5pt Times by Integra Software Services Pvt. Ltd, Pondicherry, India
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

Contents

Preface

In the mid-1990s cellular mobile radio was the only means for public wireless access on a large scale, apart from some company-specific solutions used to a very limited extent in niche markets. Since the late 1980s, researchers and developers started investigations on developing broadband wireless Internet access technology for personal computers (PCs), mainly driven at that time by Apple Corporation, taking the European Telecommunication Standards Institute (ETSI) as a platform. The first standard appeared in 1993 (ETSI, 1993) and the final version in 1995 (ETSI, 1995) specified a Wireless Local Area Network (WLAN), nowadays called HiperLAN 1, see also Walke (2001). This WLAN system was the first standard to provide a completely distributed control, able even to support differentiation of services of stations competing for the medium access. But it failed to reach market acceptance owing to its high complexity resulting from a channel bandwidth at 5 GHz of about 20 MHz using single-carrier GMSK modulation. It was not the time for such a highly sophisticated system.

Overlapping this European activity, an initiative started in the US towards WLAN standardization in IEEE 802.11, aiming at broadband transmission in the 2 GHz band with a bit rate of 1 Mbit/s on a 1 MHz channel bandwidth (IEEE, 1997). Both standards were designed with an IEEE 802.2 or ISO 8802 compatible interface, suitable as a replacement for a Local Area Network (LAN).

In those days, Asynchronous Transfer Mode (ATM) had been introduced to wired (fiber-based) telecommunication networks and it was quite natural to think of a wireless extension of ATM. Accordingly, ETSI Broadband Radio Access Networks (BRAN) published its design of HiperLAN Type 2 (H/2) in 1997 (ETSI, 1997). At that time, IEEE 802.11 had already progressed. The final HiperLAN2 standards was completed in 2000 (ETSI, 2000).

Although H/2 with its OFDM-based 20 MHz wide channels, operating in the 5 GHz band had a much better throughput and delay performance than the WLAN specified in IEEE 802.11 (with its then 2 Mbit/s throughput capacity), the 802.11 made it to the market, owing to its simplicity and cost efficiency. From about 2002 on, wireless access, as well as cellular mobile access which in those days was mainly for voice communications, became widely accepted, and laptops started to have inbuilt WLAN access from 2003/4 onwards.

Meanwhile, IEEE 802.11 has greatly improved its performance and also covers operation in the 5 GHz band, using OFDM technology standardized earlier at ETSI BRAN for H/2. This technology has also been transferred to the 2.4 GHz band, where the first 802.11 systems started operation and still operate as a majority.

Today it is true that IEEE 802.11 standardized WLAN technology has reached worldwide acceptance for wireless short-range Internet access, and similar to what the GSM (Global System for

Mobile Communications), GPRS (General Packet Radio Service) ETSI-Standards have reached for wide-range mobile Internet access. The interworking of these standard systems (besides others) is currently being standardized at IEEE 802.21.

The success of IEEE 802.11 WLAN standards has also motivated initiatives to establish standards for Personal Area Networks (PANs) and Metropolitan Area Networks (MANs). And it is quite natural to also think of solutions on how to interwork these systems, not only on the network layer but also when operating on the same radio channel. Interworking of radio systems following different standards and coexistence of radio systems are big research themes today.

The hype for an ever-increasing bit rate capacity supported by XAN standardized systems has also motivated researchers and developers to consider multi-hop communication-based solutions to extend the radio range of a Radio Access Point (RAP) beyond the range that can be covered using the transmit power permitted by the regulator. A Mesh network is the key research theme applied to improve the range of coverage and reliability of wireless technologies for private and public access.

This book explores all these themes and related systems and proposals.

Since the early 1990s, my research group at RWTH Aachen University, Communication Networks (ComNets) have specialized in the development of services and protocols for private and public mobile radio systems. It has produced an extensive set of tools for software design, modeling and stochastic simulation of wireless and mobile radio systems. By means of these tools, the mobile radio and wireless systems now being used worldwide or under discussion, or in the process of being introduced, and described in this book, have been reproduced in a highly accurate form as large simulation program packages. These tools allow us to study the existing or forthcoming systems in their natural environments with the appropriate radio coverage and channel characteristics modeling, representing mobility and typical traffic volumes of the users, and, based on this, to test our own approaches to the improvements and introduction of new services and protocols. The proposals by ComNets and the results of our work have substantially influenced the standardization of systems such as ETSI-GSM/GPRS, ETSI H/2, IEEE 802.11e/s, 802.15, 802.16 and others, especially multi-hop Mesh networks and system coexistence.

This book has its focus on existent and forthcoming IEEE 802 wireless systems, including presentation of traffic performance. It would not have been possible to present a description of the systems with the desired degree of detail without having implemented the services and protocols of all of these systems in realistic models for stochastic, event-driven simulation.

The text and many of the figures in this book are based on the input of many Diploma and Master students at ComNets whose names it would not be possible to mention individually. All I can do is convey my gratitude to all of them for their enthusiasm and for the thoroughness of their work in this collaboration. Their contribution was in modeling and implementing as software the different systems and their modifications for performance evaluation, and their input has helped my research assistants and myself to develop a deep understanding of the characteristics of the systems considered.

The individual chapters of the book have been written in close cooperation with the research assistants responsible for the respective system models and they have been named. The chapters reflect the results of extensive research and development and in many cases incorporate material from PhD theses performed at ComNets. I would like to take this opportunity to give my warmest thanks to those involved for the thoroughness of their contributions on the respective topics, for their assistance in dealing with the relevant Diploma/Master's theses that they have supervised and for their role in creating such an excellent working atmosphere. I should particularly like to mention Stefan Mangold, who proposed writing this book and provided the nucleus of IEEE 802.11 work in Chapter 5 around which the book crystallized by integrating existing work based

on other PhD theses performed at ComNets. It turned out that Stefan was not able to take responsibility for establishing the book and Lars Berlemann stepped in to take the load of editing and harmonizing large parts of the book, in close cooperation with myself. Without his devotion, this book would never have come together.

Contributions to individual sections of the book have been made by the following members of the ComNets research group:

- Lars Berlemann (PhD), Chapters 1, 2, 3, 5, 8, 12 and 13
- Ingo Forkel (PhD), Chapter 2
- Guido Hiertz, Chapters 1, 2, 4 and 6
- Christian Hoymann (PhD), Chapters 2, 7 and 8
- Ole Klein, Chapter 12
- Stefan Mangold (PhD), Chapters 2, 5, 8 and 13
- Ralf Pabst, Chapter 10
- Daniel Schultz, Chapter 10
- Matthias Siebert (PhD), Chapters 5 and 11
- Erik Weiss, Chapter 4
- Harianto Wijaya (PhD), Chapter 8
- Yunpeng Zhang, Chapter 6
- Rui Zhao, Chapter 12

May this book prove useful to you!

Bernhard H. Walke
Stefan Mangold
Lars Berlemann
Aachen

1

Introduction

Bernhard H. Walke, Guido Hiertz and Lars Berlemann

Despite the promises of feature-rich, highly interactive and high bit-rate multimedia services of third-generation (3G) systems for end-users and increased revenues for the operators, the research community has perceived the limitations of these systems in terms of user throughput and cost of operation and consequently has started to work towards next-generation (NG) systems that are also being addressed as beyond third-generation (B3G) or fourth-generation (4G) systems. NG systems, which encompass B3G, 4G systems, and so on, are expected to allow subscribers to transparently access broadband multimedia services via multiple wireless and wireline access networks as if they are connected via broadband modems to the Internet.

1.1 Standardization

The NG mobile networks are expected to be introduced according to the roadmap shown in Figure 1.1 and are being developed according to the general guidelines and rough specifications of various organizations worldwide.

The Wireless World Research Forum published its Books of Vision in 2000, 2004 and 2006 (Tafazolli, 2005), which outlined technologies suited to NG systems, and the roadmap for their introduction.

The European Union is continuously updating its Information Society Technology (IST) Framework Program, which started in the early 1990s, and is funding a large number of research projects in the domains of broadband access, mobile technologies and services. Examples of these are WINNER, Ambient Networks, E2R and Smart Mobile Life that cooperate under the Wireless World Initiative.

The ITU-R approved Recommendation M.1645 as the baseline for all activities involved in the World Radio Conference 2007 (WRC-2007) for the potential identification of new spectra for mobile and wireless communication.

At the end of 2004 3GPP initiated a study on the further evolution of 3G systems with the major requirements on the radio interface of peak data rates up to about 100 Mbit/s with very low latency. The requirements are far below the ITU-R requirements for systems beyond 3G.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

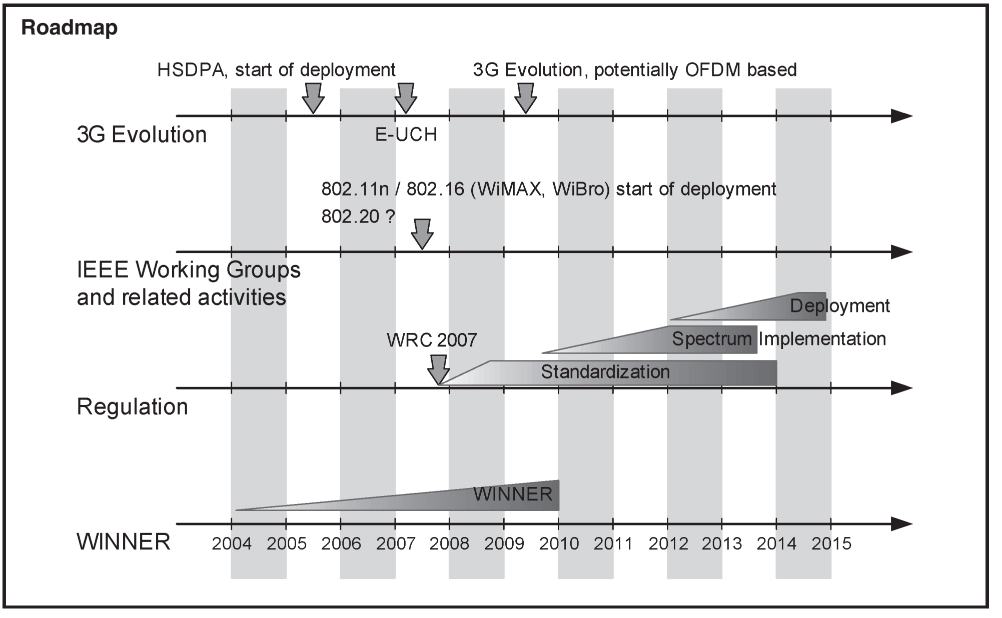


Figure 1.1 Roadmap according to the plans by 3GPP, IEEE, ITU-R and IST-WINNER. Reproduced by permission of © 2005 IEEE[1].

In Japan, the mobile IT Forum (mITF) was established in 2001 to realize future mobile communications systems and services such as the 4G system and mobile commerce services. mITF published roadmaps towards 4G in 2001 and 2004. The National Institute of Communication Technology (NiCT), a public research organization, runs the New Generation Mobile Network Project, which is developing technologies for B3G systems. NTT DoCoMo has developed a system proposal for systems beyond 3G and a demonstrator.

In Korea the Next Generation Mobile Committee (NGMC) is developing guidelines for 4G services, 4G spectra and technologies, and is coordinating 4G technology-related activities in Korea. ETRI of Korea is driving a national program on wideband mobile communications (Wireless Broadband – WiBro) to be deployed during the year 2006.

The Chinese government launched a national research project, Future Technologies for Universal Radio Environment (FuTURE), in the framework program 863 in the area of mobile communications for the time frame of the tenth five-year plan 2001–2005. Project phase 2 is planned up to 2010 and aims to achieve international leadership in mobile communications. The FuTURE project has focused on the development of radio transmission technologies, which satisfies the future needs for the time frame 2005–2010.

In the United States, the IEEE, a globally operating organization with its headquarters in the US, has developed standards for wireless local area networks and is now expanding that work to drive the development of future systems with extended capabilities. These alternative radio standards including activities on interworking and the support of mobility management may have an impact on 3G and its enhancements as well as NG systems. Another major player in North America is the Defense Advanced Research Project Agency (DARPA). The DARPA neXt Generation (XG) communications program is developing the technology to allow multiple users to share use of the spectra through adaptive mechanisms.

The wireless IT sector specifies radio interface systems on the lower layers in IEEE Project 802. In addition, industry standards and proprietary solutions are being developed for special applications. All these standards only focus on parts of the system. Their integration in existing networks is not yet solved. From a capabilities point of view these systems do not fulfill the requirements on systems beyond 3G with respect to supported throughput, latency, mobility and flexibility.

1.2 Next-generation Systems

NG systems according to Figure 1.1 will be shaped by emerging standards for wireless access from mobile terminals, like “3G evolution” and IEEE 802 wireless systems evolution, e.g., 802.11n (“a quarter Gigabit WLAN”), 802.16e (“mobile WiMax”, cf. WiBro), 802.16j (Mobile Multi-hop Relaying, MMR), 802.20 (Mobile Broadband Wireless Access, MBWA) and 802.21 (for handover and interoperability between heterogeneous network types including both 802 and non-802 networks). Figure 1.1 also shows the ITU-R regulation activities represented by WRC-2007, which is expected to specify new bands where NG systems will operate. In view of IST-WINNER, in 2012 the spectrum implementation will start permitting deployment of NG systems from 2014 on.

The main characteristics of NG systems are seamless use of a multitude of existing and forthcoming access network types by mobile terminals (not a replacement of predecessor systems by a new design). Evolution towards NG mobile networks is already taking place through the interworking of 3G cellular systems such as UMTS/HSDPA, cdma2000 and GSM/EDGE with other broadband wireless access technologies such as Wi-Fi hotspots and the forthcoming WiMax metropolitan networks to provide mobile users with anytime anywhere access to broadband multimedia services over the Internet, as shown in Figure 1.2.

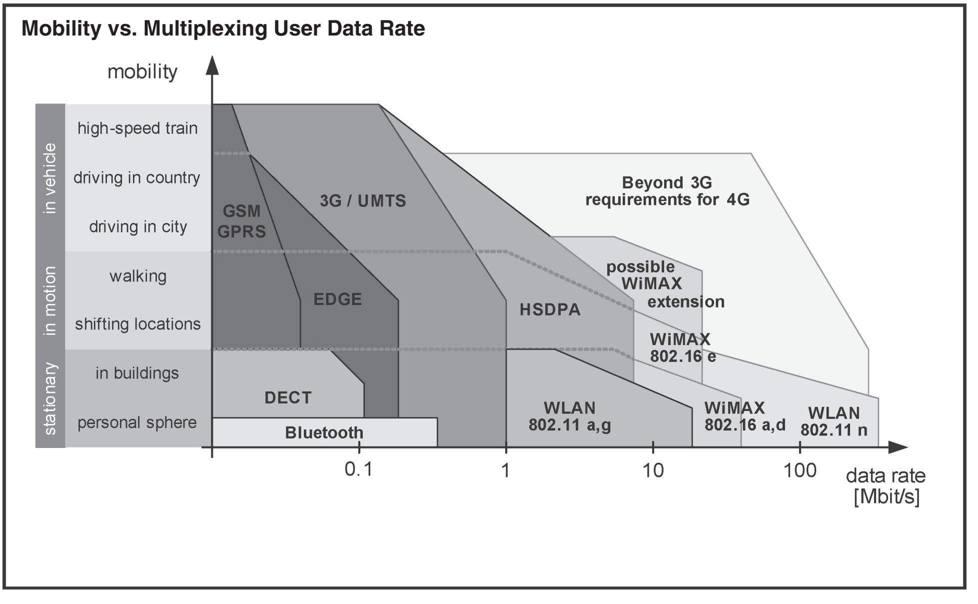


Figure 1.2 Mobility versus multiplexing user data rate of various mobile and wireless systems. Reproduced by permission of © 2005 IEEE[1].

The NG mobile networks will also be shaped by emerging standards for wireless access from mobile terminals, e.g., IEEE 802.20, and for the integration of heterogeneous wireless access networks, e.g., IEEE 802.21, which will establish interworking of IEEE wireless and wireline systems and any other mobile systems. This type of interworking may result in a converged global network and will provide many technical challenges.

Owing to the heterogeneity in the wireless access techniques and the requirements of multimedia and VoIP services, QoS provisioning and enforcement in the future generation mobile networks will become a demanding issue for wireless access equipment vendors, networks service providers and customers.

Multi-carrier based transmission and spatial multiplexing will be the dominating transmission technologies of future broadband networks. Cross-layer protocol engineering along with application-oriented signaling and creation of new services will be required for efficient network resource utilization and to bring profitability to the service providers.

1.3 The IEEE 802 Project

In 1963 the “American Institute of Electrical Engineers (AIEE)” and the “Institute of Radio Engineers (IRE)” merged to become the “Institute of Electrical and Electronics Engineers (IEEE)”. Today, the IEEE is a leading authority for standards, research and scientific exchange in the fields of aerospace, computers, telecommunications, biomedical engineering, electric power, consumer electronics, and many more.

As of 2006, IEEE has more than 365,000 members in over 150 countries. IEEE plays an important role not only for the industry but also for academia. More than 1430 student branches at colleges and universities in 80 countries prove IEEE’s presence in the research community. In 2006, IEEE published over 128 transactions, journals and magazines. Currently, IEEE sponsors more than 300 international conferences worldwide. Out of 900 active, the IEEE 802 LAN/MAN IEEE Standards Committee is one of the most well known.

Authorization for formation of Project IEEE 802 was requested in December 1979. Since then, “The IEEE 802 LAN/MAN Standards Committee (LMSC) develops Local Area Network standards and Metropolitan Area Network standards (IEEE, 2006a).” Approved on March 13, 1980, LMSC was formed by the IEEE Computer Society. Today, 22 Working Groups (WGs) in IEEE 802 consider mainly the lowest two layers of the ISO/OSI reference model. Currently, the following WGs are active:

- 802.0, Wireless Coordination Activity Group
- 802.1, Higher Layer Local Area Networks (LAN) Protocols
- 802.3, Ethernet
- 802.11, Wireless Local Area Networks (WLAN)
- 802.15, Wireless Personal Area Networks (WPAN)
- 802.16, Wireless Metropolitan Area Networks (WMAN)
- 802.17, Resilient Packet Ring
- 802.18, Radio Regulatory Technical Advisory Group (TAG)
- 802.19, Coexistence TAG
- 802.20, Mobile Broadband Wireless Access (MBWA)
- 802.21, Media Independent Handoff
- 802.22, Wireless Regional Area Networks (WRAN)

Inactive WGs hibernate until their reactivation is demanded. They have standards published, but do not work or hold meetings, currently:

- 802.2, Logical Link Control (LLC)
- 802.5, Token Ring
- 802.12, Demand Priority

Disbanded WGs have not published standards or their standards have been withdrawn:

- 802.4, Token Bus
- 802.6, Metropolitan Area Network
- 802.7, Broadband TAG
- 802.8, Fiber Optic TAG
- 802.9, Isochronous LAN
- 802.10, Security
- 802.14, Cable Modem Working Group
- QOS/FC, Executive Committee Study Group

Each of the WGs can have further subgroups. Such subgroups are referred to as Task Groups (TGs). Participants of an IEEE 802 WG may initiate the formation of a Study Group (SG), to study new problems and future extensions. Similar to a TG, the lifetime of an SG is limited. Its main task usually is the development of two documents, the Project Authorization Request (PAR) and the 5 Criteria (5C), which are used to propose the establishment of new TGs. The documents are reviewed by the New Standards Committee (NesCom), which is a body independent of the 25 IEEE societies and councils, LMSC being an example council. NesCom decides about the proposal. For decision making, the 5C used are market potential, compatibility, distinct identity, technical and economic feasibility. The PAR is used to define the charter of a new project or TG and gives the scope of the proposed standard or amendment.

1.4 Motivation and Outline

In this book, we intend to describe and evaluate the main wireless standards that have been created during the last years. We further discuss and propose improvements and regulatory constraints, which are currently targeted at the various standardization groups, and will lead to the next generation systems of the future.

Some fundamentals of wireless communication are reviewed in Chapter 2. A detailed introduction to radio transmission and medium access is given to form a basis for the common understanding of our topic of interest.

The regulation of the radio spectrum is discussed in Chapter 3. Manifold regulation bodies and global institutions are summarized followed by a discussion on approaches to spectrum regulation.

Chapter 4 gives an insight into the basics of wireless mesh networking. Challenges because of radio propagation and specific routing problems are highlighted.

WLANs of the IEEE 802.11 standard and its enhancements are introduced in Chapter 5. Initially, the structure of the 802.11 WG is described. The Medium Access Control (MAC) according to the 802.11 base standard is discussed together with an overview on the different Physical Layer (PHY) supplements for operating in frequency bands at 2.4 and 5 GHz. The capability to support QoS with the means of the 802.11e amendment is examined with the help of simulation.

The ongoing standardization efforts of the IEEE in the context of WPANs in the WG 802.15 are summarized in Chapter 6. The main focus is thereby on high-speed data WPANs of 802.15.3. Different proposals for the PHY and MAC layers are discussed rounded up through a simulation analysis of a candidate MAC proposed by the WiMedia Alliance.

A comprehensive description of IEEE 802.16 WMANs is given in Chapter 7. After introducing the scope of 802.16 and its system concept an overview on the structure of the 802.16 WG and their organization for standardization is given. The multi-carrier PHY of 802.16 based on OFDM and the MAC layer, with a focus on the TDD mode, is described in detail thereafter. Finally, the performance of 802.16 is evaluated by analytic approximation and simulation.

Approaches to wireless mesh networks in IEEE and industry are introduced in Chapter 8. Extensions of IEEE 802.11 and IEEE 802.16 for multi-hop operation to realize homogeneous multi-hop networks are discussed followed by an approach to heterogeneous multi-hop networks of 802.16 and 802.11.

The coexistence of 802 wireless networks is discussed in Chapter 9 from the perspective of medium access control. The realization of homogeneous coexistence, i.e., spectrum sharing of wireless networks of the same technology/standard, is introduced at the example of 802.11e WLANs. Further, the heterogeneous coexistence is addressed in proposing methods that allow the unlicensed operation of 802.16 in a shared spectrum with 802.11.

Chapter 10 gives an overview on and motivation for broadband cellular multi-hop networks to stimulate standardization activities of the IEEE 802 WGs. The introduced relay-based deployment concept founds the basis for standardizing a mesh enhancement of 802.16.

The integration and cooperation of complementary radio access networks with a special focus on mobility and handover procedures is highlighted in Chapter 11.

Chapter 12 reflects major drawbacks of state-of-the-art MAC protocols and proposes a solution for the architecture and protocols of a Mesh Distributed Coordination Function (MDCF) for interconnecting a large number of APs.

Cognitive radio and spectrum sharing as approaches to overcome the current scarcity and waste of radio spectrum are discussed in Chapter 13. A detailed introduction and definition of cognitive radio is given together with a comprehensive overview on flexible spectrum access and spectrum sharing.

This book is summarized in Chapter 14 with a conclusion.

As an extra resource, the companion website contains slides of the figures and tables ready for download for course presentation. Also, a complete implementation of a protocol emulator is available for download to help facilitate an understanding, visualization and performance evaluation of IEEE 802.11. Please go to www.ieee802-wireless-systems.com.

Note

1. Reproduced by permission of © 2005 IEEE. Source: Wireless World Initiative New Radio – WINNER 2005.

2

Wireless Communication – Basics

Bernhard H. Walke, Lars Berlemann, Guido Hiertz, Christian Hoymann, Ingo Forkel and Stefan Mangold

This section gives a tutorial-like description of the fundamentals of wireless communication. Additionally, literature is referenced that deepens the scope of this chapter. The aim of this section is to provide a basis for understanding the rest of the book. Topics that are important in the context of the wireless network technologies discussed in this book are therefore described in more detail.

2.1 Radio Transmission Fundamentals

In mobile radio systems, unlike wired networks, electromagnetic signals are transmitted in free space (see upper part of Figure 2.1). Therefore a total familiarity with the propagation characteristics of radio waves is a prerequisite in the development of mobile radio systems. In principle, the Maxwell equations explain all the phenomena of wave propagation. However, when used in the mobile radio area, this method can result in some complicated calculations or may not be applicable at all if the geometry or material constants are not known exactly. Therefore stochastic models were developed to determine the characteristics of radio channels, and these consider the key physical effects in different environments. The choice of model depends on the frequency and range of the radio waves, the characteristics of the propagation medium and the antenna arrangement.

The propagation of electromagnetic waves in free space is extremely complex. Depending on the frequency and the corresponding wavelength, electromagnetic waves propagate as ground waves, surface waves, space waves or direct waves. The type of propagation is correlated with the range, or distance, at which a signal can be received (see lower part of Figure 2.1). The general rule is that at a given transmit power the higher the frequency of the wave to be transmitted, the shorter the range reached.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

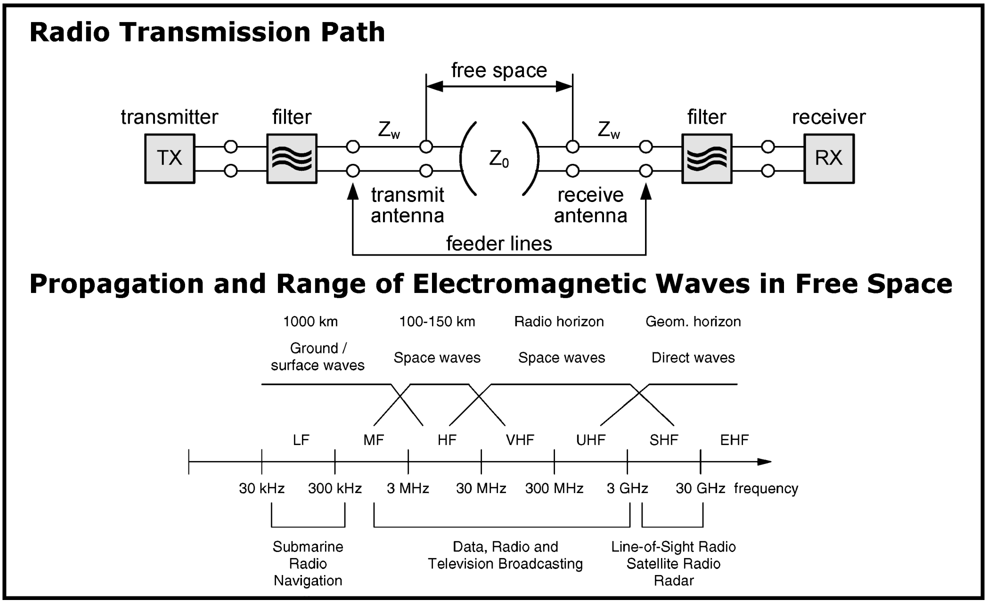


Figure 2.1 Radio transmission path: transmitter–receiver. z_0 and z_w are the radio wave resistances in free space and on the antenna feeder link, respectively (above). Propagation and range of electromagnetic waves in free space (below).

Based on the curvature of the earth, waves of a lower frequency, i.e., larger wavelength, propagate as ground or surface waves. These waves can still be received from a great distance and even in tunnels.

At the higher frequencies it is usually space waves that form. Along with direct radiation, which, depending on the roughness and the conductivity of the earth's surface, is quickly attenuated, these waves are diffracted and reflected based on their frequency in the troposphere or in the ionosphere.

The range for lower frequencies lies between 100 and 150 km, whereas it decreases with higher frequencies because of the increasing transparency of the ionosphere, referred to as the radio horizon. When solar activity is intense, space waves can cover a distance of several thousand kilometers owing to multiple reflections on the conductive layers of the ionosphere and the earth's surface.

Waves with a frequency above 3 GHz propagate as direct waves, and consequently can only be received within the geometric (optical) horizon.

2.1.1 Free-space Propagation

Another factor that determines the range of electromagnetic waves is their power. The field strength of an electromagnetic wave in free space decreases in inverse proportion to the distance to the transmitter, and the receiver input power therefore fades with the square of the distance. The received power for omnidirectional antennas can be described on the basis of the law of free-space propagation.

An ideal point-shaped source, a so-called isotropic radiator of signal energy, transmits its power P_0 uniformly in all directions Θ. The constant spatial power density is $P_{iso} = P_0/4\pi$. In this isotropic case the power density flow F through a sphere with radius d is

$$F = \frac{P_0}{4\pi d^2}\left[W/m^2\right].$$

In the normal case an antenna transmits the main part of the power P_T (index T: Transmitter) in preferred directions (main- and minor lobes). The antenna gain G_T puts this in relation to the isotropic radiation. The product EIRP $= P_T G_T = P_0$ is called EIRP (effective isotropically radiated power).

An antenna with G_T, which transmits in the mean the total power P_0, transmits into the direction Θ the power density

$$P_{TX} = \frac{4\pi}{P_0 G_{TX}}$$

The corresponding power flow density (power per unit area) through a sphere with radius d is

$$F = \frac{P_T G_T}{4\pi d^2}$$

The power P_R (Index R: Receiver) an antenna can take from the electromagnetic waves is the product of F and the effective antenna area which can be expressed as follows by the wavelength λ and the gain G_R of the receiver antenna

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi d}\right)^2$$

The term $(\lambda/4\pi d)^2$ is referenced as free-space pathloss because it describes the spatial diffusion of the transmitted energy over the path of length d.

In a logarithmic representation the difference $P_T - P_R$ corresponds to an expression $-10\log P_R/P_T$. In this representation the free-space loss L_F results (with $c = \lambda f$) in

$$L_F = -10\log(G_T) - 10\log(G_R) + 20\log(f) + 20\log(d) - 20\log\left(\frac{c}{4\pi}\right).$$

In the case of an isotropic antenna the last expression reduces to

$$L_F = -20\log\left(\frac{\lambda}{4\pi d}\right) = -20\log\left(\frac{P_R}{P_T}\right). \tag{2.1}$$

2.1.2 Two-path Propagation Over Flat Terrain

Free-space propagation is of little practical importance in mobile communications, because in reality obstacles and reflective surfaces will always appear in the propagation path. Along with attenuation caused by distance, a radiated wave also loses energy through reflection, transmission and diffraction due to obstacles.

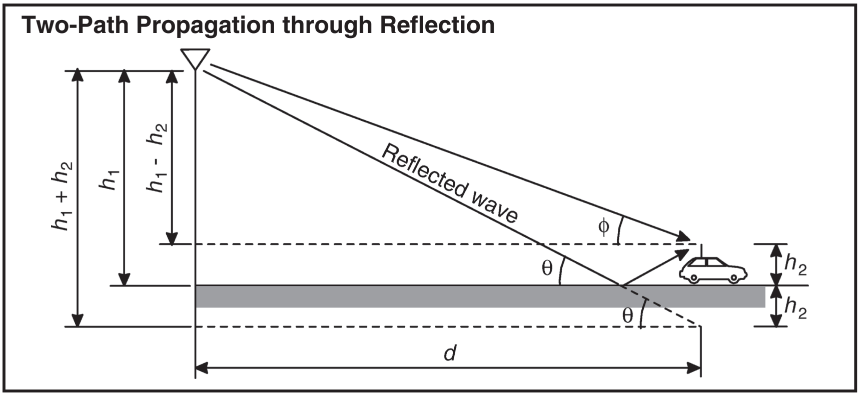


Figure 2.2 Model for two-path propagation due to reflection.

A simple calculation (Parsons, 1992) can be carried out for a relatively simple case scenario: two-path propagation over a reflecting surface (see Figure 2.2). In this case

$$P_R = P_T G_T G_R \left(\frac{h_1 h_2}{d^2} \right)^2$$

$d \gg h_1, h_2$ is a frequency-independent term. The corresponding pathloss L_P is

$$L_F = -10 \log (G_T) - 10 \log (G_R) - 20 \log (h_1) - 20 \log (h_2) + 40 \log (d)$$

and with isotropic antennas

$$\frac{L_F}{dB} = 120 - 20 \log \left(\frac{h_1}{m} \right) - 20 \log \left(\frac{h_2}{m} \right) + 40 \log \left(\frac{d}{km} \right)$$

In this model the receive power decreases much faster ($\sim 1/d^4$) than with free-space propagation ($\sim 1/d^2$). This also depicts the reality of a mobile radio environment more closely but does not take into account the fact that actual ground surfaces are rough, therefore causing wave scattering in addition to reflection. Furthermore, obstacles in the propagation path and the type of buildings that exist have an impact on attenuation.

With the introduction of the pathloss coefficient γ, the following applies to isotropic antennas:

$$P_R = P_T G_T G_R \left(\frac{\lambda}{4\pi} \right)^2 \frac{1}{d^\gamma}$$

Realistic values for γ are between 2 (free-space propagation) and 5 (strong attenuation, e.g., because of city buildings).

2.1.3 Attenuation

Weather conditions cause changes to the atmosphere, which in turn affect the propagation conditions of waves. Attenuation is frequency-dependent and has a considerable effect on some

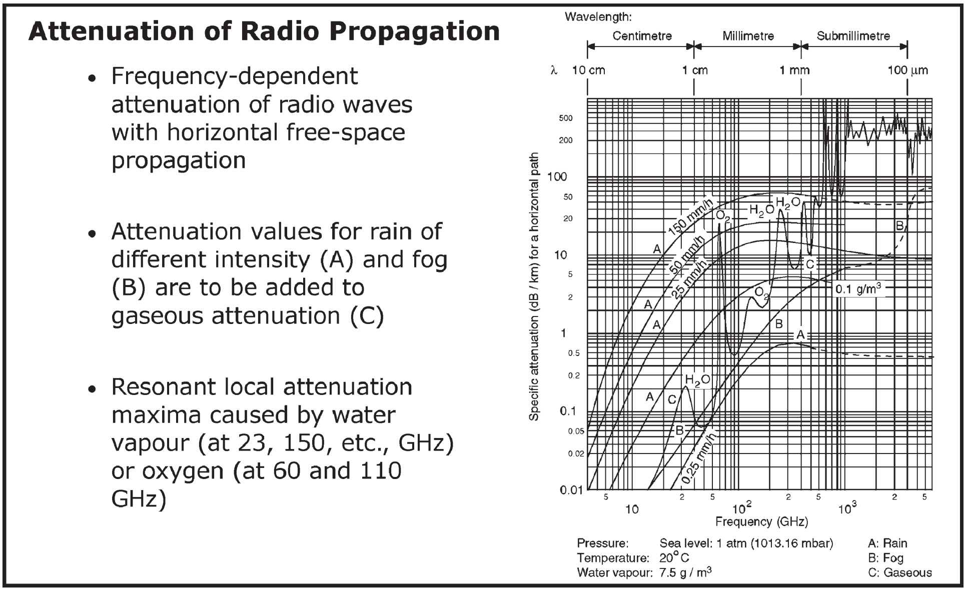


Figure 2.3 Attenuation of radio propagation depending on the frequency due to gaseous constituents and precipitation for transmission through the atmosphere (from CCIR Rep. 719, 721).

frequencies, and a lesser one on others. For example, in the higher-frequency ranges above about 12 GHz attenuation is strong when it is foggy or raining because of the scattering and absorption of electromagnetic waves on drops of water.

Figure 2.3 shows the frequency-dependent attenuation of radio waves with horizontal free-space propagation in which, as applicable, the appropriate attenuation values for fog (B) or rain of different intensity (A) still need to be added to the gaseous attenuation (curve C). What is remarkable are the resonant local attenuation maxima caused by water vapor (at 23, 150, etc., GHz) or oxygen (at 60 and 110 GHz).

2.1.4 Fading

Fading refers to fluctuations in the amplitude of a received signal that occur owing to propagation-related interference. Multipath propagation caused by reflection and the scattering of radio waves lead to a situation in which transmitted signals arrive phase-shifted over paths of different lengths at the receiver and are superimposed there. This interference can strengthen, distort or even eliminate the received signal. There are many conditions that cause fading, and these will be covered below.

In a realistic radio environment waves reach a receiver not only over a direct path but also on several other paths from different directions (see Figure 2.4). A typical feature of multipath propagation (frequency-selective with broadband signals) is the existence of drops and boosts in level within the channel bandwidth that sometimes fall below the sensitivity threshold of the receiver or modulate it beyond its linear range.

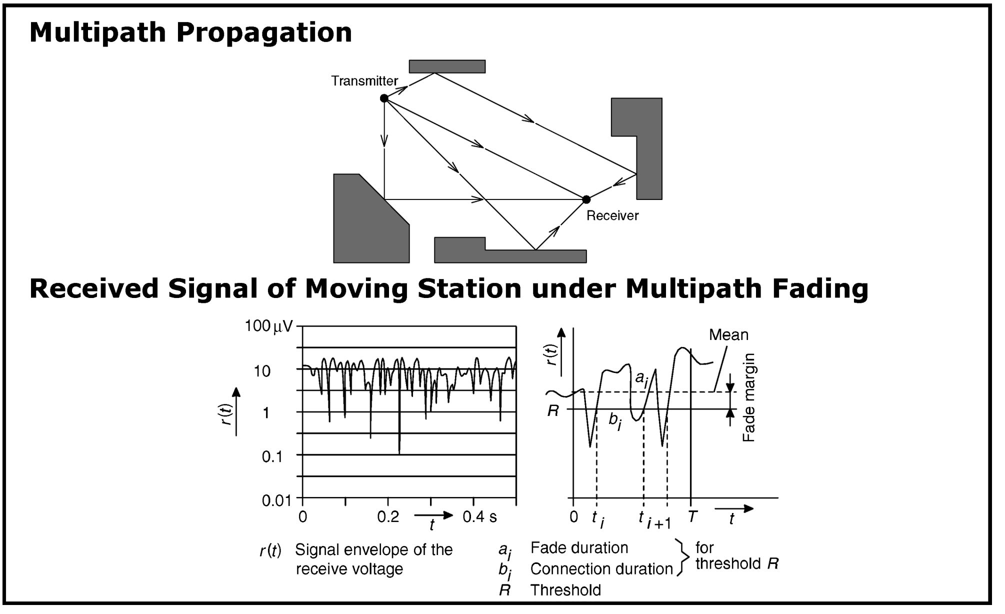


Figure 2.4 Multipath propagation and received signal voltage of moving station under multipath fading.

The individual component waves can thereby superimpose themselves constructively or destructively and produce a stationary signal profile, referred to as multipath fading, which produces a typical signal profile on a path when the receiver is moving, referred to as short-term fading, see Figure 2.4.

The different time delays of component waves result in the widening of a channel's impulse response. This dispersion (or delay spread) can cause interference between transmitted symbols (intersymbol interference).

Furthermore, depending on the direction of incidence of a component wave, the moving receiver experiences either a positive or a negative Doppler shift, which results in a widening of the frequency spectrum.

In general the time characteristics of a signal envelope pattern can be described as follows:

$$r(t) = m(t)r_0(t)$$

Here $m(t)$ signifies the current mean value of the signal level and $r_0(t)$ refers to the part caused by short-term fading. The local mean value $m(t)$ can be deduced from the overall signal level $r(t)$ by averaging $r(t)$ over a range of 40–200 λ (Lee, 1982).

2.1.5 Shadowing

Obstacles in the Line-of-Sight (LOS) path between transmitter and receiver outdoors (mountains and buildings) or inside buildings (walls) hinder direct wave propagation and therefore prevent the use of the shortest and in most cases least interfered (strongest) path between transmitter

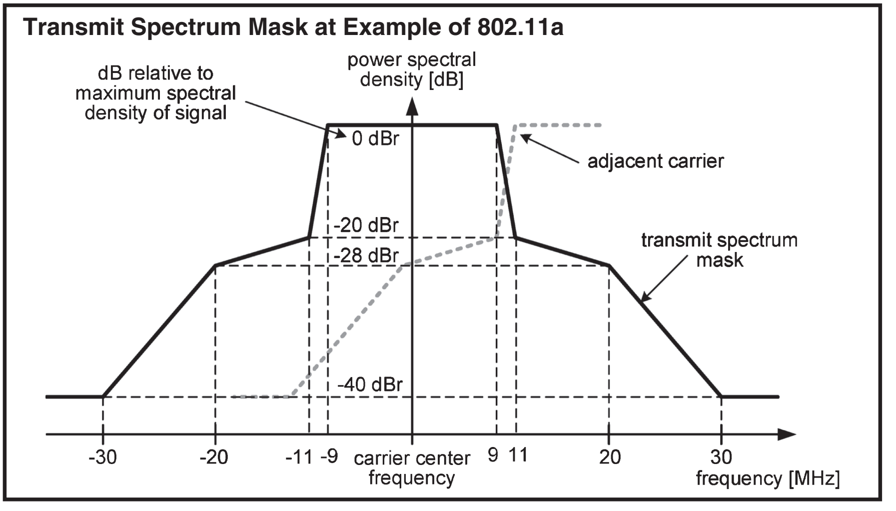


Figure 2.5 Limiting transmission power through spectrum masks (IEEE, 2003a).

and receiver, and cause additional attenuation to the signal level, which is called shadowing. Long-term fading occurs when a moving receiver is lingering for a long time in the radio shadow, e.g., for 10 to 40 s.

2.1.6 Filtering and Transmit Spectrum Masks

The 802 wireless communication standards define so-called spectrum masks for limiting their transmitter's emitted power in frequency spectrum. The transmit spectrum mask defines the maximal allowed power spectral density in dBr, i.e., dB relative to the maximum spectral density of a signal. As an example, the transmit spectrum mask of 802.11a (IEEE, 2003a) is depicted in Figure 2.5. As 802.11 does not specify a transmit filter function, filtering is implied in the spectrum mask.

The dashed gray mask indicates potential interference from an adjacent carrier. The considerable overlapping of the spectrum mask of neighboring carriers leads to an essential interference between stations transmitting on adjacent carriers. The frequency band at 2.4 GHz is for instance divided in the US into 11 channels and into 13 channels in Europe. The center frequencies of these channels have only 5 MHz in between with an overall channel bandwidth of 22 MHz. Besides this limitation from the center frequency of +11 MHz/ − 11 MHz, a transmitter may emit power in a frequency band +30 MHz/ − 30 MHz around the channel's center frequency. For an interference-free operation, typical 802.11 deployments therefore use only three nonoverlapping, independent channels (in the US with the numbers 1, 6, and 11).

2.1.7 Propagation Models

The propagation models described in the following are confined to pathloss prediction in macro-cell scenarios with a typical cell radius of several km, micro-cell environments (typical cell radius up to 1 km) and pico-cellular indoor scenarios (typical cell radius up to 500 m). A comprehensive survey of the different models is given in (Forkel and Salzmann, 2001; Garg, 2000; Laiho *et al.*, 2002). The simplest propagation situation is free-space propagation as introduced in Section 2.1.1: The attenuation is related to the distance d between transmitter and receiver. Obstacles interfering with the radio wave are not taken into account.

2.1.7.1 One-slope Model

In the one-slope model pathloss is determined by the logarithmic distance d and the power decay index γ. Typical values are $\gamma = 2$ for free-space propagation as in Equation (2.1) and $\gamma = 3.5 \ldots 6$ in buildings. Further values are presented in Andersen *et al.* (1995). It is recommended to fit γ in with the particular scenario. According to COST (1999) the one-slope model is expressed by

$$L = L_0 + 10\gamma \cdot \log_{10} \frac{d}{[m]}$$

with:

L_0: free space loss in 1 m distance $\left(L_0 = 20 \cdot \log_{10} \frac{4\pi}{\lambda/[m]}\right)$,
γ: power decay index (pathloss coefficient),
d: distance between transmitter and receiver.

2.1.7.2 Hata–Okumura Model

Based on measurements conducted at different frequencies and various environments, several correction terms have been introduced for the one-slope model. For mobile land radio systems in the 400 MHz and 900 MHz bands, Okumura *et al.* (1968) performed measurements in the area of Tokyo. On the basis of this work, Hata (1980) proposed the well-known Hata–Okumura model by

$$\begin{aligned} L = {} & 69.55 + 26.16 \cdot \log_{10} \frac{f}{[MHz]} - 13.82 \cdot \log_{10} \frac{h_b}{[m]} - a\,(h_m) \\ & + \left(44.9 - 6.55 \cdot \log_{10} \frac{h_b}{[m]}\right) \cdot \log_{10} \frac{d}{[km]} \end{aligned}$$

with:

f: carrier frequency
h_b: base station antenna height
$a(h_m)$: correction factor for mobile antenna height
d: distance between transmitter and receiver

Finally, 3GPP (1998) defines the vehicular test environment pathloss model applicable to scenarios in urban and suburban areas outside the high-rise city core, where buildings are of nearly uniform height. With this model path loss can be calculated by

$$L = 80 + 21 \cdot \log_{10} \frac{f}{[MHz]} - 18 \cdot \log_{10} \frac{\Delta h_b}{[m]} + \left(40 - 0.16 \frac{\Delta h_b}{[m]}\right) \cdot \log_{10} \frac{d}{[km]},$$

where f is the carrier frequency and Δh_b is the base station antenna height above the average roof top level that may range from 0 to 50 m.

2.1.7.3 Walfish–Ikegami Model

To accommodate radio propagation above roof tops, it is also necessary to calculate the pathloss according to the commonly known Walfish–Ikegami (or COST 231) model as described in COST (1999). This model has been used extensively in typical suburban and urban environments where the building heights are quasi-uniform. The model utilizes the theoretical Walfish–Bertoni model developed in Walfish and Bertoni (1998) to obtain multiple screen forward diffraction loss for high base station antenna heights, whereas it uses measurement-based data for low base station antenna heights. Additionally, the model takes into account free-space loss, loss due to diffraction down to the street, and the street orientation factor.

Steep transitions of pathloss occur when the base station antenna height is close to the roof top heights of the buildings in its vicinity. Therefore, the height accuracy of the base station antenna will be especially significant if large prediction errors are to be avoided. Moreover, the performance of the Walfish–Ikegami model is poor when the base station antenna height is significantly lower than the roof top heights of adjacent buildings.

Consequently, pathloss above roof tops follows

$$L = L_F - 16.9 + 10 \cdot \log_{10} \frac{W}{[m]} + 10 \cdot \log_{10} \frac{f}{[MHz]} + 20 \cdot \log_{10} \frac{\Delta h_m}{[m]} + L(\varphi) + L_{ms},$$

where

W: street width
L_F: free-space loss as defined in Equation (2.1)
f: carrier frequency
Δh_m: average roof top height above mobile antenna
L_{ms}: multi-screen loss considering the geometry of the environment

and

$$L(\varphi) = \begin{cases} -9.646 & for \quad 0^\circ \leq \varphi < 35^\circ, \\ 2.5 + 0.075(\varphi - 35) & for \quad 35^\circ \leq \varphi < 55^\circ, \\ 4 - 0.114(\varphi - 55) & for \quad 55^\circ \leq \varphi \leq 90^\circ, \end{cases}$$

includes an additional loss depending on the angle of incident φ relative to the street.

2.1.7.4 Dual-slope Model

Micro-cell measurements indicated that pathloss behaved differently at close and distant ranges (Feuerstein *et al.*, 1994). Therefore a dual-slope model as presented in Beyer and Jakoby (1997) with

$$L_1(d) = 10\gamma_1 \cdot \log_{10}\left(\frac{4\pi d}{\lambda}\right) - a_0 \quad for \quad d < d_{Br},$$

$$L_2(d) = L_1(d_{Br}) + 10\gamma_2 \cdot \log_{10}\left(\frac{d}{d_{Br}}\right) \quad for \quad d \geq d_{Br},$$

where

λ: signal wavelength
d_{Br}: breakpoint distance
γ_1: power decay index before d_{Br}
γ_2: power decay index beyond d_{Br}
a_0: difference between real loss and L_F at $d = 1\,\mathrm{m}$ as in Equation (2.1)

is more appropriate. The parameter a_0 considers e.g. wave-guiding effects and varies between 0 dB and 5 dB. In the near region, the slope γ_1 is set close to 2. In the far region the slope γ_2 is usually steeper with values up to 6.

2.1.7.5 Berg Model

For outdoor pathloss calculations along streets surrounded by buildings that are considerably taller than the height of the antennas, the so-called Berg model is proposed in Berg (1995) and introduced in the following.

Along a straight street s_0 the signal seems to originate from the actual transmitter location. The received signal strength along a perpendicular street behaves as if the wave originates from a virtual transmitter located in the proximity to the street crossing. Between these two extreme cases the model provides a continuous path loss as a function of the angle φ.

Similar to Equation (2.1) pathloss is determined by

$$L = 20 \cdot \log_{10} \frac{4\pi d_l}{\lambda}, \tag{2.2}$$

where l is the number of single street elements along the path from transmitter to receiver and wavelength λ. The "illusory" distance d_l is defined by the recursive expression

$$d_0 = 0, d_m = k_m \cdot s_{m-1} + d_{m-1} \quad with \quad l \geq m > 0,$$
$$k_0 = 1, k_m = k_{m-1} + d_{m-1} \cdot q_{m-1} \quad with \quad l \geq m > 0.$$

In Figure 2.6 the solid line describes the track of a propagating wave along an $l = 3$ segment path. The value s_m is the physical distance between nodal points, d_m is the illusory distance from the transmitter to the m-th nodal point, and l is the number of sections between the nodal points included in the calculation. The parameter q_m determines the angle dependency of the pathloss. As a simple approach to model the angle dependency of q_m, Berg proposes the expression

$$q_0 = 0, \quad q_m(\varphi_m) = \left(q_{90^\circ} \cdot \frac{\varphi_m}{90^\circ}\right)^{\nu} \quad with \quad l > m > 0,$$

where ν determines the shape of the function. Possible parameters are $q_{90^\circ} = 0.5$ and $\nu = 1.5$.

Figure 2.6 shows an example pathloss plot calculated for three different paths in a Manhattan-like urban scenario. Path one is an LoS path, path two contains a crossing after 165 m and path three bends after 57 m.

As mentioned in Section 2.1.7.4 the distance dependence of pathloss generally shows a dual-slope behavior. Due to this fact Equation (2.2) is extended to

$$L = 20 \cdot \log_{10} \left[\frac{4\pi d_l}{\lambda} D\left(\sum_{m=1}^{l} s_{m-1} \right) \right],$$

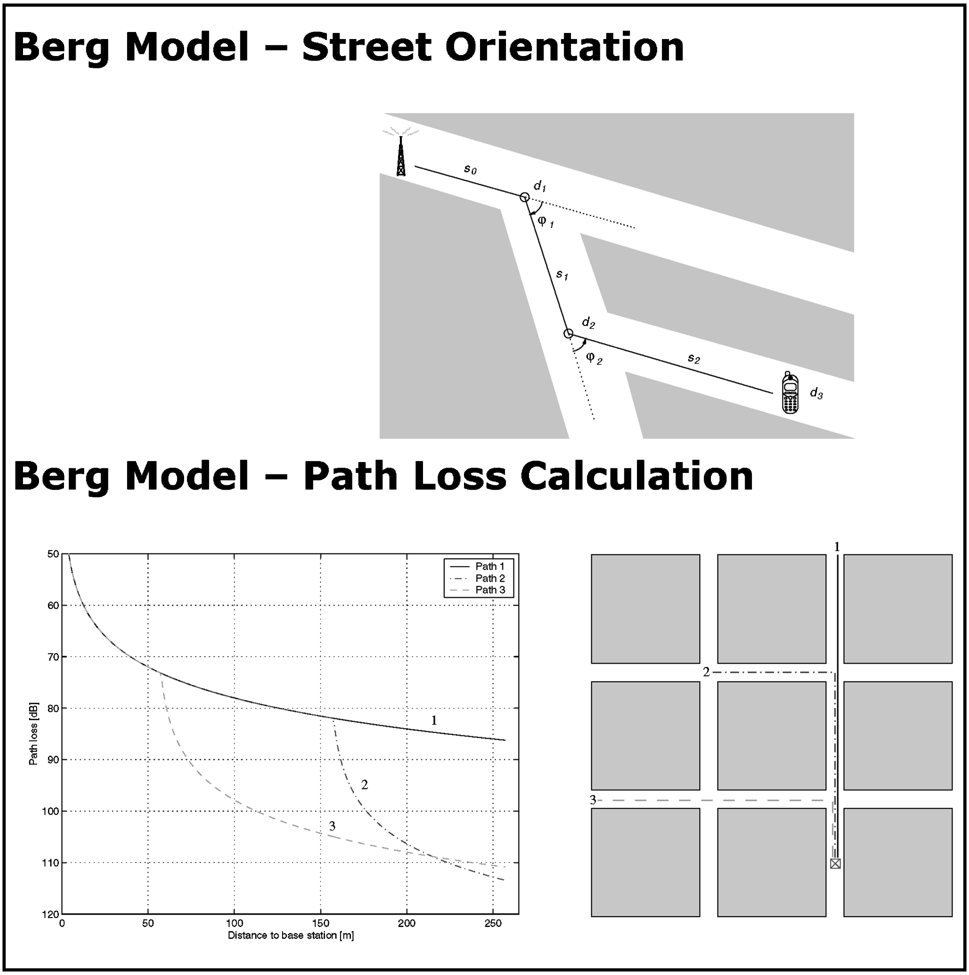


Figure 2.6 Street orientation example and pathloss calculation with the Berg model.

$$D(x) = \begin{cases} \dfrac{x}{d_{Br}} & for \quad x > d_{Br}, \\ 1 & for \quad x \leq d_{Br}. \end{cases}$$

2.1.8 Signal-to-Interference Ratio (SIR)

If the power of a received signal exceeds $P_{min}(R)$ then the receiver can decode it and retrieve information. However, $P_{min}(R)$ is not the only requirement a wireless signal needs to meet for successful reception. As the wireless medium is shared, other simultaneous radio transmissions may occur. They have impact on the possibility of successful radio reception. To successfully receive a radio transmission, the power level of the wanted signal must be several degrees higher than any other interfering signal arriving at the receiver at the same time. The ratio between power of the wanted signal and any unwanted transmissions is called the Signal-to-Interference Ratio (SIR). Let N denote the amount of devices and d_n, $n \in (1 \ldots N)$ represents a single device. Any emission of power of device d_m, $m \neq n$ on the wireless channel can be received at d_n at a level of $P_n(m)$. However, $P_n(m)$ may fluctuate and therefore depends on time, $P_n(m, t)$. For any device d_n the amount of power received at time t_0 is given by

$$P_{total,n}(t_0) = \sum_{m=1}^{N} P_n(m, t_0)$$

with $m \neq n.P_{total,n}(t_0)$ denotes the total sum of power received at d_n. For simplicity, we assume stationary conditions. Therefore, we do not consider time dependency in the following.

Let *SIR* denote the signal-to-interference ratio. It is defined as

$$SIR_n(m) = P_n(m) / \sum_{k=1}^{N} P_n(k), k \notin (n, m),$$

where d_n receives a transmission of device d_m while devices d_k concurrently emit unwanted power to the wireless medium. For any data transmission from device d_n to d_m that uses PHY rate R, signal reception solely succeeds if $P_n(m) > P_{min}(R)$ and $SIR_n(m) > SIR_{min}(R)$.

2.1.9 Noise – An Additional Source of Interference

Assuming a single transmission at a time, *SIR* would always be infinite. However, in real wireless system noise must be taken into account. It is constantly present. Noise does not depend on the existence of other simultaneous transmissions. Since it is unwanted power at the receiver it can be treated as interference. Noise is of thermal, galactic and atmospheric nature for example. The first one is the dominant source to consider. Thermal noise is an inherent effect on any wireless receiver. The term signal-to-noise ratio (SNR) $SNR_n(m) = P_n(m)/P_{n,noise,thermal}$ denotes the ratio between the wanted power $P_n(m)$ that is received by device d_n when device d_m transmits and the power of noise. The latter one can be calculated by $P_{n,noise,thermal} = k_B \cdot T \cdot \Delta f$, where k_B denotes Boltzmann's constant for a random process, T denotes the temperature and Δf is the bandwidth of the receiver under consideration. For more realistic assumptions, an additional receiver noise figure N_f may be added as constant amount – $P_{n,noise,thermal} = N_f + k_B \cdot T \cdot \Delta f$.

2.1.10 Signal to Interference and Noise Ratio (SINR)

Instead of its separate presentation, the signal to interference and noise ratio (SINR) expresses the overall ratio of the wanted signal to any other unwanted power at the receiver side:

$$SINR_n(m) = \frac{P_n(m)}{P_{n,Noise,thermal} + \sum_{k=1}^{N} P_n(k)}, k \notin (n, m).$$

It is the most important value to consider when discussing frame reception success probability. In the optimum case of a single transmitter and a single receiver with no other harmful transmission at the same time, SINR equals SNR. Hence, thermal noise is the lower boundary that limits transmission range.

2.1.11 Interference Range

Any other concurrent emission of radio power may add to the interference at the receiver side and thus reduce the SINR. SINR summarizes the overall signal reception conditions. As it is reciprocally dependent on

$$\sum_{k=1}^{N} P_n(k),$$

it is determined by the amount of other concurrent transmissions, the distance of interferers to the receiver, the power level of the interferers and propagation path conditions from each interferer to the receiver side. Hence, the interference range of a device transmitting simultaneously to other devices is not fixed and cannot be precisely determined. However, for a simple scenario consisting of a receiving device d_n, a single transmitting device d_m and a single interfering device d_k, a useful definition of the interference range may be estimated as follows. The receiver d_n is out of interference range of d_k if the distance between the receiver and the interferer is greater than $r_{interference}$. The latter value is defined as

$$r_{interference}(n,k) = \sqrt[\gamma]{\frac{P_k}{N_f(n) + k_B{}^{*}T(n)^{*}\Delta f(n)} \left(\frac{\lambda}{\pi 4}\right)^2}$$

under the assumption, that the transmission power P_k of device d_k reaches the receiving device d_n at a level that is equal to $P_{n,noise,thermal}$. Hence, $r_{interference}$ represents the distance from the location of the interferer where its power level equals the unavoidable thermal noise in the receiver. Although the distance between a receiving device d_n and an interfering device d_k equals $r_{interference}$, signal reception could be disturbed if signals of the transmitting device d_m are received close to a level of $SINR_{min}(R)$. Hence, $r_{interference}$ represents a practical definition with no means of an uninterrupted frame reception guarantee. However, with a standard IEEE 802.11 channel bandwidth ($\Delta f = 20\,\text{MHz}$), normal temperature ($T = 293.15\,\text{K}$) and $N_f = 5\,\text{dBm}$ a conservative assumption delivers $P_{n,noise,thermal} = -95\,\text{dBm}$. Supposing the latter value for noise power, e.g. in most IEEE 802.11 WLAN scenarios a distance of $r_{interference}$ between a receiver and an interferer is sufficient to avoid unwanted collisions at any time.

Under the assumption of free-space conditions and a constant pathloss coefficient γ, a theoretical definition of $r_{interference}$ can be derived. As obstacles (walls, human beings, doors, windows etc.) influence the wireless propagation path and introduce attenuation, emissions of multiple interferers sum up at the receiver side and channel conditions vary over time, the calculation of $r_{interference}$ presented here, provides a rough rule of thumb only.

2.1.12 Digital Modulation

In digital wireless communication systems a discrete clock is used to exchange symbols of fixed duration. The modulation of a carrier signal in one or more domains such as frequency, phase and amplitude provides the ability to carry multiple bits within such a symbol. Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) are for example simple and robust modulation schemes. One symbol modulated at BPSK carries 1 bit and QPSK carries 2 bits. More sophisticated modulation schemes such as Quadrature Amplitude Modulation (QAM) may carry 4 bits (16-QAM), 6 bits (64-QAM) or 8 bits (256-QAM) in each symbol. Figure 2.7

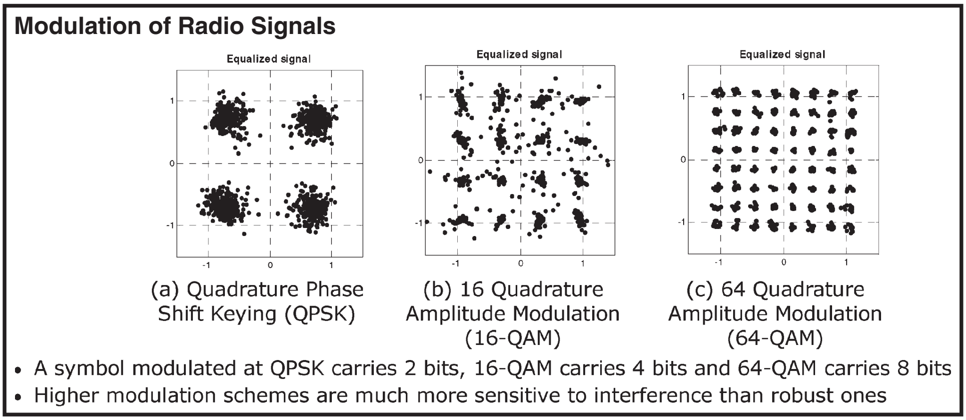


Figure 2.7 Modulation of radio signals at the example of QPSK and 16 and 64 QAM.

illustrates the modulation of QPSK, 16-QAM and 64-QAM. Today's wireless communication systems support multiple modulation schemes: BPSK, QPSK, 16-QAM and 64-QAM are for instance the modulation schemes specified by the IEEE 802.16 standard.

However, the more bits each symbol carries, the more complex the detection process is. Thus, for successful symbol reception the received power needs to be higher than the power detection level. Thus, the reception range is smaller than the power detection level. The minimum power level for reception depends on the modulation scheme. In the example of Figure 2.7, 16-QAM is for instance much more sensitive to interference since four times more symbols than with QPSK need to be discerned. $P_{min}(R)$ denotes the minimum power that is needed to decode a signal sent at the PHY rate R.

Figure 2.8 shows the mapping of bits to complex symbols being composed of the Inphase (I) and Quadratur (Q) component of a 16-QAM constellation diagram. The amplitude and phase of the complex symbol become the amplitude and phase of the corresponding subcarrier in the Inverse Fast Fourier Transform (IFFT). The amplitude of the constellation point is normalized to achieve equal average power (factor c in Figure 2.8). The first point is modulated onto the data subcarrier with the lowest frequency offset.

2.1.13 Modulation and Coding of Radio Signals

As wireless communication is unreliable, systematic redundancy is added to allow for error detection or even Forward Error Correction (FEC). A coding rate of 1/2 indicates that 2 bits encode 1 bit of information. Without redundancy, no reliable data communication would be possible since a single-bit failure would invalidate a message built up from multiple bits. Modulation and coding rate together form a specific PHY mode. For the sake of simplicity, the PHY mode R denotes in the following not only a modulation scheme but also a combination of modulation and coding rate referred to as Modulation and Coding Scheme (MCS). For each PHY mode, a specific $P_{min}(R)$ and $SINR_{min}(R)$ must be met to successfully receive a frame, exemplary values for the PHY of 802.11a are given in Figure 2.9.

The FEC options are paired with the modulation schemes to form burst profiles, i.e., PHY modes of varying robustness and efficiency. The possible PHY modes are listed at the example of the MCS available in 802.16 in Figure 2.9. The demodulator at the receiver side can convert

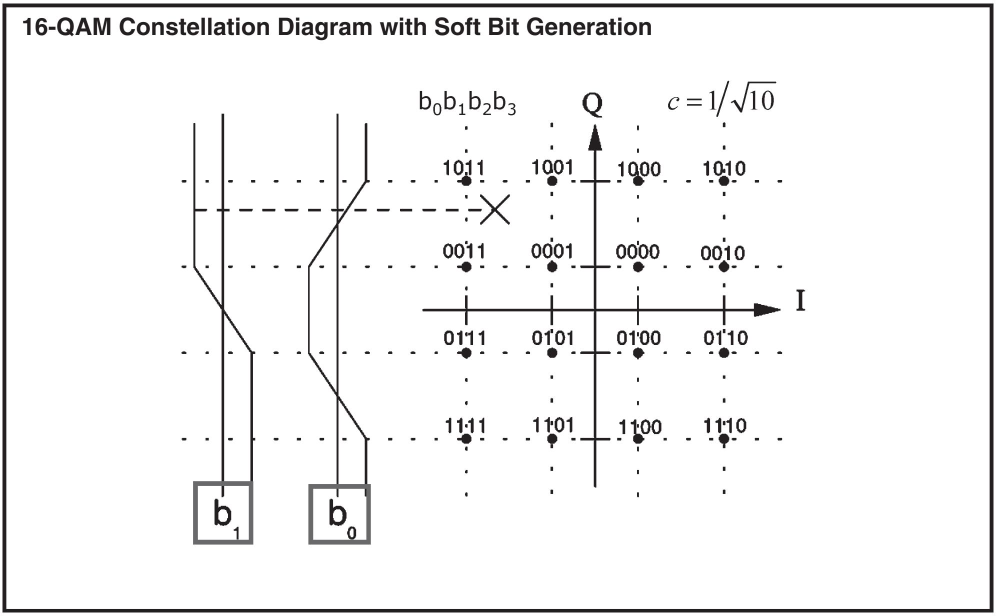


Figure 2.8 16-QAM constellation diagram with soft bit generation.

(a) Modulation and Coding Schemes

Modulation	Overall Coding Rate	Uncoded Block Size	Coded Block Size
BPSK	1 / 2	12 byte	24 byte
QPSK	1 / 2	24 byte	48 byte
QPSK	3 / 4	36 byte	48 byte
16 QAM	1 / 2	48 byte	96 byte
16 QAM	3 / 4	72 byte	96 byte
64 QAM	2 / 3	96 byte	144 byte

(b) OFDM Delay Spread Values

Type of Environment	Max. Delay Spread
In-Building (house, office)	$<0.1\,\mu s$
Large Building (factory, malls)	$<0.2\,\mu s$
Open Area	$<0.2\,\mu s$
Suburban Area LOS	$0.2-1.0\,\mu s$
Suburban Area NLOS	$0.4-2.0\,\mu s$
Urban Area	$1.0-3.0\,\mu s$

Figure 2.9 (a) Modulation and coding schemes, an example of IEEE 802.16. (b) Delay spread values in OFDM depending on radio environment.

the perceived constellation points of the subcarriers either in hard or soft bits (refer to Figure 2.8). Soft bits generated by the demodulator as well as soft bits introduced by the depuncturing module contained in the receiver may increase the performance of the convolutional decoder.

2.2 Duplexing Schemes

In general, duplexing schemes provide a separation of the send and receive signals of a terminal or Subscriber Station (SS). Partitioning of the wireless medium by duplex separation prohibits the self-interference of a station so that it does not receive its own transmitted signal. The typical signal pathlength between sender and receiver in cellular networks and related pathloss leads to a much higher signal strength at a station when sending compared to the received signal strength.

2.2.1 Time Division Duplex

The alternate transmitting and receiving of data on a single frequency channel is referred to as Time Division Duplex (TDD). In TDD, the times for transmitting and receiving are periodically alternating as depicted in Figure 2.10. Thus the uplink (UL) and downlink (DL) directions are separated in the time domain. The point of time when switching from reception to transmission is referred to as switching point. The position of the switching point allows an asymmetric separation of a frequency channel into DL and UL. Thus, asymmetric services can be effectively supported that transmit more data into one direction (DL) than into the other.

In TDD, radio signal propagation durations have to be taken into account in order to avoid a partial time overlap and potential interference of data packets originating from different senders. Therefore, guard times as part of the time slots shown in Figure 2.10 are introduced so that packets

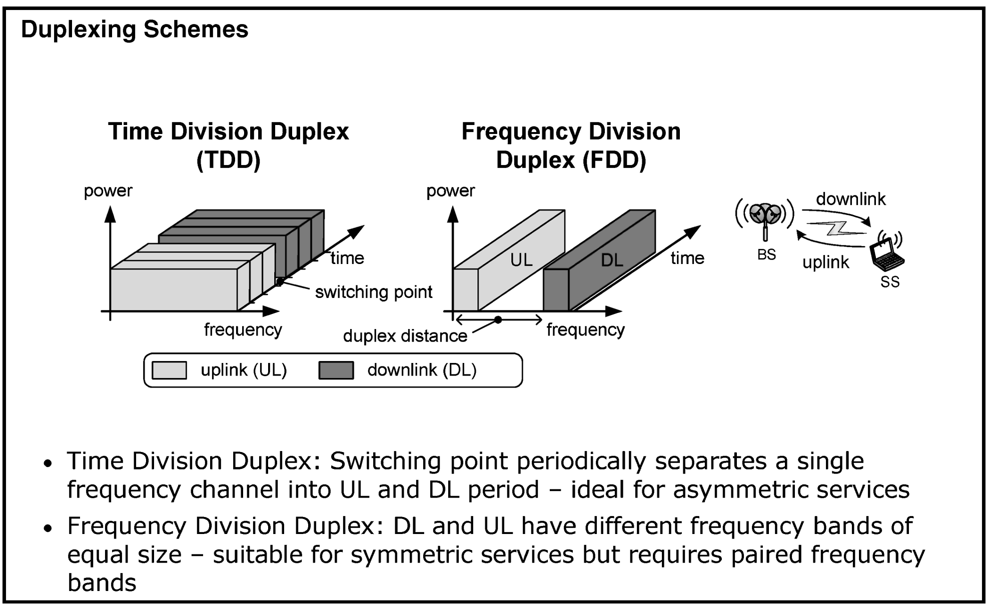


Figure 2.10 Duplexing schemes of time division and frequency division duplex.

cannot use the full slot for data transmission resulting in less efficient channel utilization. Such guard times are also required to switch the transceiver from transmission to reception. Prominent examples for wireless communication systems using TDD are IEEE 802.11 and 802.16.

2.2.2 *Frequency Division Duplex*

In the case of a Frequency Division Duplex (FDD) scheme, the UL and DL of a wireless network are separated in the frequency domain. The stations transmit and receive in different frequency bands as illustrated in Figure 2.10. The frequency channel used by one station for transmitting is used by another station for receiving and vice versa. In centrally controlled systems, the upper frequency band is usually assigned to the base station to transmit in DL direction, as the pathloss of radio signals increases with frequency. To communicate over the same distance, the base station therefore requires a higher transmission power (on the upper frequency band) than the subscriber station (transmitting on UL in the lower frequency band). Since battery power is limited at mobile devices, the lower (less power intensive) frequency band is used for UL transmission. The distance in MHz between paired frequency bands used with FDD is referred to as duplexing distance. The FDD scheme is ideal for producing a symmetric data flow in both directions, UL and DL. This facilitates an equal dimensioning of the used frequency bands and substantially eases cellular network planning. Today's cellular radio networks like GSM, UMTS and cdma2000 apply FDD, but also IEEE 802.16.

2.3 Multiplexing

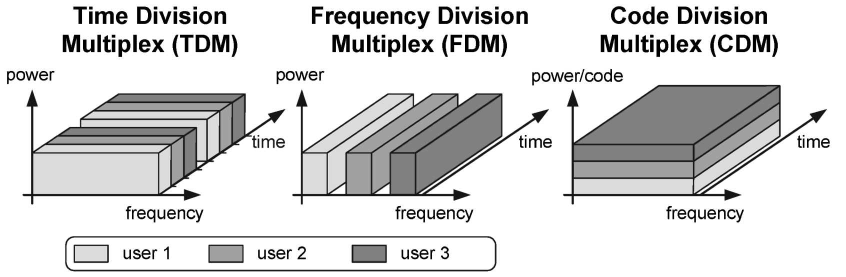
Multiplexing serves to share the radio channel capacity between competing stations. A multitude of simultaneous transmission requests is multiplexed to a common channel. A wireless medium seen as a transmission resource can be divided into multiple dimensions, namely frequency, time, code and space. Access to a multiplexed resource is by means of a multiple access rule so that the multiplexing scheme and the multiple access rule are strongly related. In addition, the duplexing scheme needs to be specified to fully characterize a transmission system. Today's wireless communication protocols combine several multiple access rules and duplexing schemes. For instance the current enhancements of IEEE 802.16 that combine TDD, TDMA with SDMA. GSM can be regarded as an FDMA/TDMA/FDD system while one version of UMTS combines for instance FDMA/CDMA/TDMA and TDD. These multiplex and multiple access schemes are discussed in the following sections.

2.3.1 *Frequency Division Multiplex*

In a Frequency Division Multiplex (FDM) based system the spectrum is divided into frequency channels that each may be simultaneously used by multiple users. The FDM principle is illustrated in Figure 2.11: The transmission signals of multiple stations are separated in the frequency domain. A single station may transmit and receive on a dedicated frequency channel. In most FDM-based systems a Multiple Access (MA) rule is applied to allow multiple stations to share a number of FDM channels being called Frequency Division Multiple Access (FDMA). The overall system bandwidth is divided by FDM among the communicating stations. Thus, only a fraction of the overall capacity of a multi-channel band is available to a single station using one channel.

The partitioning of spectrum into frequency channels is realized by modulating the corresponding carrier frequencies with the data messages of stations to be transmitted. At the receiver, a

Multiplexing Schemes



- Application examples:
- Time Division Multiplex: IEEE 802.16, GSM, DECT
- Frequency Division Multiplex: UMTS
- Code Division Multiplex: UMTS, CDMA 2000

Figure 2.11 Time, frequency and code division multiplexing schemes.

separation of stations' signals is reached through filtering. Due to the slope of real filters, guard bands are required between channels in order to avoid adjacent channel interference between neighboring frequency bands. The separation of multiple users through FDM implies many guard bands and thus spectral inefficiency. Therefore, FDM is usually combined with another multiplexing scheme for separating users such as CDM or TDM.

2.3.2 Time Division Multiplex

The capacity of a frequency channel may be higher than required for a single communication link and therefore may be divided in the time domain among several stations. This principle is referred to as Time Division Multiplex (TDM). The stations sequentially transmit on a common frequency channel that is separated into periodic time slots forming time channels. The assignment of time slots to stations may be under central or decentral control according to a multiple access rule (here: TDMA). Time slots might not be periodic in time but might be considered as individual time slots each accessed according to a Medium Access Control (MAC) protocol. The complete capacity of a frequency channel may be used by a single station under TDMA, which allows great flexibility in assigning capacity to specific stations. The separation of the wireless medium into time slots may lead again to inefficient channel utilization if the data packets do not completely fill a time slot. The available capacity per station is lower than the overall system capacity.

TDMA might lead to a more efficient resource utilization than FDMA as a frequency channel may be shared by multiple stations. However, TDMA requires a system wide synchronization in order to guarantee in time transmission of bursts in slots and thereby introduces more complexity. Guard times as part of slots reduce the user data capacity but help to avoid interference between time slots/channels.

2.3.3 Code Division Multiplex

With Code Division Multiplex (CDM) a radio signal of small bandwidth is transmitted in a wide frequency channel. The small bandwidth signal is spread through an adequate code to become a wide bandwidth signal. The spreading codes are bipolar code sequences, orthogonal to each other. CDM allows the simultaneous transmission of multiple stations in the same frequency band. User signals are separated at the receiver through correlating the received signal with the channel specific code sequence. Orthogonal spreading codes aim to realize minimal mutual interference in spite of simultaneous transmission, which can only be reached in a synchronized system. Access to CDM channels is under control of the CDMA multiple access rules specifying the codes available to be used for transmission at a station.

The number of stations transmitting in the same frequency channel is also limited in CDM: With an increasing number of code channels per frequency channel, the signal-to-noise ratio (SNR) is increased and the required threshold for successful decoding at the receiver might not be reached.

2.3.4 Space Division Multiplex

The spatial reuse of the wireless medium through multiple users is possible owing to pathloss, shadowing, sectorized antenna patterns and use of sophisticated phased antenna arrays combined with advanced signal processing able to support multiple transmissions/receptions at the same time, e.g. by using beamforming. Such multiplexing is referred to as Space Division Multiplex (SDM). Simultaneous transmission to multiple users without causing too much interference can be applied, depending on the antenna characteristics (mainly the number of antenna elements). Mutual interference can not always be avoided because of sidelobes of the antenna radiation pattern when using beamforming. If receive stations are located close by, spatial separation is difficult to realize. Space Division Multiple Access (SDMA) describes how the spatial resources may be accessed by multiple stations at the same time.

SDMA can be applied by any transmitter owning an antenna array making this technology suitable for Access Points (APs) of a communication system. In the UL direction, joint detection algorithms at the AP receiver can provide a separation of signals of multiple stations, similar to the DL.

2.3.5 Orthogonal Frequency Division Multiplex

Radio transmission outdoors and within buildings typically is subject to multipath propagation as described in Section 2.1.4. In a conventional serial data transmission system, short data symbols are being sequentially transmitted and the frequency spectrum of each data symbol transmitted is allowed to occupy the entire bandwidth of the frequency channel. As described in the previous section, under multipath propagation the received signal arrives as an unpredictable set of reflected and direct radio waves each with its own degree of attenuation, delay and phase shift. This leads to Inter-Symbol Interference (ISI) of consecutively transmitted data symbols at the receiver due to the signal delay spread.

Multi-carrier Modulation (MCM) techniques such as Orthogonal Frequency Division Multiplexing (OFDM) transmit data by dividing the high symbol rate stream into several parallel low symbol rate streams, and by using these substreams to modulate equally spaced subcarriers of a given radio channel. OFDM is the basis for many wireless communication technologies discussed in this book. The symbol duration of each substream is aimed to be higher than the channel delay spread and the maximum excess delay of delayed signals, and these parameters define the number of subcarriers to be used in a channel. By using a large number of OFDM subcarriers, immunity against the multipath effect can be provided (Cimini, 1985). OFDM as introduced by Weinstein and Ebert (1971), having densely spaced subcarriers with overlapping spectra of the modulated subcarriers, abandons the use of steep bandpass filters to separate each subcarrier as it is used in conventional FDM schemes. It offers therefore a high spectral efficiency. There are extensions of OFDM, which are not used in 802.11a, towards more flexible MCM schemes based on Wavelet transforms, where the individual subcarriers have different bandwidths. This aims to provide an accurate adaptation of the transmission scheme, in terms of data-throughput per subcarrier, to the time-variant radio channel. Signals received from multiple indirect paths added to the direct path do not meet the condition of orthogonality between subcarriers resulting in Inter-Channel subchannel Interference (ICI).

The part of the OFDM symbol that carries the information is in the following referred to as block instead of symbol, to distinguish it from the symbols transmitted on subcarriers. Insertion of a guard interval T_g of sufficient length given by the channel excess delay before a block of length T_b serves to avoid ICI and ISI. An OFDM block duration, $T_b' = T_g + T_b$ is then obtained representing what is known as OFDM symbol duration. The guard interval is typically chosen smaller than $T_b/4$. If T_g is longer than the maximum channel excess delay, the subcarriers are still mutually orthogonal inside the effective block interval $(T_g \ldots T_b')$. Adding a guard interval means that a cyclically extended OFDM symbol is transmitted.

In an Orthogonal Frequency Division Multiple Access (OFDMA) scheme a subset of OFDM subcarriers is assigned to transmit to individual stations. Thus, OFDMA can be regarded as an extension of OFDM in introducing a multiple access scheme. The base station transmits symbols on subchannels built from subcarriers that are orthogonal to those of other stations. Thus a simultaneous transmission on multiple subchannels on the DL can be realized. In the case of imperfect synchronization, which applies when multiple stations transmit uplink on different subchannels, the stations might nevertheless interfere at the base station. Flexibility in capacity assignment can be realized in assigning the DL more than one subcarrier to a single station. The subcarriers assigned from the multitude of available subcarriers to one station of a system may be clustered or distributed in the channel. Clustered subcarriers are adjacent in the time and/or frequency domain, while distributed subcarriers are separated by multiple subcarriers assigned to other stations. Clustering allows interference avoidance and multi-user diversity while distributed subcarriers enable an averaging of interference at the stations (Einhaus *et al.*, 2005).

2.3.5.1 Pilot Tones and Preambles

In between the subcarriers used for data transmission, subcarriers are inserted to transmit pilot tones (in 802.16 for instance 8 pilot tone subcarriers are inserted between 192 data subcarriers). Usually, these subcarriers are modulated with the most robust PHY mode (BPSK) in order to guarantee error-free reception over a large transmission distance. Thus, in 802.11 and 802.16 these subcarriers are BPSK modulated with a pseudo-random binary sequence. The actual BPSK constellation point depends on the random sequence initialization and on the pilot tone subcarrier index. Since the receiver knows the modulation of the pilot tones very well, this knowledge can be used for channel estimation, synchronization, and tracking of frequency offsets during the regular reception of an OFDM symbol. The remaining subcarriers, i.e., 55 guard carriers and the DC carrier are not modulated at all.

Preambles can be appended either in the frequency domain, i.e., before processing a symbol in the IFFT transmitter module or in the time domain, i.e., thereafter. Preambles and midambles are predefined training symbols and are therefore used for frame time and frequency synchronization and for channel estimation purposes. Preambles are composed of QPSK modulated subcarriers. Only every second or every fourth subcarrier is used so that the resulting time-domain waveform consists of two, respectively, four repetitive parts. The long preamble consists of two consecutive OFDM symbols. A long preamble is appended at the beginning of each DL subframe to indicate the start of the MAC frame and it is used to transmit initial ranging messages in contention-based uplink slots during the network entry procedure. Short preambles consist of only one OFDM symbol. Short preambles precede UL bursts and, optionally, DL bursts. As UL midambles, they may be included periodically in UL bursts. Short preambles might be cyclically shifted by an integer number of samples. By means of the cyclic shift, concurrently transmitted preambles become separately detectable. The preamble modification can be used by SDMA techniques such as joint detection and pre-distortion. The cyclic shift can also be leveraged by other advanced antenna techniques, e.g. space-time codes.

2.3.5.2 Fast Fourier Transformation (FFT)

Only with (nearly) periodic discrete-time signals, a convolution of two signals is equivalent to multiplying the Fourier transforms (frequency responses) of the respective signals, here an OFDM symbol and the channel impulse response. OFDM symbols are created by an Inverse Fast Fourier Transform (IFFT) of the data to be transmitted. The frequency response of an OFDM symbol generated with an IFFT is the original data, thus each original data symbol is multiplied by a single complex number for transmission over the radio channel. Equalization at the receiver becomes very simple, or even can be avoided. The IFFT creates the OFDM waveform by transforming the complex constellation points from the frequency domain (see Figure 2.8) into the time domain.

This transformation is discussed in the follwing at the example of OFDM in IEEE 802.16 using 256 subcarriers. Since a 256 Fast Fourier Transformation (FFT) is used in 802.16, the addition of 256 individual subcarrier waveforms composes the resulting OFDM waveform. As an example, the upper graph in Figure 2.12 shows three separated subcarriers modulated with the

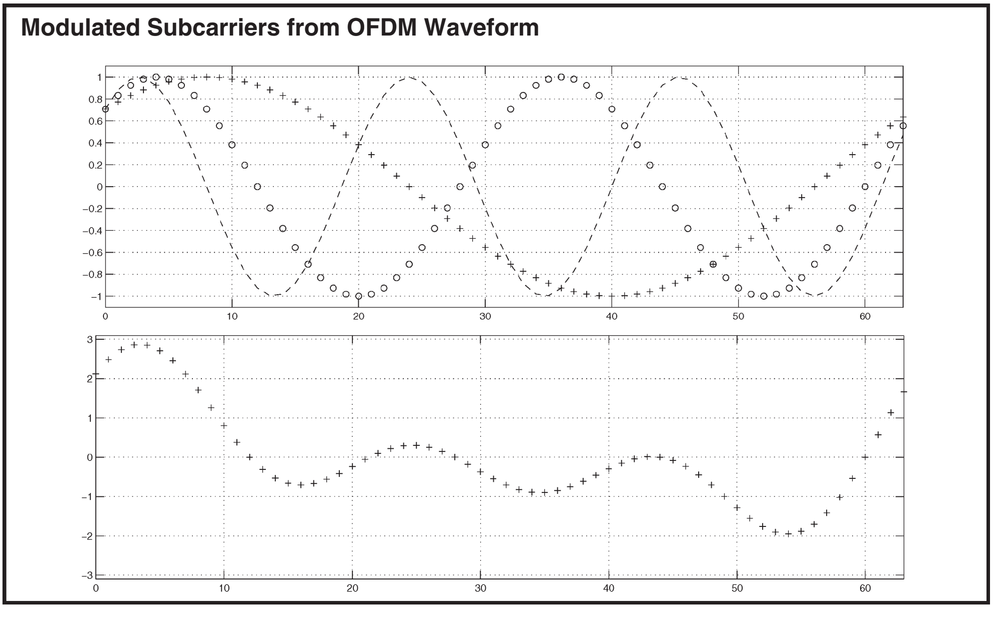


Figure 2.12 Modulated subcarriers from OFDM waveform.

complex values $1/\sqrt{2}^{*}(1-i)$ (phase shift of $-45°$ representing the QPSK modulated bit sequence bx10). The lower graph in Figure 2.12 shows the addition of three subcarriers. In general, the addition of all 256 sinusoidal waves results in a high Peak to Average Power Ratio (PAPR). For instance, the worst case PAPR of a QPSK-modulated OFDM symbol (with an oversampling factor $L=1$) is 23 dB. Typical PAPRs vary between 8 and 13 dB. PAPR reduction techniques can reduce this value by 2 to 4 dB (Han and Lee, 2005). Still, the transeiver's power amplifier has to be linear over a wide range of power, which can be solved now but was a challenge for a long time.

2.3.5.3 Cyclic Prefix

Delay spread from multipath signal propagation causes ISI and ICI, see Section 2.3.5. In order to preserve the orthogonality of subcarriers a Cyclic Prefix (CP) is introduced in front of every useful part of an OFDM symbol (van Nee and Prasad, 2000). The CP in an OFDM symbol helps to remove the impact of multipath signal delay spread by making the OFDM symbol appear periodic in time. As Figure 2.13 indicates, the CP extends the useful OFDM symbol duration by copying the last portion of the OFDM waveform to the front of the symbol. Different CP lengths, measured in ratios of CP duration to useful symbol duration, e.g., values of 1/4, 1/8, 1/16, and 1/32 are specified in 802.16. By extending the OFDM symbol, the samples required for performing the FFT at the receiver can be taken anywhere over the length of the extended symbol. Thus, the CP provides multipath immunity as well as a tolerance for symbol time synchronization errors. As a drawback, the overhead introduced by the cyclic extension results in a reduced SINR (Engels, 2002).

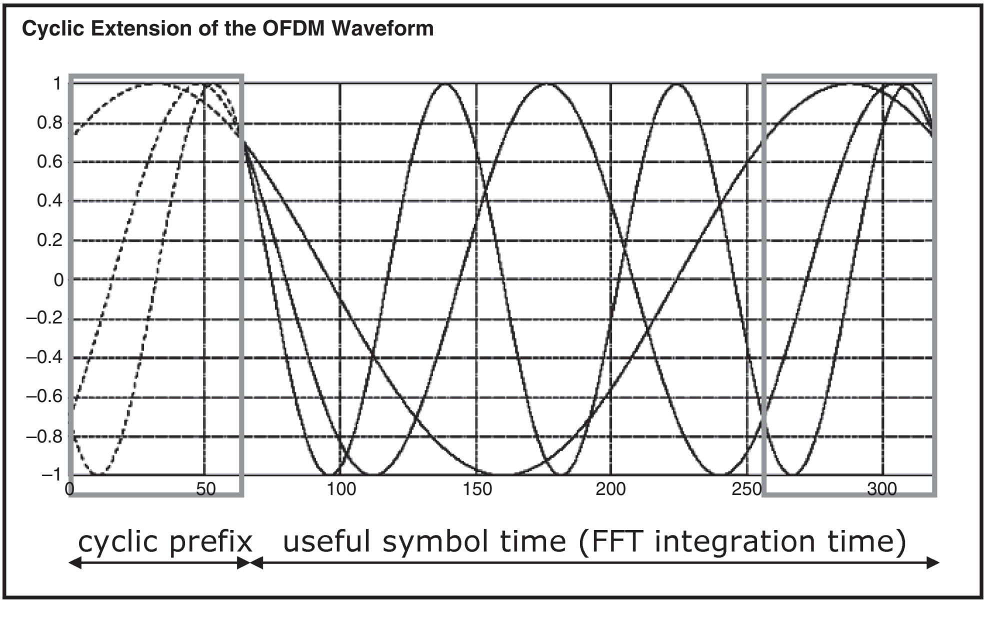


Figure 2.13 Cyclic extension of the OFDM waveform.

In the targeted frequency bands, radio communication benefits significantly from the ability to operate under obstructed LOS and NLOS conditions. It is therefore necessary to choose a CP larger than the maximum delay spread but still as short as possible. Table (b) of Figure 2.9 lists common maximum delay spread values in different types of environment. These delay spread values remain unchanged for any operating frequency above 30 MHz, since the wavelengths become much smaller than manmade architectural structures. Recent measurements confirm the values for frequency bands between 800 MHz and 6 GHz (Kepler *et al.*, 2002; van Nee and Prasad, 2000).

Two drawbacks of the cyclic prefix are worth mentioning (Mangold *et al.*, 2001a). One is that redundant data is transmitted over the radio channel reducing the net data throughput. The other is that the CP of duration T_g leads to a power loss, as the receiver only uses the energy received during the time T_b. The energy corresponding to T_g is being discarded. A power loss α_g must be taken into account:

$$\alpha_g = \frac{T_b}{T_{b'}}$$

2.4 Switching in Communication Networks

2.4.1 Circuit Switching

Circuit-switched communication implies a dedicated communication link that might be a frequency, time or code based channel between data source and sink. In circuit switching a connection management function is required responsible for establishing a connection, before transmitting data and releasing a connection afterwards. Circuit switching implies a permanent reservation of channel capacity for the duration of a connection independent of the actual load of the channel. Thus, circuit switching may be rather inefficient. The establishment of a connection that might run across a number of intermediate router nodes results in an initial delay. However, the delay for user data transmitted on the channel is negligibly low. Signal propagation delay and channel data rate are the parameters that govern the time duration required to transmit a data packet. It is worth mentioning that most IEEE 802 systems use reservation-based transmission, which is in a sense circuit switching but with short live duration of the channel reserved.

2.4.2 Packet Switching

In packet switching, communication is based on data segmented into packets for transmission. A packet is individually transmitted over the wireless medium and packet transmissions are independent of each other, i.e., each packet has to compete on its own for medium access applying some multiple access rule, which might cause considerable initial access delay. Packets might require a virtual connection to be established in advance, as is usual with ATM transmission systems, or might just be self-contained in that all information to route a packet from source station to sink is contained in the packet header, as is usual for IP datagrams in the Internet or for packets transmitted via a wireless link or LAN. Transmission delay of packets may be shorter than in circuit-switched systems, if the packet transmitting medium has the full capacity of the respective frequency channel available, which is not the case in most systems applying circuit switching. The network may drop some packets due to congestion, while under circuit switching the network provides a reserved channel according to the capacity needs of communicating stations.

Packet and circuit switching may require additional functions in order to guarantee reliable communication. Error control as discussed below is necessary to assure the correct transmission of packets. Then, data packets are stored for a certain time duration at the transmitter to enable retransmission on request. Under packet switching, segmentation at the transmitter and reassembly at the receiver is required to fit the packet size used for transmission and sequencing at the receiver to guarantee the correct sequence of received packets. Further, flow control is necessary with packet switching in order to handle overload situations at the receiver and increase the packet throughput with the help of a transmit window.

2.5 Channel Coding for Error Correction and Error Detection

2.5.1 Forward Error Correction

The correction of transmission errors with the help of channel coding is called Forward Error Correction (FEC). Redundancy is added by the sender to the data packet to allow the receiver to correct errors. The more redundancy added at the transmitter the higher the number of errors that can be corrected at the receiver. Different to ARQ protocols introduced in the next section, FEC requires no return (feedback) channel from receiver to transmitter. The codes used for FEC are either linear block codes or convolutional codes. Linear block codes are systematic, i.e., the redundant bits are contained separated from the redundant data in a packet. Contrary, convolutional codes combine a number of sequential bits so that the input information is not visible in the encoded output packet. Block codes are used for coding packets (blocks) with a predetermined size while convolutional codes are applied to streaming data of bits or symbols. A prominent example related to convolutional coding is decoding by means of the Viterbi algorithm. For further details on FEC, please see Lin and Castello (2004).

FEC enables a constant throughput and packet delay independent of the current quality (bit error ratio, BER) of the radio channel. Major drawbacks of FEC are: (i) a substantially reduced throughput owing to a code rate of 2/3 up to 1/2 to be able to correct most of the possible transmit errors and (ii) a residual bit error rate essentially depending on the actual channel quality.

2.5.2 Automatic Repeat Request Protocols

Different from forward error correction of data packets at the receiver, Automatic Repeat Request (ARQ) protocols apply systematic redundancy at the transmitter in order to detect errors at the receiver, not to directly correct them. The receiver identifies erroneous packets based on an error-detecting code and requests repeated transmission from the sender. Systematic block codes are used for error detection and the redundant bits are being attached as a checksum to the data packet called Cyclic Redundancy Check (CRC) or Frame Check Sequence (FCS). The ARQ protocols require a feedback channel from the receiver to the sender to be able to notify the sender of a successful or non-successful transmission using an acknowledgment. A successfully received data packet is acknowledged positively by sending an acknowledgment (ACK) otherwise a negative acknowledgment (NACK) is sent back. This principle implies that the sender keeps a transmitted data packet as long as the acknowledgment is pending.

ARQ protocols imply a flow control function whereby the data flow of the transmitter can be regulated from the receiver. Therefore, transmitted data packets are sequentially numbered at the sender, say modulo 8. A window mechanism helps to synchronize two communicating stations with respect to the successfully transmitted packets. The maximum number of data packets that may be transmitted by the sender without having received an acknowledgment is specified by the window size. Since with duplex transmission both communicating stations may be transmitter and receiver, different window sizes may exist in the two directions. Examples for different window sizes on duplex transmission are for instance discussed in Tanenbaum (1988) or Stallings (1991).

In the following three prominent ARQ protocols are discussed. Their throughput in terms of radio channel utilization is compared based on Walke (2002).

2.5.2.1 Send-and-Wait

The simplest but most inefficient ARQ protocol is the Send-and-Wait ARQ, in the literature also referred to as the Stop-and-Wait ARQ. The sender waits for an acknowledgment (positive or negative) of a transmitted data packet before transmitting the next one. If a NACK is received the packet transmission is sent once more, otherwise the next packet in sequence is scheduled for transmission. Thus, both stations have a window size of 1. The principle of the Send-and-Wait ARQ is illustrated in the timing diagram (a) of Figure 2.14. The throughput of such a system is very low, especially for long signal propagation delay and small data packet size, as the sender must wait a relatively long time until an ACK/NACK has arrived. The throughput of the channel is dependent on the round trip delay t_{rd} and the packet error ratio (PER). The round trip delay is composed from two elements, the time required to transmit a packet to the receiver and the time needed for traveling back the acknowledgment.

The throughput S_{SW} of the channel is given as

$$S_{SW} = \frac{n \cdot (1 - PER)}{n + t_{rd} \cdot v},$$

where n represents the packet length in bits, and v the data rate of the channel in bit/s.

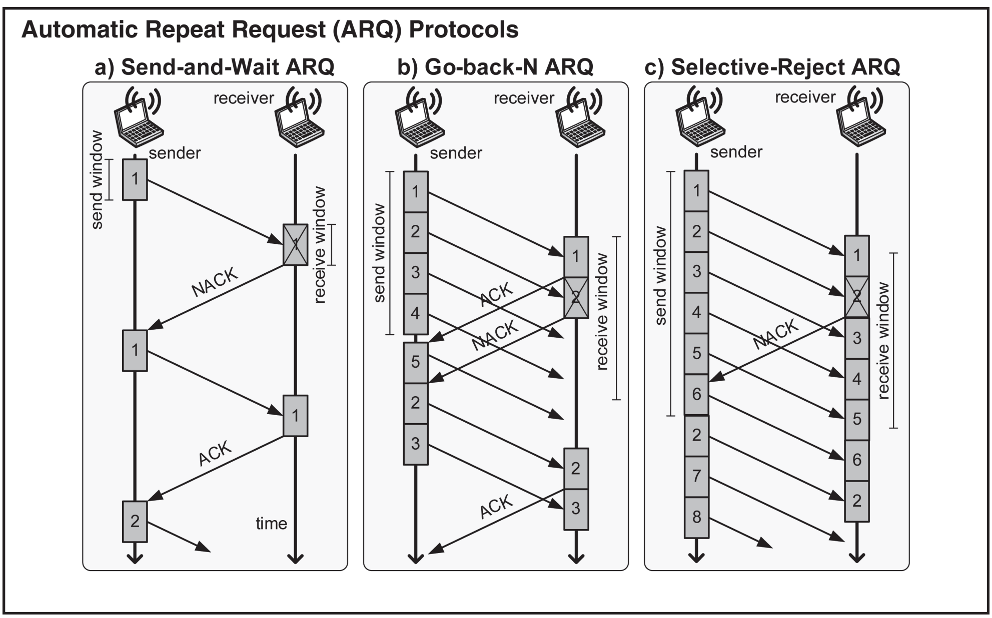


Figure 2.14 Different versions of Automatic Repeat Request (ARQ) protocols. (a) The Send-and-Wait ARQ. (b) Go-back-N ARQ. (c) Selective-Reject ARQ.

2.5.2.2 Go-back-N

The Go-back-N ARQ protocol leads to a continuous data flow on the channel and in the literature it is also known as Reject-ARQ, Cumulative ARQ or Continuous ARQ. When applying the Go-back-N ARQ protocol, the sender transmits in the limits of its window sequentially numbered data packets. If a NACK is received from the receiver, the sender retransmits all packets staring from the packet number specified in the NACK message and considers packets with lower sequence numbers as acknowledged. The receiver ignores all data packets following a packet received in error and only accepts packets in correct sequence that are judged error free. The Go-back-N ARQ is illustrated in timing diagram (b) of Figure 2.14. The protocol requires a sufficiently large buffer size at the sender in order to retransmit data packets on request. If the sender receives an acknowledgment for an already transmitted data packet, all packets with an equal or lower sequence number (modulo, say, 8) are regarded as successful and are cleared from the buffer. This mechanism is referred to as Cumulative ACK. The throughput S_{GBN} of the Go-back-N ARQ protocol can be calculated from

$$S_{GBN} = \frac{n \cdot (1 - PER)}{n + PER \cdot t_{rd} \cdot v}.$$

2.5.2.3 Selective-Reject

Similar to the Go-back-N ARQ, the Selective-Reject ARQ protocol permits in the limits of the window a continuous transmission of data packets on the channel. After detecting a packet error at the receiver, the Selective-Reject ARQ protocol sends a reject message to the sender identifying this data packet. Any other correctly received packets are stored in a receive data buffer. Only the rejected packet is repeated by the sender. All packets with sequence numbers higher than the rejected one must wait in the receive buffer until successful retransmission of the failed packet in order to provide an in-sequence delivery of data packets to the higher layer. The sender also must store all packets until receiving an acknowledgment for each. The principle of the Selective-Reject ARQ is illustrated in the timing diagram (c) of Figure 2.14. The throughput S_{SR} of the Selective-Reject ARQ protocol is given under the assumption of an unlimited receive buffer size

$$S_{SR} = 1 - PER.$$

Similar to the Go-back-N ARQ, cumulative ACKs are allowed whereby the sequence number contained in an ACK message acknowledges all packets sent earlier with lower sequence numbers.

In the literature, the Selective-Reject ARQ is often equivalently referred to as Selective-Repeat ARQ. A data packet is selectively repeated after receiving a NACK.

2.5.2.4 Summary

The Selective-Reject ARQ protocol leads to the highest throughput of the three ARQ protocols discussed in Figure 2.15. Here, the radio channel has a data rate of $v = 6\,\text{Mbit/s}$, a packet data size of $n = 1514$ byte and an unlimited window size is assumed, and the round trip delay t_{rd} is 1 ms.

ARQ protocols result in a high system throughput under good channel quality and a very low residual bit error ratio can be guaranteed. The major drawback is the need for a feedback channel and the overhead needed for signaling ACKs. The throughput degrades at low channel quality and packet delay jitter increases with increasing packet error ratio.

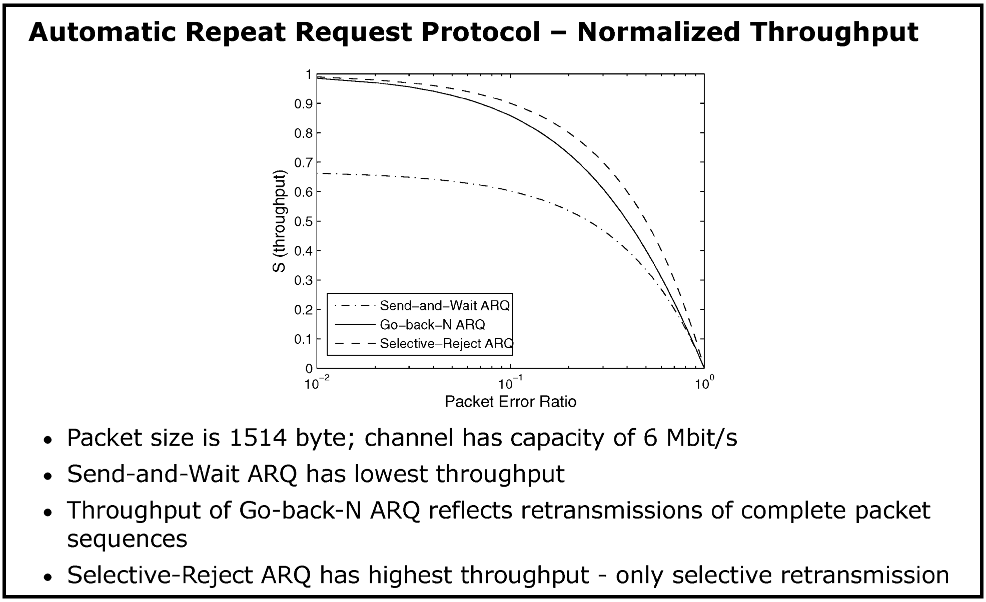


Figure 2.15 Normalized throughput of Send-and-Wait ARQ, Go-back-N ARQ and Selective-Reject ARQ is dependent on the packet error ratio.

The discussed ARQ protocols may be combined. Mixed mode ARQ protocols that incorporate a selective repeat mode are analyzed in Miller and Lin (1981).

2.5.3 Hybrid Automatic Repeat Request

In wireless systems, FEC and ARQ are often combined to take maximum benefit from both approaches: The ARQ protocol is applied to data transmission using FEC, i.e., packet errors are first corrected and error detection is applied thereafter. FEC reduces the number of retransmissions needed, while the ARQ protocol eliminates the remaining error Hybrid ARQ protocols apply soft combining at the receiver of a packet received in error and its copy retransmitted by the sender on request, called *chase combining*. Instead of a copy, advanced HARQ protocols may send redundant bits instead that had been punctured (eliminated and not sent) during FEC coding of the original packet. Using this *incremental redundancy*, the receiver may be able to correct errors contained in a packet.

2.6 Medium Access Control (MAC) Protocols

Medium access control protocols are used in communication networks when multiple users compete for using the same channel. In the following, approaches for realizing TDMA are discussed. The transmissions of multiple users on a shared frequency channel are separated in the time domain. Many MAC protocols used in wireless communication have their origin in fixed communication networks. When comparing a wired with a wireless radio channel,

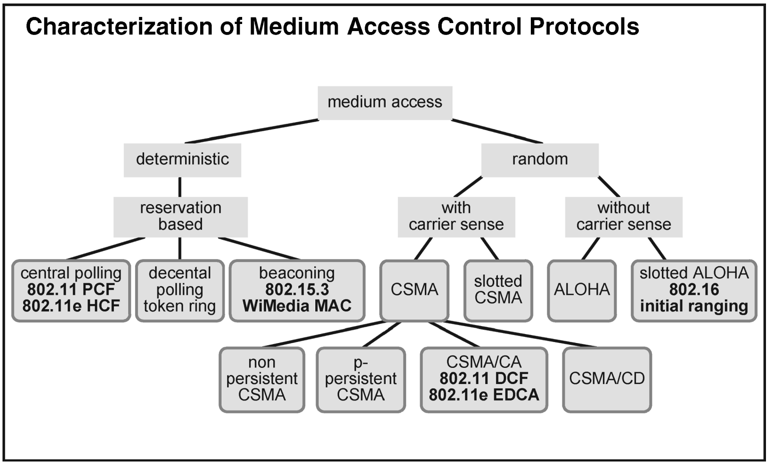


Figure 2.16 Overview of different medium access control protocols. Standards discussed in this book that use these protocols are in bold.

different propagation conditions for transmissions can be assessed. They have to be taken into account and require additional mechanisms for a proper protocol operation.

In general, MAC protocols can be divided into two classes: deterministic and random, contention-based access as illustrated in Figure 2.16. The main characteristic of random-based medium access is that no predictable or scheduled point in time for medium access of stations exists. QoS is therefore difficult to support under strong competition for medium access as for instance shown in Section 5.5.5 in the context of decentralized operation of IEEE 802.11e.

Deterministic medium access is realized through reservation. Then, coordination of reservations is required that may either be under central or decentral control. With decentral access control, both reservation messages and user data may collide if access is not controlled properly. Under proper reservation control, the capacity lost due to collisions may be small since only the reservation packets may collide and not the data packets. Under central control a strict QoS support is possible. QoS support may be limited when decentral access control is applied.

Random access is used in wireless networks operating under central or decentral control for different reasons. The initial access to the network in order to associate and for establishing a logical connection to the networks is always based on random access. Radio resources may also be requested from a central controller based on random access. The contention slots used for initial ranging and bandwidth request in the UL subframe of 802.16 using slotted ALOHA are examples for how mobile stations can get in contact with a base station. Under decentral access control, random access to the medium is used to transmit user data packets, too, e.g., as specified in IEEE 802.11.

2.6.1 ALOHA

The ALOHA random access protocol was the first to realize uncoordinated access of multiple users to a shared channel for packet-based computer communication (Abramson, 1970).

The rationale behind an ALOHA system is simple: User terminals transmit their data as soon as they have data ready to transmit. This may lead to collisions, since users may transmit simultaneously or packets may partly overlap. The resulting interference may prohibit error-free decoding of a packet at the receiver. If the sender does not receive an acknowledgment for its transmission the data packet is transmitted anew after a random time duration reducing the probability of multiple collisions of the same packet.

The introduction of a slotted medium leads to a time discrete version of the ALOHA protocol called slotted ALOHA.

2.6.1.1 Pure ALOHA

With pure ALOHA, data packets are transmitted at random time instants over the medium. In the following, analytic results for system throughput and packet delay are derived and evaluated. The basic variables and assumptions are also used for analyzing slotted ALOHA and CSMA MAC protocols in the subsequent sections. Pure and slotted ALOHA are illustrated in Figure 2.17 by means of examples and their throughput and delay performance is summarized in Figure 2.18.

The traffic offered from user packets, normalized to the channel capacity is denoted G. Successfully transmitted (non-collided) packets contribute to the normalized throughput (goodput) S.

When assuming an infinite number of independent source stations the arrival process of new packets can be modeled as a Poisson process with parameter λ, where $\lambda = S$ in a non-saturated system. If all data packets have the same size, service time T is the normalization constant that is set to $T = 1$. Collided packets are transmitted after independent, random backoff times. Consequently, the offered traffic G, comprising new and repeatedly transmitted packets, also follows a Poisson process. With a normalized packet service duration of 1, the vulnerable period, during which packets may collide partly or as a whole, has a length of 2. Thus in the worst case

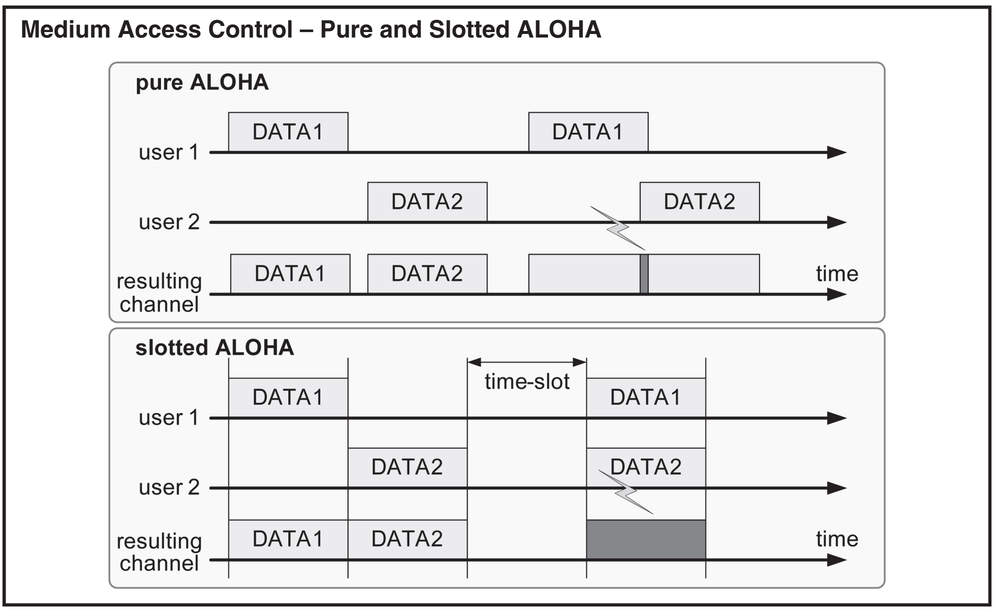


Figure 2.17 Random medium access according to pure and slotted ALOHA principles.

the duration of an interfered transmission is 2. The probability for a packet arrival during the vulnerable period is $1-e^{-2G}$ and the throughput of a pure ALOHA system is (Kleinrock, 1976)

$$S_{Pure\ Aloha} = Ge^{-2G}.$$

The maximum possible throughput of the system is $S = 1/2e = 0.18$ at an offered traffic load of $G = 0.5$. Thus, the maximum throughput is limited to 18% of the total channel capacity.

An approximation of the packet delay, the time duration from ready to transmit a packet until successful reception at the receiver, can also be calculated. The delay is the sum of the queuing, propagation and transmission time of a packet. The expected number of retransmissions per packet can be expressed as $G/S - 1 = e^{2G} - 1$. With a as normalized propagation delay, a normalized packet duration of 1 and an average delay for a retransmission of δ the delay is calculated to

$$D_{Pure\ Aloha} = \left(e^{-2G} - 1\right)\delta + a + 1.$$

The average retransmission delay δ (of successful transmissions) can be derived as follows: When drawing the backoff time before initiating a retransmission from a uniform distribution between 1 and k, the average delay is $(k+1)/2$. The time required for noticing that a transmission is not successful must be added to this value. It is given by the transmission time from the receiver to the sender $(1+a)$ plus the time needed for answering with an acknowledgment $(a+w)$. The variable w represents the time required for generating an acknowledgment. For simplicity it is assumed that acknowledgments do not collide. The overall delay is then

$$D_{Pure\ Aloha} = \left(e^{-2G} - 1\right)\left(1 + 2a + w + \frac{k+1}{2}\right) + a + 1.$$

Typically, a is very small and acknowledgment packets usually are essentially shorter than data packets leading to small values for w.

2.6.1.2 Slotted ALOHA

To increase the efficiency of medium utilization, Roberts (1975) suggested logically dividing the channel into equally sized time slots with a duration of the packet transmission time thereby creating a slotted ALOHA MAC protocol. A central system clock for realizing synchronization is required. In slotted ALOHA, a packet must be transmitted at the start of a time slot whereby a synchronization of packet transmissions is reached and the vulnerable period is reduced to 1. With the definitions from above, the probability for packets colliding is then given by $1-e^{-G}$ and the throughput is

$$S_{Slotted\ Aloha} = Ge^{-G}$$

The maximum value for the supported throughput, i.e., the capacity of the system, is $S = 1/e = 0.37$. The delay is calculated similar to pure ALOHA with one exception. Due to slotting, a station has to wait when a packet is ready until the next slot begins before initiating its transmission. On average this is half a slot time leading to

$$D_{Slotted\ Aloha} = \left(e^{-G} - 1\right)\left(1 + 2a + w + \frac{k+1}{2}\right) + 1.5a + 1.5.$$

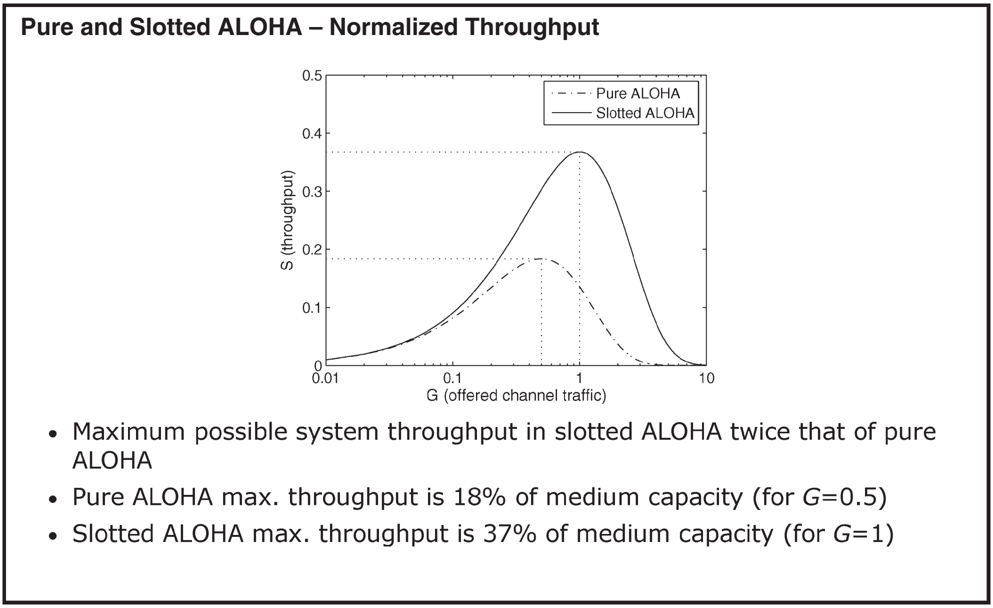


Figure 2.18 Normalized system throughput and delay for pure and slotted ALOHA.

A Markov model for evaluating the throughput and delay performance of slotted ALOHA multi-hop packet radio networks has been published by Takagi and Kleinrock (1985).

2.6.1.3 Comparison of Pure and Slotted ALOHA

Throughput and delay of pure ALOHA and slotted ALOHA are compared in Figure 2.18 for $a = 0.01$, $w = 0.01$ and $k = 10$. The graph illustrates the instability of a contention-based protocol. When increasing traffic load G an increased system throughput S results, until the system capacity is reached, namely $G = 0.5$ for pure ALOHA and $G = 1.0$ for slotted ALOHA, respectively. When further increasing load the system collapses and the throughput approaches zero. The reason for this is the increased collision probability. Under high traffic load, retransmissions overload the system. Even if in such a situation no new packets would arrive, no successful transmissions are possible until all pending packets are discarded and the system restarted. Various techniques for controlling random access in slotted ALOHA have been proposed as for instance in Corleial and Hellmann (1975), Cunningham (1990) and Kleinrock and Lam (1975).

2.6.2 Carrier Sense Multiple Access

The maximum reachable channel utilization of slotted ALOHA is 37%. The efficiency of channel utilization can essentially be improved when taking a prominent characteristic of wireless communication into account: Signal propagation delay on the radio channel is relatively small compared to the packet transmission time ($a \ll 1$).

Listening to (sensing) the radio channel before deciding to transmit is useful to avoid collisions. Such a medium access scheme is referred to as Carrier Sense Multiple Access (CSMA) or

Listen-Before-Talk (LBT). Compared to ALOHA, the channel utilization is much improved, as no ongoing transmissions are being interfered. If the channel is detected idle, a station may initiate its transmission. More than one station may attempt to access the channel at the same time, find it idle and transmit, and a collision will then occur. A transmitting station must wait a significant duration of time before it can regard a transmission as successful: The round trip propagation delay plus the time duration for contending for medium access spent by the destination station to retransmit an acknowledgment must be waited for.

Similar to ALOHA, the medium access based on the CSMA principle can be differentiated into unslotted and slotted CSMA. The organization of the radio channel based on time slots known to all stations competing for medium access leads to slotted CSMA. The CSMA used in IEEE 802.11 can be characterized as follows: The 802.11 OFDM symbol length introduces a value for a time discrete organization of the channel for distributed medium access that is significantly smaller than the length of a user data packet. 802.11 stations using the DCF have in common a slotted time structure for medium access that is not absolutely fixed but a slotting individual to any station after detecting an idle channel, see Section 5.4.1.

The station's behavior after detecting a busy channel, and in the case of a collision, leads to different variants of CSMA that will be discussed in the next sections. A detailed analysis of the throughput and delay characteristics of CSMA variants can be found in Kleinrock and Tobagi (1975).

2.6.2.1 CSMA Variants

In CSMA, rules are required to specify the reaction to a busy channel if a station has a packet ready waiting for transmission. Different variants of reaction are illustrated in Figure 2.19. The maximum achievable throughput and delay of some CSMA variants are presented in Figure 2.20. The formulas used for calculating the delay in the right-hand side of Figure 2.20 can be found in Kleinrock and Tobagi (1975).

Non-persistent CSMA

If a station would like to transmit and the medium is sensed busy, the station may wait for a random time duration before attempting a retransmission. This is referred to as non-persistent CSMA. With the definitions given above, the system throughput is

$$S_{non\text{-}persistent\ CSMA} = \frac{Ge^{-aG}}{G(1+2a)+e^{-aG}}.$$

The variable a represents the ratio of propagation delay to packet transmission time and G is the normalized offered traffic rate.

1-persistent CSMA

In order to reduce channel idle time, a station may initiate medium access as soon as the medium is idle. In the case of a collision a backoff period is applied by any station involved (a random waiting time is drawn by each from a specific time interval) before repeating medium access. This principle is called 1-persistent CSMA. The throughput S can be calculated as

$$S_{1-persistent\ CSMA} = \frac{G\,[1+G+aG(1+G+aG/2)]\,e^{-G(1+2a)}}{G(1+2a)-(1-e^{-aG})+(1+aG)e^{-G(1+a)}}.$$

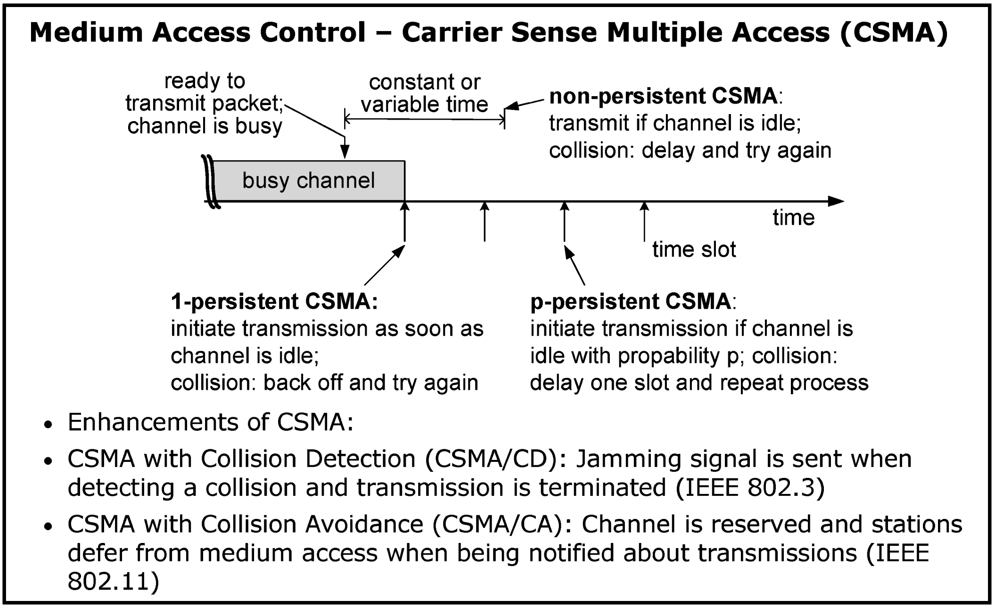


Figure 2.19 Medium access according to the Carrier Sense Multiple Access scheme.

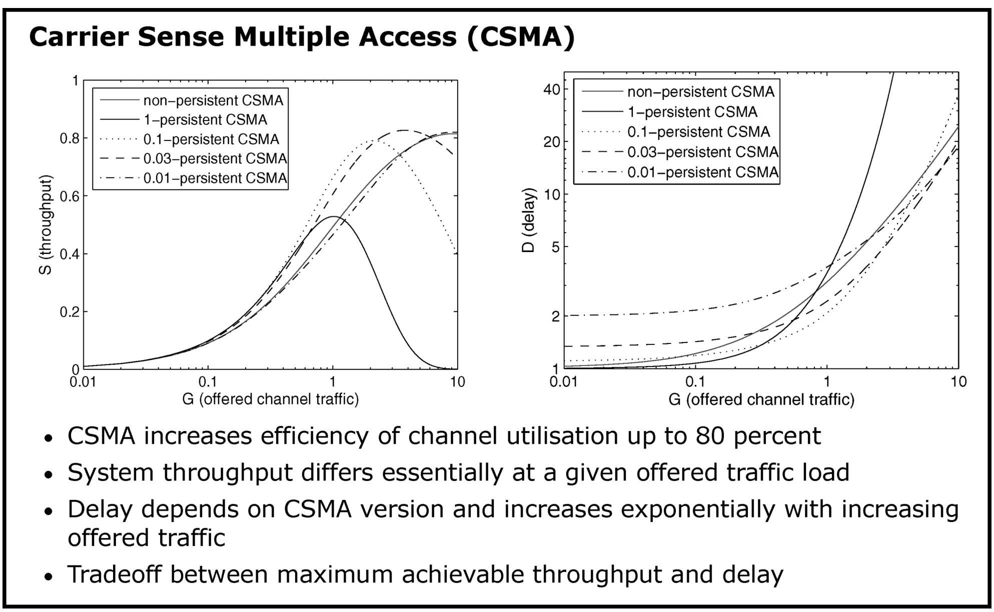


Figure 2.20 Comparison of the normalized system throughput of multiple ALOHA and CSMA versions.

p-persistent CSMA
The radio channel may be divided into logical time slots. These time slots do not necessarily have the same duration as the packet transmission duration. If an idle time slot is detected, the station may transmit its packet with probability p and defers from medium access with probability $(1 - p)$. In the case of a collision this procedure is repeated after a random backoff duration. A derivation of the system throughput when using p-persistent CSMA can be found in Kleinrock and Tobagi (1975).

Summary
The p-persistent CSMA appears to be a compromise in terms of high throughput and low delay, as illustrated in Figures 2.20 and Figure 2.21. For small p a high throughput with a long delay is gained, while a large p leads to a small throughput and short delay. By varying p and/or the interval from which the backoff duration is drawn, the behavior of the p-persistent CSMA can be adapted to the specific needs. An increase of the backoff interval by a factor of 2 after a collision is referred to as binary exponential backoff, which is also used in the DCF of 802.11.

2.6.2.2 CSMA/CD

The performance of CSMA can be improved when introducing means for Collision Detection (CD) resulting in CSMA/CD. On a wired medium, stations may observe the shared channel to detect if two transmissions have been started simultaneously, or at least overlap. If such a collision is detected, a jamming signal is transmitted by the respective station and its transmission is terminated immediately. The capacity of the medium wasted due to a collision can be reduced substantially, since the time spent to detect a collision and issue the jam signal, typically, is much

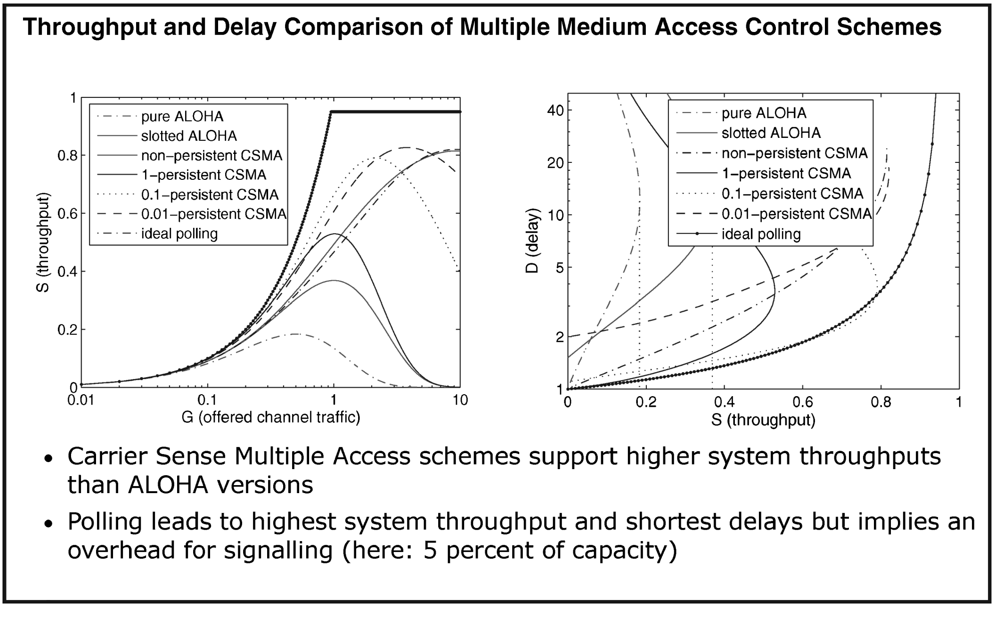


Figure 2.21 Throughput and delay comparison of various medium access control schemes.

shorter than the time needed to transmit a complete data packet. CSMA/CD is used in the LAN Ethernet (IEEE 802.3).

2.6.2.3 CSMA/CA

Collision Avoidance (CA) is an enhancement of CSMA-based radio channel access leading to CSMA/CA. Collision avoidance aims at reserving the channel in advance by broadcasting reservation-related information. Stations that sense the carrier may receive a channel reservation message containing a specific time duration that other stations are being asked to defer from channel access. After expiration of the time reserved all stations may compete again for medium access. This concept is used for instance in the DCF of 802.11 when using the RTS/CTS frame sequence for channel reservation in the coverage range of both, transmitter and receiver. The hidden station problem described in Section 4.4.1 can be essentially mitigated in this way.

2.6.3 Polling

In centrally controlled networks, the controlling device may invite (poll) associated stations one by one to transmit its data packets. Then, no random access is allowed and the channel access is deterministic, allowing the support of QoS requirements of the stations. Polling of stations may also be performed under decentral control, like in token based LANs (IEEE 802.4, 802.5). Polling introduces some overhead, since polling messages are being issued to stations and they might not have data ready to transmit. With short user data packets and many stations associated, polling can be very inefficient owing to high overhead.

The PCF/HCF functions of IEEE 802.11/802.11e are prominent examples for applying polling techniques in wireless networks with central control.

2.6.4 Summary

The performance of the multiple access schemes discussed in this chapter is summarized in Figure 2.21: System throughput versus offered traffic and mean delay versus system throughput are presented. The assumptions made are as follows: ratio of propagation delay to packet transmission duration is $a = 0.01$, time required for generating an acknowledgment is $w = 0.01$, the backoff is drawn randomly from an uniformly distributed time interval of length $k = 10$.

In the left part of Figure 2.21, pure and S-ALOHA lead to the lowest channel capacities of 0.18 and 0.37, respectively. The CSMA variants achieve much better capacities (up to 0.8) and the polling protocol appears optimum introducing an upper capacity bound of 0.95. For implementing an optimal polling-based medium access control protocol, an omniscient central coordinator would be needed able to poll each station just in time, when it has a data packet ready for transmission. When assuming a polling overhead for transmitting poll messages of 5%, the maximum achievable capacity results in 0.95.

In the right part of the figure, the tradeoff between delay and throughput of the CSMA variants considered is visible. A higher maximum achievable system throughput implies higher delay values. The deterministic medium access of optimum polling again establishes the upper bound giving shortest mean delay at highest throughput.

3

Radio Spectrum Regulation

Lars Berlemann and Bernhard H. Walke

The loss of the *Titanic* in April 1912 was a milestone in the introduction of radio spectrum regulation: Many people were rescued due to the reception of the radioed SOS but also lives were lost due to the absence of the radio operator of the nearest ship. This event indicated the relevance of radio communication for public safety and the dangers from unreliable communication (Lucky, 2001). Regulation of the radio spectrum has its origin in the economic regulation of railroads: In the US, the Communications Act 1934 and its predecessors were principally concerned about control of monopoly power when the market is served by a single provider (May *et al.*, 2005). Since the beginning of the nineteenth century, communication services such as telephony, radio and television have been regulated according to this model in order to provide service to the public on a nondiscriminatory basis with fair and reasonable prices and conditions.

From a technical point of view, the radio spectrum is a public resource that can be used without many limitations. Use of the spectrum implies constant interference with neighboring radio receivers sharing the same spectrum. Therefore, radio spectrum regulation is required to allow reliable and efficient spectrum usage. Regulators determine how particular bands of spectrum can be used, make rights available to licensees or unlicensed users and define rules constraining access to this spectrum. The regulators' decision making thereby targets at increasing public welfare and reflects public interest.

In the context of radio spectrum regulation there has to be a distinction between "trading", as transfer of spectrum usage rights and "liberalization", as weakening of restrictions and limitations associated with spectrum usage rights related to technologies and services.

Regulation of the radio spectrum has different characteristics:

- **Licensed spectrum for exclusive usage** is enforced and protected through the regulator. Frequency bands licensed to be used 20 years exclusively for the Universal Mobile Telecommunication System (UMTS) in Europe are such an example.
- **Licensed spectrum for shared usage** is restricted to a specific technology. The frequency channels assigned to Digital Enhanced Cordless Telecommunication (DECT)/Personal Communications Service (PCS) are an example here. The secondary usage of under-utilized licensed

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

spectrum through intelligent radio systems is a different kind of sharing licensed spectrum and will be discussed in Section 13.3.

- **Unlicensed spectrum** is available to all users operating in conformance to regulated technical etiquette or standards, like the Unlicensed National Information Infrastructure (U-NII) bands in the US at 5 GHz, or the Industrial Scientific and Medical (ISM) bands, worldwide, e.g., the band at 2.4 GHz where IEEE 802.11 and Bluetooth are operated.
- **Open spectrum** allows anyone to access any range of spectra without permission under consideration of a minimum set of rules from technical standards or etiquette that are required for sharing spectra.

The report of the Federal Communications Commission (FCC)'s Spectrum Policy Task Force (FCC, 2002) defines spectrum regulatory mechanisms in a similar way. There, the assignment of spectrum rights is differentiated into an "exclusive use" model, a "command-and-control" model and a "commons" or "open access" model. The "command-and-control" is currently the most often used regulation model and refers to the "licensed spectrum for shared usage" and "unlicensed spectrum".

Radio spectrum regulation has to take account of the development of access protocols and standards to balance the following goals (Peha, 2000):

- An adequate Quality-of-Service (QoS) should be possible for all radios depending on the supported applications.
- No radio should be blocked from spectrum access and from transmission for extended time durations.
- Spectrum management policies and standards should not slow down innovations in an economically successful, but rapidly changing, radio communication sector.
- The limited available spectrum should be used efficiently, including spatial reuse of spectrum and solving the "tragedy of commons" described in Section 3.2.
- The spectrum should be used in a dynamically adaptive way, taking the local communication environment such as spectrum usage policies into account.
- The costs of devices should be not increased significantly through techniques prescribed by regulation.

3.1 Regulation Bodies and Global Institutions

3.1.1 International Telecommunication Union

Frequency spectrum allocation is the concern of the Radiocommunication Sector of the International Telecommunication Union (ITU-R). ITU-R ensures the rational, equitable, efficient and economic use of the radio-frequency spectrum by all services using radio communication. ITU agreements on spectrum allocation are set out in the ITU Radio Regulations (ITU RR), which have treaty status. The ITU RR regulate the use of the radio spectrum internationally and form the global framework for regional and national planning (although nations remain sovereign in their use of the radio spectrum in their own territory and Article 48 of the ITU Constitution states that ITU members may retain their freedom with regard to military radio use). The ITU RR are regularly revised through World Radio Conferences (WRCs) that take place every 2–3 years. The next WRC is scheduled for 2007 in Geneva.

3.1.2 Europe

More detailed spectrum planning is conducted at the European level through the Electronic Communications Committee (ECC) of the European Conference of Post and Telecommunications Administrations (CEPT), which currently has 46 member countries. Frequency allocation issues are handled through the Frequency Management Working Group (FM) and there are also committees dealing with Radio Regulatory (RR) and Spectrum Engineering (SE) issues. The ECC structure was significantly changed in October 2001 when the responsibilities of the radio and telecommunications sides of the CEPT, previously handled separately, were combined in a new Electronic Communications Committee (ECC).

Each of the ECC's three main working groups (FM, RR, and SE) also has a number of project teams which are charged with detailed examination of specific areas. The ECC produces Reports, Recommendations and Decisions on spectrum usage. When implemented by member countries, these form the basis for European harmonization of spectrum usage at the allocation level.

The ECC coordinates long-term spectrum planning in Europe and has produced the European Common Allocation Table (ECA), a harmonized frequency table for Europe, covering the entire usable spectrum. The ECC also provides the forum for coordinating European preparations for WRCs through its Conference Preparatory Group (CPG).

The EU has also adopted spectrum harmonization measures. In the early 1990s, three frequency harmonization Directives were adopted, on GSM, DECT, and ERMES. Unlike ECC Recommendations and Decisions, these Directives are binding on member states. In 1992, it was agreed that European spectrum harmonization should normally be carried out through CEPT and that EU member states should commit themselves to full participation in the development and implementation of ERC/ECC Decisions.

In recent decades Europe privatized formerly state-owned telecommunication companies to overcome market entry barriers and to harmonize the member states' laws. The "New Regulatory Framework" targets at establishing conditions for a fair and efficient competition in eliminating sector-specific regulation (EU, 2002a). The regulatory framework assumes a "significant market power" as a basis for regulation in most cases. The framework sees a need for minimum regulation to enable the interconnection of all public communication networks and providers aiming at a single interoperable network. Under this EU framework, the finding of a significant market power permits the national regulators of the member states individual spectrum regulation. The "Radio Spectrum Decision" (EU, 2002b) aims at the harmonization of conditions under consideration of availability and efficient use of the radio spectrum necessary for the establishment and functioning of internal markets such as electronic communications, transport and research and development.

3.1.3 Germany

In Germany the Bundesnetzagentur (BNetzA) formerly referred to as Regulatory Authority for Telecommunications and Posts (RegTP) is responsible for spectrum regulation. The emphasis of Germany's frequency regulation in the next few years is presented in Bundesnetzagentur (2005): The status quo and strategies for the future are introduced for many frequency bands. This includes for instance the global harmonization of frequencies assigned to WLANs and a frequency refarming. The future of UMTS/GSM/DECT frequency bands, frequency trading, and strategies for UWB applications are planned for a future revision of this document. The BNetzA targets at a transparency of future frequency regulation and aims at giving guidelines for future research, innovation and investment decisions.

3.1.4 Japan

In Japan in 2004, the Ministry of Public Management, Home Affairs, Posts and Telecommunications (MPHPT) changed its name to the Ministry of Internal Affairs and Communications (MIC). The MIC is responsible for planning, designing, and promoting general policies for using communication devices and is in charge of frequency allocation. The Association of Radio Industries and Businesses (ARIB) is designated by the MIC as "the Center for Promotion of Efficient Use of the Radio Spectrum" and "the Designated Frequency Change Support Agency". The ARIB conducts studies, establishes standards, and provides consultation services for radio spectrum coordination with overseas organizations. Currently the ARIB has 277 members from the field of telecommunications, broadcasting services and research, development and manufacturing of radio equipment. The Telecom Engineering Center (TELEC) provides services related to technical regulation conformity certification under designation of the MIC.

The MIC has identified a bandwidth of at least 1.5 GHz below 6 GHz by 2013 as requirement for future needs in its *Radio Policy Vision* (MIC, 2003). There, a review of frequency assignments, establishment of rules enabling a fast as well as easy reallocation of the spectrum, and a reconsideration of the fee system for using the spectrum has been identified as cornerstones of future spectrum management. The MIC is establishing a legal basis for a compensation system in spectrum reallocation. The economic benefits for new spectrum users are used to collect an additional usage fee in order to compensate the economic losses incurred for the incumbent radio system. The freeing of the 100 MHz band from 4.9 to 5.0 GHz for outdoor WLAN in metropolitan areas from being licensed for fixed microwave stations for telecommunication business by 2005 is a first example for implementing such a compensation system.

3.1.5 China

In China, the radio spectrum is regulated on two levels: The Ministry of Information Industry (MII) and its Radio Regulatory Department are responsible for nationwide regulation, while provincial radio regulation institutions realize state regulation and administrate regional regulation. Additionally, China has the Army Radio Regulatory Commission which regulates spectrum usage in the military systems. The MII is technically supported by the China Radio Monitoring Center, which supervises spectrum usage while looking for interference and unauthorized radio stations.

Spectrum regulation in China is restrictive and driven from interests to protect national economic markets. For operation of telecommunication equipment, the MII issues Network Access Licenses. Without such a license no operation is allowed. In the recent years, the MII has not followed international efforts of harmonizing standards and regulation, and has rather favored individual solutions. There was for instance an attempt in 2003 that all Wi-Fi products require a Chinese encryption standard. Nevertheless, China is separating its administrative and enterprise functions related to the spectrum to enable competition and to break monopoly.

3.1.6 United States

In the US the responsibility for spectrum management is shared between the FCC and the National Telecommunications and Information Administration (NTIA). The NTIA is responsible for the spectrum used by the government and the FCC is responsible for the spectrum used for non-government purposes.

The Spectrum Policy Task Group of the FCC has identified three approaches to the improvement of spectrum utilization (FCC, 2002): (i) improve access in the space, time and frequency domains: (ii) enable flexible regulation in permitting controlled access to licensed spectra; and (iii) stimulate efficient spectrum usage through policies. Therefore, the FCC is rethinking its licensing policy. This can be seen in the fact that no major spectrum allocations for licensed usage have been released since the assignment of frequencies for PCS in 1993. On the other hand the FCC is initiating multiple approaches to liberalization of spectrum usage. Thereby, market orientation and flexibility is the target to provide opportunities for new technologies and to stimulate investments. The additional unlicensed spectrum at 5 GHz (FCC, 2003a), the new unlicensed band at 3.7 GHz (FCC, 2004a) and the opportunistic usage of TV bands (FCC, 2004c), as depicted in Figure 13.7, are examples of this. Additionally, the FCC is also focusing on cognitive radio technologies for flexible and reliable spectrum usage (FCC, 2003c, 2005a) and the Spectrum Policy Task Group of the FCC introduced the Interference Temperature Concept for underlay spectrum sharing (FCC, 2003b) as outlined in Section 13.4.3.

3.2 Licensed and Unlicensed Spectrum

3.2.1 Licensed Spectrum

Large parts of the radio spectrum are allocated to licensed radio services in a way that is often referred to as command-and-control. The licensing spectrum covers exclusive access to the spectrum and spectrum sharing within the licensed spectrum.

In the case of exclusive spectrum usage, a license holder pays a fee to have this privilege. Exclusive access rights have the advantage of preventing potential interference, which implies dangers to reliable and thus chargeable communication. In the case of spectrum scarcity, a licensed spectrum is highly valuable leading to economic profits, as consumers need to pay to use it. Having an immense commercial impact, spectrum licenses can be bounded to requirements such as a concrete transmission technology or a certain percentage of population to be reached by the network using the spectrum for wireless communication. The auctions in countries of Europe for licensing spectrum to 3G systems in 2000 are an example.

Today's most often used licensing model is to license spectrum for shared usage restricted to a specific technology. Emission parameters such as the transmission power and interference to neighboring frequencies such as out-of-band emissions are restricted. Regulation takes care of protection against interference and sometimes supports limited coexistence capabilities such as Dynamic Channel Selection (DCS) in DECT, which are mandatory and part of the standard.

3.2.2 The Problem with Licensing

Licensing spectra takes a lot of time and is very difficult, therefore licensing is also very expensive. The licensing process constrains innovation, as it is a barrier difficult to overcome when introducing new technologies. The inflexibility of exclusive usage rights from licensing spectrum leads to inefficient spectrum utilization, as the license prohibits the usage of the spectrum if it is underutilized or even unused by the license holder.

In 1959, R. H. Coase (Nobel prizewinner in economic science in 1991) criticized spectrum allocation by the government and spectrum licensing in the "public interest" (Coase, 1959). He suggested the issuance of clearly delineated rights of spectrum ownership as a more efficient method of allocating the spectrum to users. Coase observed that "government control" is not required for economic development of spectrum usage but actually constrains it. This led to the more general "Coase Theorem" which shows that well-defined property rights and moderate trading costs for these rights lead to an efficient allocation of resources (Coase, 1960). Coase is therefore often referred to as the "father" of reforming spectrum policy.

Another problem with licensing is the duration of a license. Typically licenses in the US expire after 10 years but can be renewed. Temporal licenses give the regulator the possibility to intervene if the spectrum is underutilized or wasted. The regulator is able to answer market demands in shifting, extending or reissuing licenses and it can in this way accelerate the introduction of new technologies. The danger of temporal licenses on the other hand is that uncertainty about future regulatory decisions may constrain investments (Peha, 1998).

Licenses are often issued based on auctions. The advantage or disadvantage of this procedure has been discussed in economic science for many years: Auctions for selling spectrum access rights (referred to as trading at the primary market as introduced below) are primarily a revenue tool for governments and do not reflect social values. The one-time payments from auctions are often used from shortsighted governments to pay for annual expenditures. For a discussion on auctions of licenses for using spectrum, the reader is referred to Hazlett (1998) and Noam (1998).

3.2.3 Unlicensed Spectrum

Unlicensed spectrum is available to all users but is strictly regulated. An unlimited number of users share the same unlicensed spectrum. Spectrum usage is allowed to all devices that satisfy certain technical rules or standards that mitigate potential interference resulting from transmission powers or advanced coexistence capabilities. The usage rights of unlicensed spectrum are flexible and no concrete spectrum access method is specified.

In general, two main frequency bands can be identified that are designated for unlicensed operation: The Industrial, Scientific and Medical (ISM) bands from 2400–2483.5 MHz and the frequency band between (and below) 5 GHz and 6 GHz. The frequency band at 5 GHz is referred to as Unlicensed National Information Infrastructure (U-NII) band. While the regulatory restrictions for the ISM bands are similar throughout the world, spectrum regulation of unlicensed operation in the 5 GHz band differs essentially when comparing Europe, the US and Japan as depicted in Figure 3.1(a).

3.2.3.1 Europe

A general overview for the regulation of using short-range devices in Europe can be found in ERC (2005). The regulations for the usage of the 2400–2483.5 MHz band are described in ERC (2001b): The Equivalent Isotropic Radiated Power (EIRP) is limited there to 100 mW. Further, for Direct Sequence Spread Spectrum (DSSS) based transmission, the maximum spectrum power density is limited to -20 dBW/MHz and for Frequency Hopping Spread Spectrum (FHSS) the maximum spectrum power density is limited to -10 dBW/100 kHz. The operation in the 5 GHz frequency bands is regulated in ECC (2004): The 5150–5350 MHz band is designated for indoor usage with a mean EIRP of 200 mW and DFS together with TPC are additionally required for unlicensed operation above 5250 MHz. The frequencies from 5470–5725 MHz may be used indoors as well as outdoors and the mean EIRP is limited in this band to 1 W. The use of DFS and TPC is mandatory.

Compared to the US, the multitude of different regulation authorities present in Europe results in a slower liberalization process of radio spectrum regulation.

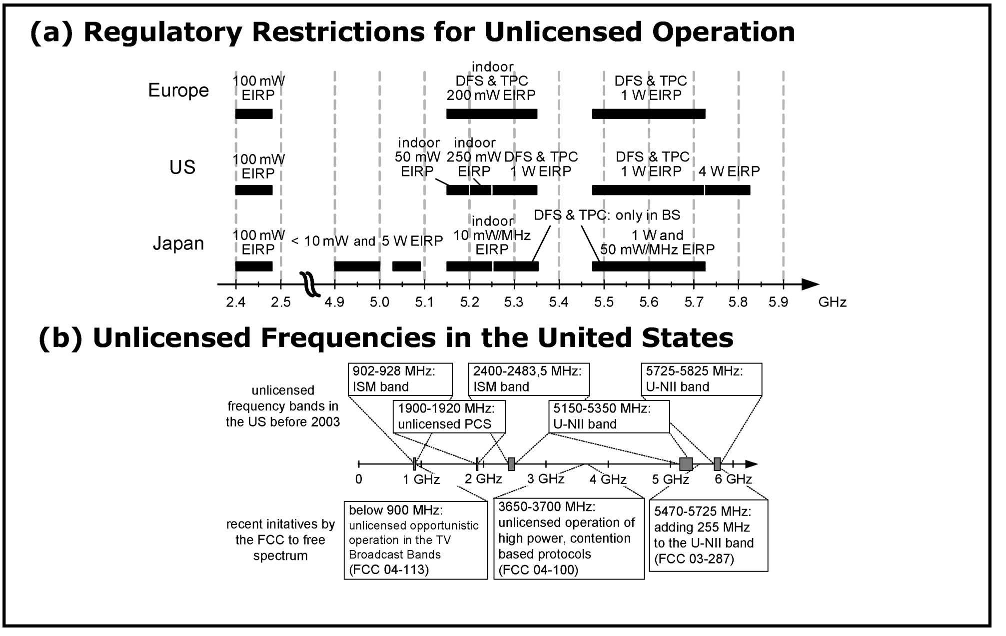


Figure 3.1 Frequency bands and regulatory restrictions of unlicensed operation (a) and available spectrum for unlicensed operation in the US (b). Reproduced by permission of © 2006 IEEE[1].

3.2.3.2 United States

Figure 3.1(b) illustrates the allocation of unlicensed spectrum in the US. Besides the ISM bands at 900 MHz and 2.4 GHz, the FCC opened in the US for unlicensed usage in 1990 during PCS rule making 20 MHz at 1.9 GHz for Unlicensed PCS (UPCS). Additionally, the FCC reserved in 1997 300 MHz and in 2003 255 MHz at 5 GHz for unlicensed operation. Contrary to the ISM bands, use of the U-NII band is more restricted: There, limited coexistence capabilities such as DFS and TPC, as introduced below, are also mandatory (FCC, 2003a).

As TV bands in the US are often underutilized, the FCC proposed in 2004 to allow unlicensed systems the secondary usage of this spectrum (FCC, 2004c). This principle of vertical spectrum sharing is introduced in detail in Section 13.3.3.

In 2004, the FCC also initiated opening new spectrum for wireless broadband communication in the 3650–3700 MHz band for fixed and mobile devices transmitting at higher power (FCC, 2004a). It is envisaged that multiple users will share this spectrum through the use of "contention-based" protocols to minimize interference among fixed and mobile operations. New fixed and mobile stations will therefore be required to use contention-based protocols, which will reduce the possibility of interference from co-frequency operation by managing each station's access to the spectrum. The FCC regards this approach as a reasonable, cost-effective method for ensuring that multiple users can access the spectrum. Besides a few regional constraints, at radar sites and frontiers of the US, fixed stations will be allowed to operate with a peak power limit of 25 watts per 25 MHz bandwidth, and mobile stations with a peak power limit of 1 watt per 25 MHz bandwidth. For further details on the FCC's understanding of contention-based protocols in the context of IEEE 802.11y see Section 13.5.3. The licensing, service, and operation provisions for this spectrum will be placed in Part 90 of the FCC's rules taking the non-exclusive nationwide nature of the spectrum into account. The status of this frequency band is currently subject to intensive lobbing from different protocol fractions in the wireless industry.

3.2.4 Part 15 Regulation

The Part 15 rules of the FCC (FCC, 2005c) describe the regulations under which a radiator may be operated without an individual license. It also contains the technical specifications and administrative requirements of Part 15 devices. ISM low-power devices such as garage openers for instance are allowed to transmit at 1 W if using spread spectrum technologies. Three basic principles describe in general the rules of Part 15 regulation: "listen before talk", "when talking, make frequent pauses and listen again" and "don't talk too loud". When detecting a busy channel, either another unused channel is chosen or the radio waits until the channel is idle again. These simple etiquettes require no interoperability or information exchange between spectrum sharing devices. IEEE 802.11, as being designed for operation in frequency bands subject to Part 15 regulation, realizes spectrum access corresponding to these three basic principles. The analysis of the QoS capabilities of 802.11 (Mangold, 2003) shows that with the current Part 15 regulation a QoS support is impossible under mutual interference of multiple coexistent 802.11 systems and an additional distributed coordination is required to allow QoS support. The accidental background noise emitted from consumer electronics such as personal computers operating at 2–3 GHz is also restricted in the Part 15 regulation.

3.2.5 Tragedy of the Commons in Spectrum Regulation

The success of unlicensed spectrum draws to a close, as the severe QoS constraints to spectrum access imposed by the upcoming multimedia applications cannot be fulfilled by today's means available for supporting coexistence (Mangold, 2003).

In the case of short-distance wireless communication spectrum demand is extremely localized and often sporadic. In such a scenario, the competition for shared spectrum is limited. Therefore the regulatory instrument of restricting transmission, e.g. limiting the maximum emission power, is sufficient.

In all other scenarios, as for instance WLANs, unlicensed spectrum usage is a victim of its own success: Too many parties and different technologies are using unlicensed spectrum so that it is getting overused and thus less usable for all. In economics the phenomenon is referred to as the "tragedy of the commons". Hazlett (2005) additionally introduces the "tragedy of the anticommons": Contrary to the overuse of spectrum due to missing regulations of access, the "tragedy of the anticommons" refers to inefficient spectrum utilization because of too restrictive regulation. The "tragedy of the commons" and the associated inefficient overuse of spectrum results in underinvestment in technology and questions "open access". Therefore, to anticipate the "tragedy of the commons", regulators impose restrictions such as transmission power. As a consequence, many alternate systems are not allowed to operate in such a spectrum, which again results in inefficient underutilization of the spectrum. In Hazlett (2005) it is concluded, that limiting spectrum sharing in spectrum regulation is the only way out of this tragedy. Section 13.7.4 discusses this statement and indicates some better alternatives.

3.3 Open Spectrum

Open spectrum allows anyone to access any range of spectrum without permission under consideration of a minimum set of rules from technical standards or etiquettes that are required for sharing spectrum (Weinberger *et al.*, 2005). Open spectrum targets at the complete liberalization of radio communication in overcoming regulatory roadblocks. The core concept of open spectrum is that technologies and standards are able to dynamically manage the access and spectrum sharing, replacing the static spectrum assignments resulting from bureaucratic licensing (Berger, 2003).

New, upcoming radio transmission technologies suggest that radio spectrum can be treated as an open access commons rather than a collection of stringed-together access rights (Noam, 1998). It is correct to say that just squeezing more wireless communication out of a given bandwidth does not solve the tragedy of commons (Hazlett, 1998). But rules and etiquettes on the one hand and technologies allowing more than one user simultaneously using the same frequency on the other hand enable open spectrum. Thus, the spectrum differs from other public resources such as highways that are overcrowded by too many cars.

Three primary technologies, which have seen great development progress in recent years, can be identified to realize open spectrum:

- Low-power, ultra-wideband underlay spectrum usage (see Sections 13.3.2 and 13.4.2).
- Cognitive, frequency-agile radios based on SDR (see Section 13.1).
- Cooperation-based, self-configuring meshed networks (see Section 13.2).

As open spectrum considerably impacts the way spectrum is regulated, regulation authorities have to face new challenges. The promising advantages of open spectrum impose a fundamental rethinking of regulating spectrum (Werbach, 2002):

- The scarcity of spectrum is overcome: A dynamic spectrum sharing through cooperative techniques improving efficiency makes regulatory limitation of spectrum access obsolete.
- The competition through innovation while sharing spectrum as a common resource is in many cases superior compared to auctioning licensed spectrum through the market.
- Investment costs for exploiting spectrum are essentially reduced as no tremendous fees for temporal licenses are required resulting into essentially lower service costs.
- Wireless broadband technologies are a solution for the last-mile bottleneck. The special challenges of the last-mile access can be solved in using long-range communications, wideband underlay or meshed architectures.
- Wide-area 3G systems can be complemented by more efficient short-range and meshed unlicensed technologies.

Noam (1997, 1998) approaches open spectrum under the term “open access”. There “open access” allows many users from different radio-based applications to access the spectrum without license by buying access tickets, referred to as “access tokens”, whose prices depend on congestion. Transmitted data packets piggy-backed carry these electronic tokens and a decentralized payment is possible at tollgates such as access points. Such a system would reduce infrastructure costs to a marginal value, reduce market access barriers and encourage competition. Contrary to current unlicensed systems, which depend on etiquettes to manage competition, such a system guarantees access under congestion conditions, as individual values of the spectrum is reflected in the price (Noam, 1997).

In the literature, open spectrum has sometimes a less revolutionary meaning when referring to the vertical sharing of licensed spectrum (Zhao, Q. *et al.*, 2005) as introduced in Section 13.3.3.

3.4 Summary

The availability of a sufficient amount of radio spectrum is crucial for the economic success of future radio communication systems. The triumph of wireless technologies operating in unlicensed frequency bands, especially of the 802.11 standard family, indicates the economic advantage of unlicensed spectrum. The scarcity of free available spectrum on the one hand and underutilized licensed spectrum on the other hand leads to a rethinking of spectrum regulation in general.

National regulation authorities indicate their willingness of liberalizing the regulation of radio spectrum. However, the concrete realization of liberalization differs extremely when comparing the approaches and public statements of national regulation authorities.

The reliable operation of wireless communication systems in unlicensed frequency bands is a central issue in this book and will be discussed in several chapters. Approaches to coexistence, interworking, and spectrum sharing are introduced in the context of 802 standards in Chapters 5–7.

The potential and current trends in the field of intelligent radio communication systems capable of efficiently using spectrum (either unlicensed or underutilized licensed spectrum) are discussed under the umbrella of cognitive radio in Chapter 13.

Note

1. Reproduced by permission of © 2006 IEEE. Source: L. Berlemann, C. Hoymann, G. R. Hiertz, and B. H. Walke, "Unlicensed Operation of IEEE 802.16: Coexistence with 802.11(a) in Shared Frequency Bands," in Proc. of 17th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2006, Helsinki, Finland, 11-14 September 2006. (Fig. 3.1).

4

Mesh Networks – Basics

Guido Hiertz, Erik Weiss and Bernhard H. Walke

Wireless networks have become ubiquitous. With cheap and reliable products Wi-Fi and Bluetooth (Bluetooth, 2006) have created and developed new mass markets. Similar to the evolution of wired networks, current wireless networks form isolated communication islands without any interconnection between them. In market-relevant wireless technologies, network control is often centralized. Bluetooth (IEEE 802.15.1), IEEE 802.16 (aka WiMAX) and Wireless Personal Area Networks (WPANs) according to IEEE 802.15.3 standard form star topologies with a Central Controller (CC), Base Station (BS) or Piconet Controller (PNC) in the center.

At present most of the deployed IEEE 802.11 Wireless Local Area Networks (WLANs) operate in infrastructure mode where a central Access Point (AP) is present. Although channel access in such configurations is decentralized, all traffic in the network flows via the AP, see Figure 4.1. In contrast in the sporadically used IEEE 802.11 ad-hoc mode, stations send their traffic directly to the target destination, which must be within the ad-hoc network. The wireless networks proposed by the WiMedia group (see Section 6.7) and a recent proposal for a mesh-distributed coordination function (see Section 12.2) are the only ones that operate under decentral control. All others use a single central node that takes care of traffic relaying to destinations in and out of the local network. Figure 4.1 provides examples for IEEE 802.11 WLAN and IEEE 802.15 WPAN.

In its current state all existing wireless standards need bridging (relaying in layer 2) or routing (relaying in layer 3) functionality to connect with other networks that may be based on wire or be wireless. Transparent in-band bridging is not foreseen. As a common way to bridge data in current IEEE 802.11 networks most existing WLAN APs provide an Ethernet port to interconnect the WLAN segment with the wired IEEE 802.3 segment. Figure 4.2 provides an example. The wired network provides connectivity to other APs. Data can be forwarded from the source to the final destination with the APs working as bridging devices that use the wired network for frame exchange.

Products that use radio to forward (relay) data between different 802.11 Basic Service Sets (BSSs) work similarily as shown in Figure 4.2, with the wired link replaced by a wireless link. However, the function to establish an Extended Service Set (ESS) is not part of any APs of the interconnected BSSs, since it is not part of the 802.11 standard.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

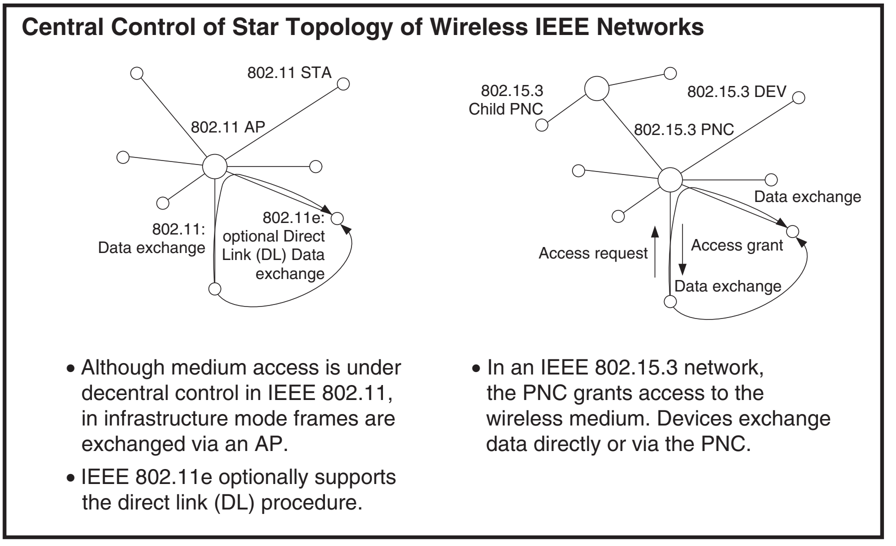


Figure 4.1 All existing IEEE standards as the most used option apply central control of star topologies, although medium access is decentrally controlled and frames are relayed between stations (STA) via a central node (AP, PNC).

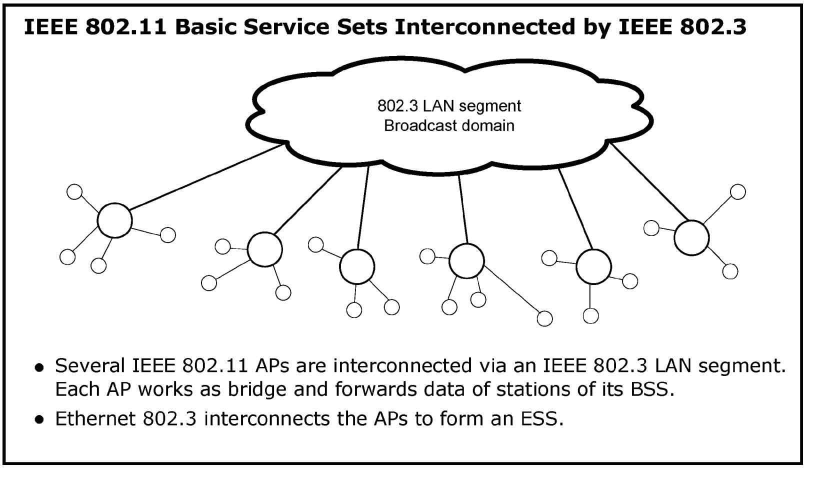


Figure 4.2 The current IEEE 802.11 standard makes use of bridging capability to interconnect several APs and their BSSs.

4.1 Introduction

The basic characteristic of a wireless mesh network is the capability to relay frames from one device to another. Figure 4.3 shows an example of the coverage extension of Internet access via relaying devices. To be able to relay data from a source device to the ultimate destination device, sufficient address information must be provided. In IP networks, the network address is used to

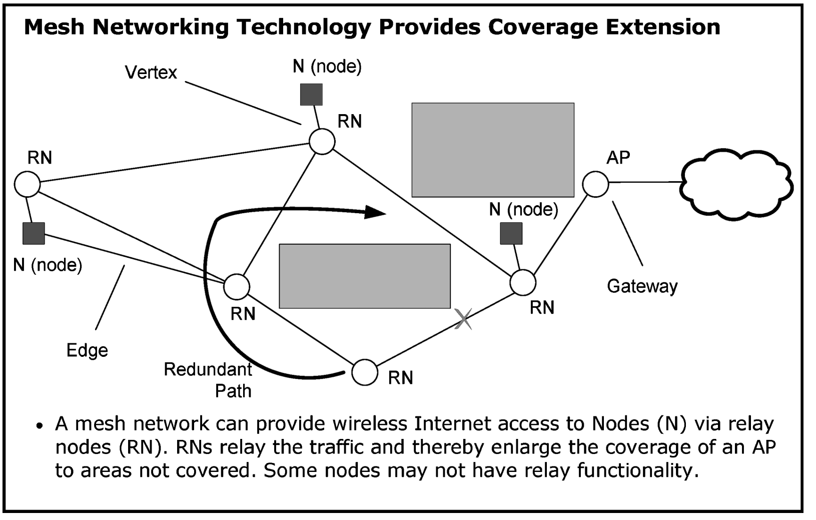


Figure 4.3 Wireless mesh networks extend the reach of an AP/gateway, provide easy deployment and redundant paths. Nodes in the mesh may be referred to as vertex, while links form edges of a graph representing the mesh topology.

forward data multi-hop from source to destination. IP routers exchange information on their attached networks and advertise known routes. Routers are interconnected and provide the relaying service for devices and in their attached networks. In an IP network, the source device requests relaying of a frame by its local network serving router (a Relay Node, RN) – also referred to as the gateway. Several routers along the routing path forward the frame. At the final router, the frame is delivered to the destination inside an attached local network. In an IP network, devices use the network mask to identify devices outside the local network. Figure 4.4 shows a single router that interconnects several IP networks. Routing decision are based on network addresses.

For destinations inside the local network, direct frame exchange is possible. The local network forms a subset of all existing devices. Thus, devices are able to communicate with a subset of all devices only. To mutually address each other, devices use broadcast messages in a local network. Hence a local network is also referred to as a broadcast domain. Within a broadcast domain, each device may communicate with other devices without the help of any intermediate node. For destinations outside the broadcast domain, devices rely on routers to forward their frames. Hence, routers represent a set of devices. In terms of addressing, they work as proxies for their attached networks. A set of devices can be addressed through a single router. This hierarchy of relaying-capable and non-capable devices helps to keep the routing tables small.

In contrast to single-hop networks, where usually most of the traffic is directed to and received from a central device, mesh networks, potentially, have no hierarchy. The wireless medium is a shared resource that is used by all entities of the mesh network. In some cases the wireless medium may even be shared with non-mesh capable devices that are also served by the mesh network. Other neighboring or co-located non-mesh networks may be present, especially, in WLAN and WPAN environments, which usually use unlicensed frequency bands. The common resource that is shared among the devices participating in the mesh network is therefore a harsh environment.

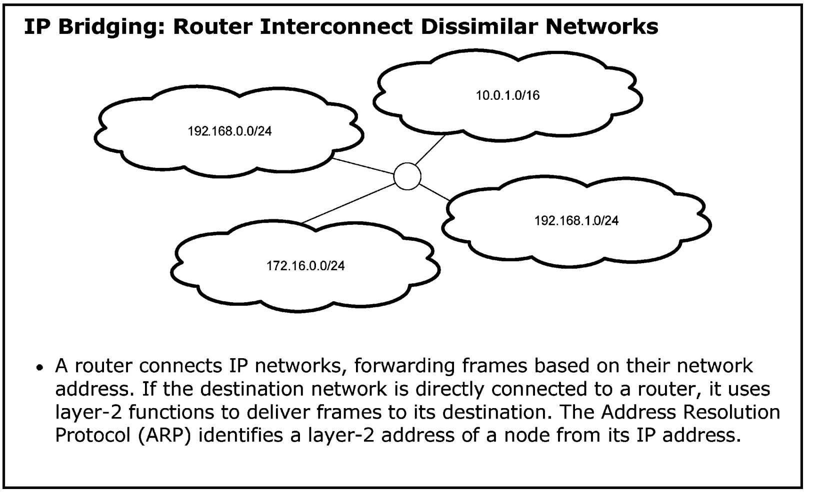


Figure 4.4 Devices use routers as default gateway for any device they cannot communicate directly with by means of layer-2 functionality. This hierarchy enables the Internet. Using layer-2 addresses only would be impossible to handle.

As most wireless technologies solely define layer 1 and 2 of the ISO/OSI Reference Model, their topology is flat. Therefore, wireless standards usually define a single broadcast domain only and no routing function is defined. Any frame relaying needs to be handled by higher layer protocols. While in traditional wireless single-hop networks all devices are in either mutual reception range or have a common central neighbor (the AP), in wireless mesh networks multiple direct and indirect neighbors may exist that do not necessarily have an intersection of their sets of neighbors. The mesh supporting the Medium Access Control (MAC) protocol needs to take this into account. Furthermore, a wireless mesh network introduces multi-hop links inside the broadcast domain. To enable higher protocols to work transparently over a wireless mesh network, routing needs to be handled by each relay device in the broadcast domain. The identification of possible hops from source to destination is called routing in the IP layer. To distinguish routing from the respective function in the mesh-able MAC layer, in the following it is referred to as *path selection*. However, the basic function remains the same: Devices determine their neighbors and propagate the information in the network to be able to relay frames.

Mesh networks may consist of devices mutually unknown to each other. Those devices may mutually provide services in terms of path selection and frame forwarding. Therefore, security support in wireless mesh networks is more complex than in single-hop networks. Trustful relationships between devices may not always exist. Hence, end-to-end security support differs from single-hop link security. If unknown devices participate in the mesh network, path selection may become impossible. Invalid path information may be provided by attackers; hence frames may be relayed to false devices. For highest security, military services use mesh networks, where each device floods each received frame. In all wireless mesh networks, a hop counter may prevent infinite frame forwarding and loops.

Security, path selection and MAC adaptation are the most important elements to consider when designing a wireless mesh network. The following sections will give an introduction to these topics.

4.2 Classification of Wireless Mesh Networks

Wireless mesh networks may operate with or without a hierarchical structure. In a flat hierarchy, any wireless device in the network is able to forward frames. In such networks, a device does not solely operate as a sink or source of traffic, but may accept packets that are not directed to itself to be relayed to neighbor devices. Each device in such networks needs path selection functionality and the capability to support multi-hop traffic.

In hierarchical mesh networks, only mesh-able devices provide the mesh networking service to other non-mesh-able devices that do not have relaying capabilities. Non-mesh-able devices associate with the mesh-able devices. Typically, mesh-able devices are APs. A hierarchical mesh network, especially, is sufficient for static mesh networks, where fixed APs form the backhaul mesh network to provide ESS service for mobile client devices. Only the mesh-able devices need extra resources such as memory, computing power and multiple transceivers to be able to operate the wireless mesh network, besides the BSS. As APs are fixed and connected to the mains, power-saving algorithms are not a concern, this is different from mobile devices that need to optimize use of their battery power. In addition, location aware packet relaying protocols may be applied to exploit the fixed nature of the network.

With regard to the frequency channels used, wireless mesh networks with respect to the mesh function in comparison to the BSS support function may operate in band or out of band. Wireless mesh networks may operate on single or multiple frequency channels. In a single channel mesh network, single-hop frames (inside a BSS) as well as multi-hop frames (in an ESS) travel in the same channel. Coexistence support is necessary then, and fine traffic segregation is needed to provide the mesh network with the necessary resources to relay frames generated remotely when competing with frames that travel locally to a BSS.

Multiple channels may be operated using a single or multiple radios in mesh-able devices. Then, traffic segregation is possible, where single- and multi-hop frames may travel on different

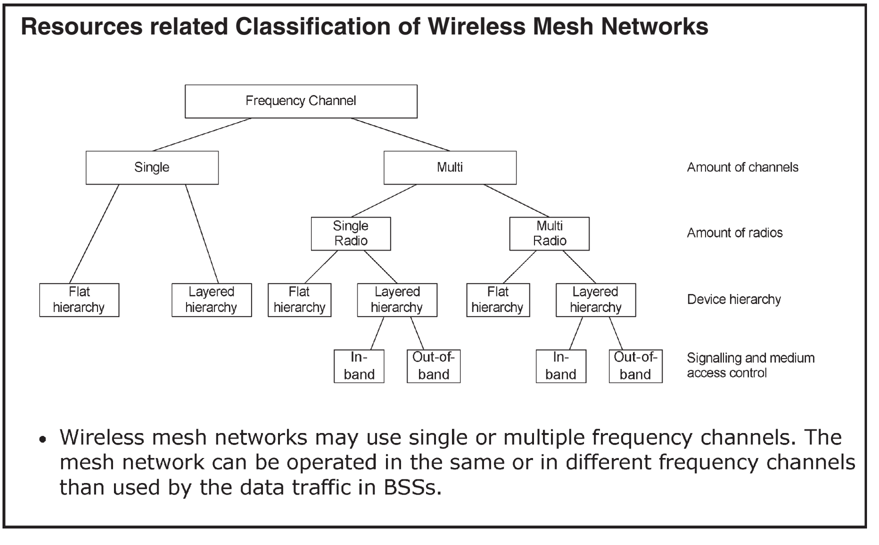


Figure 4.5 Characteristics of wireless mesh networks vary depending on the available resources. Resources are the wireless medium and hardware capabilities. Protocols for wireless mesh networks need to consider these resources.

frequency channels. Separation by operating BSSs on channels different from that used for meshing (ESS) may be inferior in terms of overall capacity to dynamic channel assignment or even sharing common channels. Figure 4.5 provides a classification of wireless mesh networks based on the numbers of channels and radios used and the way of how to share channels for BSS and ESS services. Cellular radio networks applying mesh are discussed in Chapter 10.

Another classification of mesh networks may be derived from the MAC protocol used. IEEE 802.11 applies an asynchronous medium access under decentral control, while IEEE 802.16 and 802.15.3 are based on synchronized medium access.

4.3 General Problem Statement

New phenomena emerge from wireless mesh networks. In contrast to single-hop networks, transmission of a frame multi-hop reduces the end-to-end data throughput while the overall latency/delay increases. Besides that, problems resulting from relaying, not known in single-hop networks, severely impact the performance of wireless mesh networks. While routing algorithms are well known from wired networks, self-interference of relayed frames and unpredictable path metrics are the new challenges.

4.3.1 Path Selection

Bellman (1958), Dijkstra (1959) and Floyd (1962) provide generic routing algorithms that form the basis of most existing routing protocols in wired networks. Example metrics used by these protocols to calculate the optimum route are hop count, link speed, cost for transiting traffic, and delay. These are being used to weight edges when applying graph theory.

Since the bit error ratio and its fluctuation are of minor concern in wired and optical networks, data rate and delay tend to be relatively constant, compared to route updating time intervals or frame transmission duration. Consequently, routing algorithms in the wired Internet do not take frequent changes of topology and link speed into account.

With wireless networks, topology (connectivity of nodes) and link speed change quickly. Roaming devices and moving obstacles cause frequent topology changes in both infrastructure-based and ad-hoc wireless mesh networks, causing changes in load of relay nodes and mutual interference of network internal nodes and with nodes of foreign networks. Path metrics of wired networks appear insufficient for wireless mesh networks. Vertices (wireless devices) and edges (links between wireless devices) cannot be considered very stable in wireless networks resulting in frequent change of the topology. Contrary, in wired networks, status, availability and characteristics of vertices and edges change slowly, and so does the routing graph. Depending on the network size and application, routing graphs in wired networks have long periods of stability ranging from hours to months or even years. In wireless networks such stability cannot be achieved.

Path metrics for mesh networks providing more accurate path selection decision may consider in addition:

- Packet error probability that depends on SINR reflecting the current PHY mode used for a given link, antenna gains, transmission power, background noise level, and frame length used by the MAC protocol.
- Congestion status of receiving relay node in the mesh network.
- Availability of relay node on a certain frequency channel.
- Bandwidth needed for transmission.

In wireless networks the path selection metrics mentioned are all time variant and may essentially change within a short duration. Therefore, to calculate the optimum path at time t_0, estimates of the parameter values of each metric are required that can hardly be made. Furthermore, the respective information may be available in the MAC layer only. Existing standards do not provide an interface to support information exchange with the routing layer to provide these parameters. Hence, ad-hoc path selection (routing) in wireless mesh networks must operate blindly.

Different from the working assumptions made by the Internet Engineering Task Force (IETF) group "Mobile Ad-hoc Networks" (IETF, 2006b), wireless mesh networks developed at the IEEE cover only layers 1 and 2 and must provide transparency to higher layers. The ad-hoc routing protocols developed at MANET cannot be used in wireless mesh networks, since frame forwarding is performed in the IP layer. The IEEE aims at path selection (routing) protocols realized in the MAC ("layer 2.5") to provide multi-hop forwarding of uni-, multi- and broadcast frames in the MAC layer. The wireless mesh network is considered a single LAN segment that forms a unified broadcast domain.

The ad-hoc routing protocols for wireless mesh networks developed by MANET of the IETF reside in the IP layer. Since no interfaces for parameter values exchange exists within the MAC layer, routing decisions are based on a small set of routing metrics. Since the IP layer lacks these metrics for decisions on the preferable paths, the MANET routing protocols use frequent IP broadcast frames to exchange topology information between the relay nodes involved, where IP (layer 3) broadcast frames are mapped onto MAC frames with the receiver field set to the broadcast address. Inside the broadcast domain, which is limited by the actual transmission range of the broadcasting device, other devices are periodically being informed about the transmitter's routing tables and its topology view. For more details see Section 4.6.

4.3.2 Medium Access Control

Different from single-hop networks, relaying in multi-hop networks introduces new problems that cannot be solved by just applying single-hop MAC protocols multiple times in sequence. A wireless mesh network may be seen as the sum of a number of continuously overlapping neighboring single-hop networks. In single-hop wireless networks all devices in the network are either in mutual receive range or have a common central device that is in receive range of all other devices. However, a wireless mesh network provides frame exchange among devices that are not in mutual receive range. The source and the final destination nodes may not be able to exchange information directly. Hence, coordination of their channel access in an area larger than that of a single-hop network is needed. The hidden and exposed node problems, when not handled properly, may cause much more severe problems in wireless mesh networks than with single-hop networks and therefore are discussed in subsequent sections.

4.4 Exploiting the Capacity of the Radio Channel by Spatial Reuse[1]

A frame directed by a node to a non-direct neighbor must be relayed, thereby increasing the spectrum load and the overhead. Each relay node operates for each relayed frame as both receiver and transmitter. Hence, the wireless medium in the vicinity of a relay node is occupied once for frame reception and a second time for frame transmission, during which times the relay node may not transmit or receive its own or other frames. When assuming a string topology of equidistant nodes, under reciprocal channel propagation conditions, a spatial frequency reuse distance can be

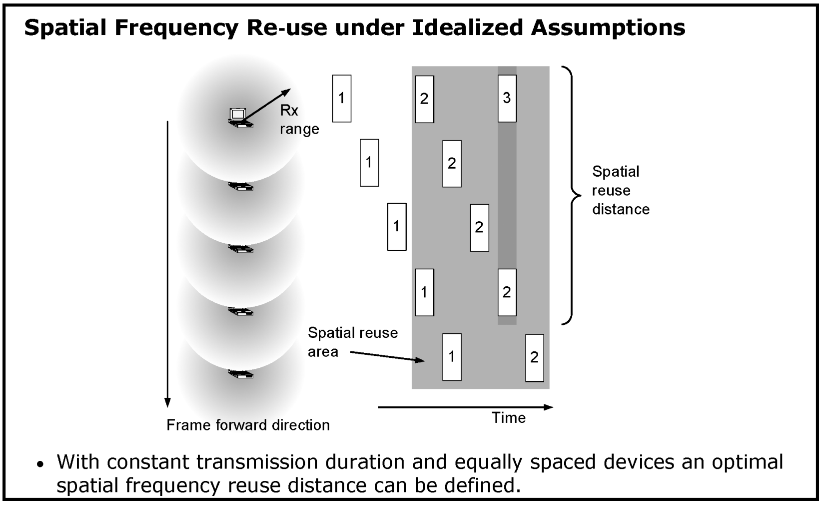


Figure 4.6 Spatial frequency reuse is an important aspect for wireless mesh networks. Multiple devices in mutual receive range aim at a highly efficient spectrum usage. Spatial frequency reuse, even when applied to a multi-hop link, increases spectrum capacity.

easily explained, see Figure 4.6. There, only devices in mutual neighborhoods may receive each other. Assuming free-space radio propagation, the interference range may be assumed to be less than twice the distance between two neighboring devices. Thus, the minimum spatial frequency reuse distance is three hops.

Real deployments of relay nodes in wireless networks are not regular and not limited to two dimensions. A comprehensive overview on the wrong assumptions frequently made when discussing wireless networks is presented in Kotz *et al.* (2003). Accordingly, WLANs cannot be easily described. Mesh network topologies add even more complexity. Indirect neighbor devices are a major source of heavy interference, whose influence on the actual channel conditions cannot be easily determined or even predicted. Since the actual SINR value determines success or failure of a frame reception, interference prediction is crucial to be able to exploit spatial frequency reuse. Without spatial reuse, capacity is wasted. The assumption of a fixed interference range is imprecise as in real-world scenarios attenuation highly depends on obstacles in the direct and indirect propagation paths, which may be of any nature, e.g., walls, doors, humans, buildings.

In the following a basic wireless mesh network is considered to comprise several fixed devices all sharing the same channel for frame exchange, with direct and indirect neighborhoods of each device being time-invariant. We allow moving obstacles to introduce some time dependency of radio propagation. The set S_n of devices d_m with $n \neq m$ and transmit power $P_n(m) > P_{\min}(R)$ describes all devices that can receive transmissions from device d_n. Assuming reciprocity of radio propagation, d_n is part of set S_m if device d_m is part of S_n. Then S_n is the set of devices which d_n can directly exchange frames with. Channel reciprocity is mostly applicable, but not always given. For any transmission of a device d_m that is part of set S_n to d_n, other simultaneous power emissions by devices $d_k, k \notin (n, m)$ on the same channel do not interfere if

$$\frac{P_n(m)}{P_{n,\text{Noise,thermal}} + \sum_{k=1}^{N} P_n(k)} \text{SINR}_{\min}(R), k \notin (n, m),$$

when device d_m transmits using MCS R. The determination of possible simultaneous transmissions, apparently, is a combinatorial problem.

The assumptions made are very stringent and lack several aspects of wireless communication. For any $n \neq m$, the necessary power $P_n(m)$ is known at the receiver side only. Thus, solely device d_n does know the power value $P_n(m)$. To identify opportunities for concurrent transmissions, each device $d_k, k \notin (n, m)$ needs to determine its impact on $\text{SINR}(P_n(m))$. If simultaneous transmissions of devices d_k lead to $\text{SINR}(P_n(m)) < \text{SINR}_{\min}(R)$, the devices shall not be allowed to transmit concurrently. However, the PHY mode $R(m)$ used for transmission of device d_m to d_n may not be known at d_k. Hence, it cannot determine the value of $\text{SINR}_{\min}(R(m))$. Furthermore, the latter value may dependent on the receiver of device $d_n : \text{SINR}_{\min}(R(m, n))$. With some transceiver manufacturing tolerances, even similar transceiver hardware may apply different powers and have different sensitivities. Therefore, $\text{SINR}_{\min}(R(m, n))$ cannot easily be determined. A common worst case $\text{SINR}_{\min}(R)$ must be assumed, sufficient for all devices in the mesh network, greatly reducing the chance of spatial frequency channel reuse. In a non-synchronized packet-based wireless mesh network, frames of arbitrary size are transmitted at any time. Due to unpredictable packet lengths and unknown PHY mode $R(m)$, it may be impossible for devices neighboring d_n to identify start time and duration of frame reception. Since any device d_k must avoid interference to any of its direct and indirect neighbors, opportunities for concurrent frame transmission are very unlikely if neighboring devices start receiving at arbitrary time instants for non-predictable time durations. As a wireless mesh network is defined by multiple devices being in a mutual neighborhood, in a fixed interference range of a device d_k multiple other devices may be affected. Since a fixed interference range is an unrealistic simplification, exploiting all possibilities for spatial frequency reuse appears to be very difficult. Section 12.2 presents a mesh MAC protocol superior to any design known.

4.4.1 Hidden Devices – Potential Interferers

Since a device in the mesh network can communicate to only a subset of the total of all devices but may interfere many more others, hidden devices must be taken into account. A given device may be in reception range of a minor fraction of all other devices in the mesh, having only a few direct, but many indirect neighbor devices that may mutually not be aware of each other. Then, only intermediate relay devices can inform the indirect neighborhood about the existence of such hidden devices. As explained earlier, a hidden device transmits close to the receiver device during a frame exchange, but out of detection range of the transmitter. The terms *close* and *far* relate to the wireless signal propagation range. Hence, there is no direct relationship between range and distance, amendment IEEE 802.11e (IEEE, 2005b) of the WLAN standard 802.11 defines:

Hidden station (STA): An STA whose transmissions cannot be detected using Carrier Sense (CS) by a second STA, but whose transmissions interfere with transmissions from the second STA to a third STA.

A frame transmission of a hidden device may interfere with other frame exchanges, see Figure 4.7. Since it cannot sense the transmitter, it cannot detect an ongoing transmission to the receiving device. Since the hidden device is close to the receiving device, it may interfere frame reception. A usual way to inhibit hidden devices is a handshake establishing channel reservation by transmission of short frames between transmitter and receiver, prior to data frame exchange, e.g., RTS/CTS in 802.11. The RTS frame sent by the transmitter contains the transmission duration of the data frames scheduled that will follow. The receive device responds by transmitting Clear-To-Send (CTS) containing the same transmission duration value. Any devices in the neighborhood

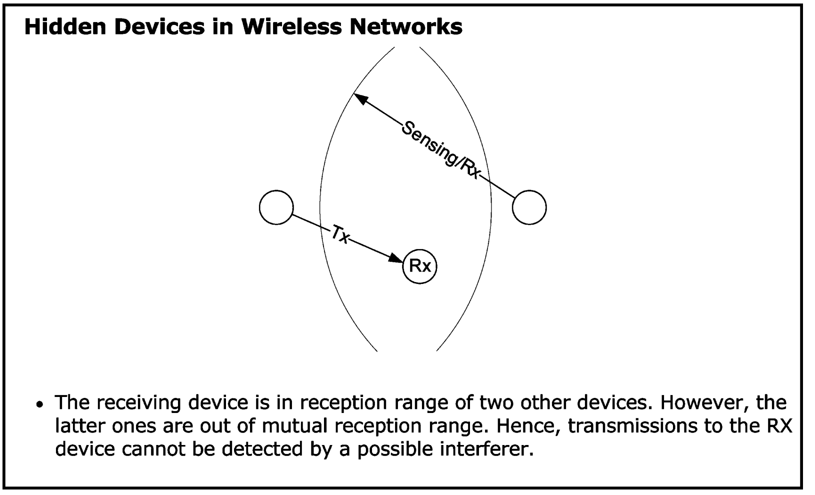


Figure 4.7 In wireless mesh networks, each device has more indirect than direct neighbors. Hidden devices have high potential of interference, therefore.

of both, transmitter and receiver, learn about the upcoming data frame exchange and shall keep silent for the time span communicated, so that any source of possible interference is being silenced. The channel reservation works optimum if the most robust PHY mode is used to transmit RTS/CTS messages.

However, with large differences between interference and reception range, the reservation frames may not silence a sufficiently large area: Collisions by hidden nodes may then still occur. The reservation frames are most likely not received by all devices which can harmfully interfere with the receiver. This fact is specifically serious when using high-speed PHY technology, where reception range is only a minor fraction of interference range, or when taking high mobility devices into account.

Other methods are known to reserve the wireless medium and secure it from interference of hidden devices by using dedicated signaling channels. One is to use a standard IEEE 802.11 transceiver and an additional narrowband transceiver to transmit a busy tone with long reach to avoid any interference from hidden devices. Whenever a device is receiving, it sends a non-modulated busy tone in the narrow band channel, see Figure 4.8. All devices in the mesh network must listen to the narrowband channel prior to any transmission attempt. If a busy tone is sensed, no transmission is allowed. Like the Morse code, where short, long and no power on the wireless channel establish an alphabet, presence of the busy tone indicates a transmission in the neighborhood, while absence indicates a free user channel.

Another method is to pair periodic slots used for data frame transmission with slots on the same channel to transmit a busy tone, as proposed in MDCF, see Section 12.2.1.

4.4.2 Exposed Devices – Unused Capacity

A device is called exposed if according to the protocol applied, the device decides that the channel is not available, so that it refrains from channel access, although its transmission simultaneously to another ongoing transmission would not cause harmful interference. Since exposed devices

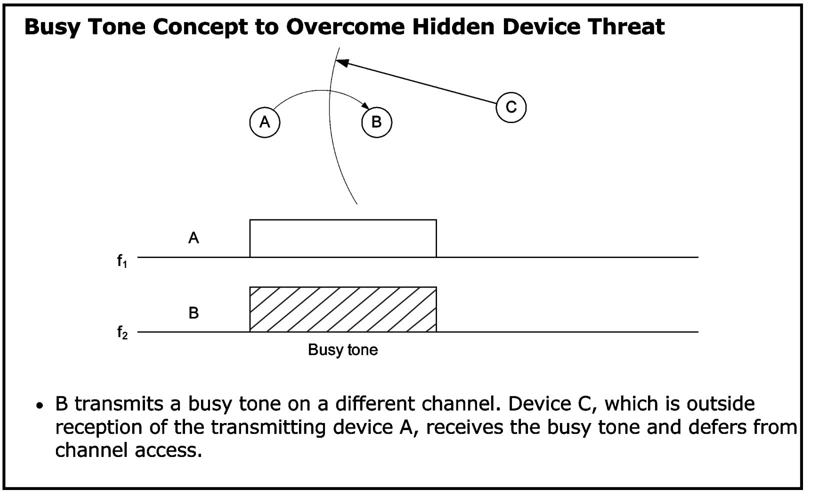


Figure 4.8 With busy tone, a receiving device can signal an ongoing frame reception to its neighborhood.

are not harmful to other devices, most wireless standards do not take them into account, although capacity of the wireless medium is wasted. With increased interest in mesh networks, proper handling of exposed devices appears attractive, since multi-hop communication in mesh networks tends to load the available spectrum much more than single-hop networks.

Protocols that use immediate acknowledgment of each variable length data frame reception like 802.11, cannot control exposed devices properly: Since a transmitter device S1 is required to wait for reception of an acknowledgment to its data frame, no exposed device may transmit while concurrently neighbored to S1.

Obstacles, walls, and buildings provide sufficient shadowing that may allow interference-free simultaneous transmission in the same channel. Detection and identification of opportunities for simultaneous transmission are important for the design of MAC protocols of dense wireless mesh networks.

4.5 Fairness and Congestion Avoidance

In networking, fairness denotes a specific means of resource sharing. If several traffic flows equally share a link, total fairness may be achieved. Different characteristics of fairness may be distinguished. Standard IEEE 802.11 provides fairness based on frames. No matter what size the payload and which PHY mode is used for transmission, all frames have equal chance to get access to the Wireless Medium (WM). With standard IEEE 802.11e, a paradigm shift is introduced. Fairness here is based on the capacity of the WM and fairness among frames of equal priority is based on transmission duration, e.g., for voice and video services, default durations for 802.11e Transmission Opportunities (TXOP) are defined. Hence, devices that use different PHY modes for frame transmission achieve different throughput, although they may have equal share in terms of transmission duration.

When the number of flows is large and the capacity is given, none of the traffic flows may be able to fulfill its QoS requirements. Flow Admission Control (FAC) helps to prevent sharing of link capacity by a too large number of devices by preserving existing traffic flows and denying access of additional flows to a WM. Thus, unfairness is applied to newly arriving flows.

In addition to FAC, traffic flow prioritization may be used to discriminate some traffic flows in favor of others, e.g., high priority traffic may even starve lower priority flows. Then, capacity of the WM is shared among flows of unequal priority. Both FAC and prioritization are needed in wireless mesh networks to establish fairness and support QoS traffic flows.

The capacity of a wireless mesh is given by the relaying capacity of the highest loaded relay node that forms the bottleneck in the network. The network capacity, typically, results from the share of the channel capacity available to it, sometimes its processing speed might limit its performance. In most cases a portal device connecting a mesh network to the Internet forms the bottleneck, but relay nodes in a mesh that must carry many routes may also limit the mesh network capacity. A bottleneck device in a single channel mesh network most of the time is receiving or transmitting frames, leaving the remaining channel capacity to non-neighbor stations for feeding its neighbor nodes with frames.

The Transmission Control Protocol (TCP) greatly impacts the performance of wireless communication systems (Bottigleliengo *et al.*, 2004; Kherani and Shorey, 2004; Leith *et al.*, 2005; Zhenghua *et al.*, 2005). TCP is the de-facto standard for Internet-based applications and hence is omnipresent. Since TCP was designed for wired communication networks, its congestion detection and avoidance algorithm is based on the assumption of frame losses in congested routers having insufficient capacity. The frame error ratio is assumed low and constant as usual with fixed networks. These assumptions do not apply in wireless networks. Fluctuation in channel conditions and device mobility result in a high variance of the frame error ratio owing to frequently changing interference. Link adaptation changes PHY modes and thus the link capacity. Frame error ratios may vary orders of magnitude. TCP draws the wrong conclusions and throttles down the window size making wireless links appear weak in capacity.

Most radio communication systems use ARQ protocols to provide error-free data transmission, see Section 2.5.2. IEEE 802.11 applies an ARQ mechanism, for example, that sends non-acknowledged frames several times again, see Section 5.4. With TCP working on top of IEEE 802.11, two different ARQ protocols mutually interfere. While TCP may retry sending a missing frame, IEEE 802.11 may still increase its contention window to start another back-off period for a collided frame. In multi-hop mesh WLANs such retries may occur several times on each link. TCP tries to avoid end-to-end congestion by reducing the frame generation rate at the source. Independent loops (TCP congestion avoidance and 802.11 back-off algorithm) seek to control frame generation and transmission rate, independently working against each other, resulting in non-predictable performance behavior. With MANET mesh technology working on top of IEEE 802.11 WLAN, the achievable throughput is known to sharply decrease with each additional hop.

Congestion avoidance is important in wireless mesh networks, since frames should be discarded or rejected by the network under local overload as early as possible. Since most wireless mesh networks are packet based, end-to-end connection set-up is not considered. Hence, resources cannot be reserved along an end-to-end path from source to destination. At the border of a wireless mesh, more capacity might be available than in the center, since fewer neighbor devices at the border means more capacity per device. If the load level of the wireless medium in the core mesh cannot be communicated to the border devices, no proactive but only reactive flow admission control is possible (see Figure 4.9). With frames that have already been relayed across several hops, any late discard operation close to the final destination is a severe waste of network capacity. Hence, an early congestion detection and avoidance is necessary in wireless mesh networks.

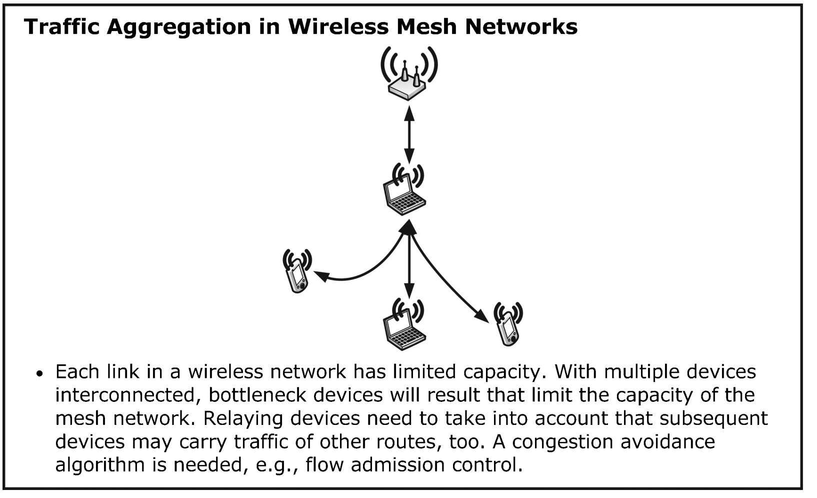


Figure 4.9 A relaying device carries the traffic aggregated from three other devices. Prioritization of the forwarding device is necessary to ensure sufficient performance.

4.6 Routing

This section describes routing algorithms applicable for the next-generation WLAN technologies. In general, networks operating at high carrier frequencies suffer from small coverage areas. Devices in close vicinity will be able to communicate using throughput optimized modulation and coding schemes. However, when increasing the distance of the devices the throughput and service quality will decrease. In order to support mobile devices with continuous service a wireless network must apply fast and effective routing algorithms. Wireless routing algorithms will consider the overall network capacity and topology rather than only the number of hops, like they do today. This section presents routing algorithms and strategies well suited to improve today's routing in order to meet the requirements for future wireless technologies.

4.6.1 Routing Algorithms

MANET ad hoc routing protocols can be categorized into proactive or reactive, and hybrid variants combining aspects of both, see Figure 4.10. Proactive routing protocols use periodic flooding to broadcast information to devices about routes, known neighbors and others. This enables short path set-up times and ensures that the latest parameter values of the routing metrics are always present in the transmit range. Increased overhead is a drawback of proactive routing protocols. OLSR (IETF, 2006a) and DSDV (Perkins and Bhagwat, 1994) are two examples of proactive routing protocols.

Reactive routing protocols establish a route on request only, reducing path selection overhead but introducing high delay for the first frame to be transmitted as known for AODV (Perkins *et al*., 2003) and (IETF, 2004). The reason is that the path selection procedure must be executed before data frames can be exchanged. As several paths may lead from the source to the same destination, the path selection frame that arrives first at the final destination usually determines the preferred path. Reactive protocols avoid maintaining unused routes, but pay for this by a higher

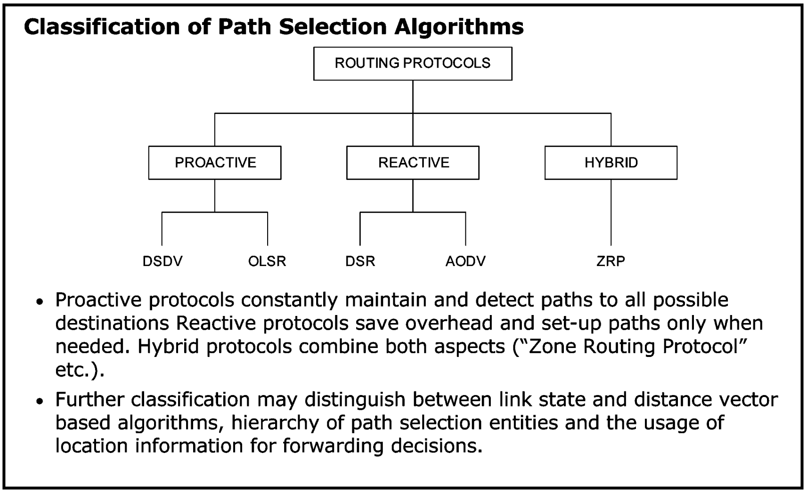


Figure 4.10 Ad hoc routing protocols can be classified by adaptation strategy.

route discovery and packet transmission delay. Proactive protocols provide a route when requested and therefore minimize the packet delay. Furthermore, hybrid approaches have been developed.

When a route in use breaks, both proactive and reactive protocols need to recover by establishing a new route. Most approaches inform the source device and it then starts a completely new route discovery process, thereby a large number of signal messages are exchanged, and the network is flooded. Some routing protocols try to keep the route discovery process local around the devices involved in the breakage; hence the load to the network from flooding is limited. Nevertheless all approaches only react in a proper manner, when the route is already broken. This typically leads to a high number of lost packets as well as increased route rediscovery and packet transmission delay. In the following we present the most discussed routing protocols, namely the Ad-Hoc On-Demand Distance Vector Routing (AODV). Classification of ad-hoc routing protocols may also differentiate between source and non-source, hierarchical and flat routing protocols, etc., see (SECAN-LAB, 2006).

4.6.1.1 Ad-hoc On-demand Distance Vector Routing (AODV)

AODV is a severe candidate for wireless mesh networks. Owing to its reactive nature the protocol avoids maintaining unused routes introducing a higher delay than proactive protocols when establishing a route (Perkins *et al.*, 2003; Perkins, 2000; Weiss *et al.*, 2004).

4.6.1.2 Route Discovery

A device intending to send a packet to some destination device checks its routing table to determine whether it has a valid route available. If a route is known the packet is forwarded to the next hop towards the destination, otherwise, device D1 initiates a route discovery process. It then broadcasts a Route Request (RREQ) message and floods the network (cf. Figure 4.11(a)). The RREQ contains the source IP address, the destination IP address, a sequence number from the source device, the last known sequence number from the destination device and a broadcast

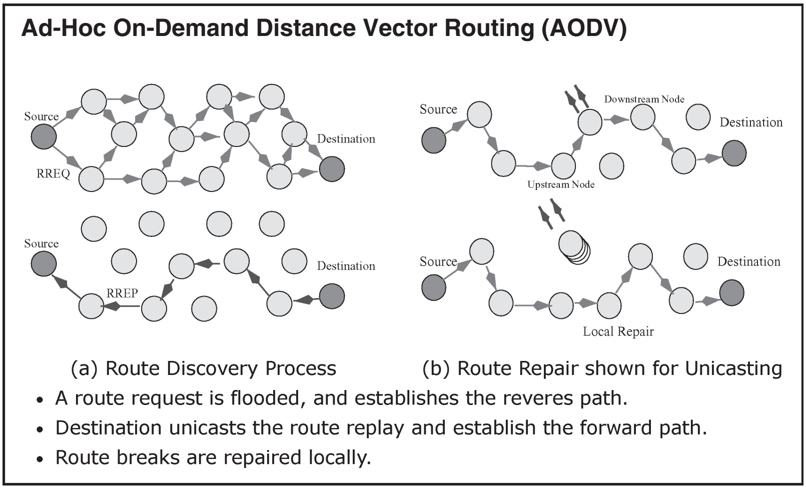


Figure 4.11 AODV routing.

ID. The broadcast ID is incremented with each broadcast sent by the source device D1. The source IP address and the broadcast ID form a unique identifier for the RREQ. After receiving the RREQ a device D2 checks whether it is the destination or if its routing table contains a valid route to the destination device. If the destination is unknown device D2 builds a new reverse route entry for the source device D1. This reverse route entry contains the source IP address of the RREQ, the according sequence number, the hop count towards the source D1 together with the IP address of the neighbor device where the RREQ was received from (i.e. the next hop towards the source device). Afterwards, device D2 increments the RREQ hop count and re-broadcasts the RREQ.

Hence, device D2 has a route to forward a Route Reply (RREP) to source device D1 if one arrives from other devices. If device D2 itself is the destination or if it knows a valid route entry for that destination, it compares the received sequence number from the RREQ with the last received sequence number stored in its routing table. If the stored sequence number is at least as great as indicated in the RREQ, device D2 responds to the request by transmitting a RREP packet to the device D1. The RREP contains source and destination IP addresses. If the responding device D2 is the destination it inserts the current sequence number and sets the hop count to zero in the RREP.

If an intermediate device D2 replies, it inserts the sequence number and hop count saved in its routing table for the respective destination into the RREP. Each device receiving the RREP builds a forward path entry containing the destination IP address, the neighbor IP address of the last RREP sender and the hop count towards the destination. Routing table entries are associated with a certain lifetime. Each time the entry is used the lifetime is updated. If the route is not used within the specified lifetime the respective routing entry is deleted.

To limit the overhead an extending-ring search mechanism is proposed in Perkins *et al.* (2003). The extending-ring search algorithm sends repeated RREQ messages with an increased Time-To-Live (TTL). Thus flooding of the whole network is avoided if the destination device is in close vicinity of the source device, although extending-ring search delays route discovery if the destination is far away.

4.6.1.3 Route Maintenance

Once a route has been discovered, it is maintained as long as needed. If the source device moves during an active session it reinitiates the route discovery process. When either the destination or an intermediate device moves and the route breaks a Route Error (RERR) packet is sent to the source device. This RERR is sent by the device on the source side of the break (upstream).

When the neighbors receive the RERR, they mark the affected route entries as invalid (hop count is set to infinity) and send the RERR to all neighbors that are affected by the broken link. The source device recovers the route when it receives the RERR. In addition, AODV also has a method to repair broken routes locally.

4.6.1.4 Local Repair

AODV by route maintenance is able to repair broken routes. When a route breaks the upstream device of the broken link decides either to repair the route or to send an RERR message (cf. Figure 4.11(b)). A device using local repair sends an RREQ searching for a new route to the destination device. It is worth noting that the repair RREQ will not reach the source device thus preventing creation of loops (Perkins *et al.*, 2003).

If the repair attempt fails, an RERR message is sent back to the source device. Otherwise the initiator of the local repair updates its routing entry and compares the stored hop count with the recently received one. If the new hop count to the destination is larger than the former hop count, the device creates an RERR message for the source device with the N flag set to distinguish it from a standard RERR. A device receiving the repair RERR updates the respective route entry instead of setting the entry invalid. However, when the source device receives a repair RERR, it may decide to initiate a completely new route discovery process in order to avoid inefficient routes.

Based on link adaptation related information available from the PHY, a device may be able to predict from the history of MCSs used the current link state so that a device is able to rearrange a route before it breaks (Weiss *et al.*, 2004; Weiss *et al.*, 2005; Weiss *et al.*, 2006). Two link state prediction based route re-arrangement algorithms are proposed: Early Route Rearrangement (ERRA) and Early Route Update (ERU), see Sections 4.6.5 and 4.6.6.

4.6.2 Common Link Layer Behavior (Link Adaptation)

Most wireless communication standards do not specify rules for Link Adaptation (LA) but leave this open to vendor-specific implementation. In general, most LA algorithms react to degrading channel conditions by switching to a lower data rate (more robust) PHY mode. If the SINR of the wireless medium improves again, a higher data rate (less robust) mode is chosen. To investigate the potential of use of LA related information on the routing performance the IEEE 802.11a PHY and MAC have been taken as an example (Weiss *et al.*, 2004; Weiss *et al.*, 2005). The LA algorithm developed is an enhanced version of the Auto Rate Fallback (ARF) (Kamerman and Moneban, 1997) algorithm addressing the following design principles:

Quick reaction to fast channel condition changes is a major requirement for an LA algorithm, while slow channel condition change must be considered on a long-term basis. To fulfill these requirements, LA uses three buffers (called packet window, PW_s) labeled short PW_s, medium PW_m and long PW_l , of size 5, 10 and 25 to store the PHY mode ID used in the past, along with information about whether packet transmissions were successful (acknowledged) or not (Figure 4.12). For each PW a packet error ratio (ER_s, ER_m, ER_l) is calculated separately.

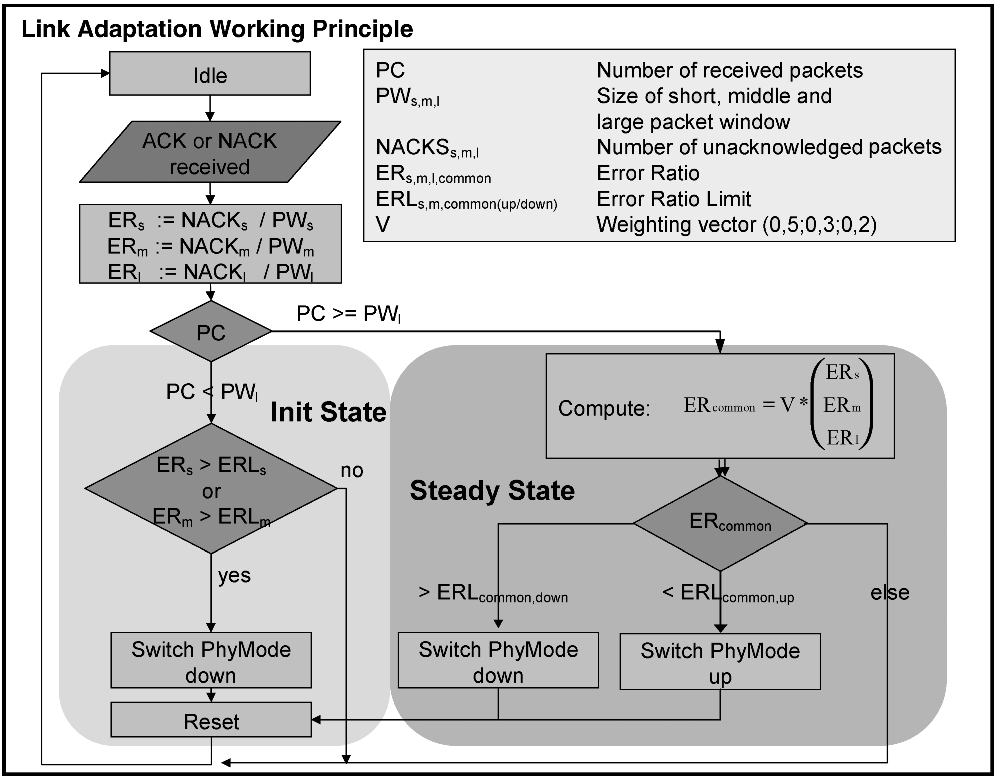


Figure 4.12 Link adaptation working principle.

The LA algorithm operates from an initial state *Init State* until all packet windows are filled. Then, LA switches from the PHY mode used initially to the current link condition. If the packet error ratio in buffers PW_s or PW_m falls below an Error Ratio Limit (ERL) the PHY mode is decreased, otherwise it is increased. It is important to switch to a suitable PHY mode before the IEEE 802.11 MAC starts discarding erroneous packets.

The Modulation and Coding Scheme (MSC) QPSK 3/4 is chosen as an initial PHY mode for a forward link. To further enhance the LA algorithm performance, after entering *Init State* the LA checks availability of the backward link. If it is available but its PHY mode is lower than the initial default PHY mode chosen for the forward link, the PHY mode of the backward link is also applied as the initial forward link PHY mode, see upper part of Figure 4.13. Device A transmits using BPSK 1/2 to device B having already adapted its PHY mode to the link conditions. Hence, device B uses the same MSC for its initial PHY mode to transmit to device A.

Once all buffers are filled, the LA switches to the steady state, where all buffers are considered. The LA algorithm applies a weighting vector $\vec{V}^T = (\nu_s, \nu_m, \nu_l)$ to combine the different error ratios. The resulting error ratio is used to decide whether the PHY mode must be increased, decreased or left unchanged. To be able to react to fast changing channel conditions, PW_s should get a high weight. The simulation results presented used as weights ($\nu_s = 0.5$, $\nu_m = 0.3$, $\nu_l = 0.2$) weighting PW_s by 50 %, PW_m by 30 % and PW_l by 20 %. If the link to a certain neighbor device

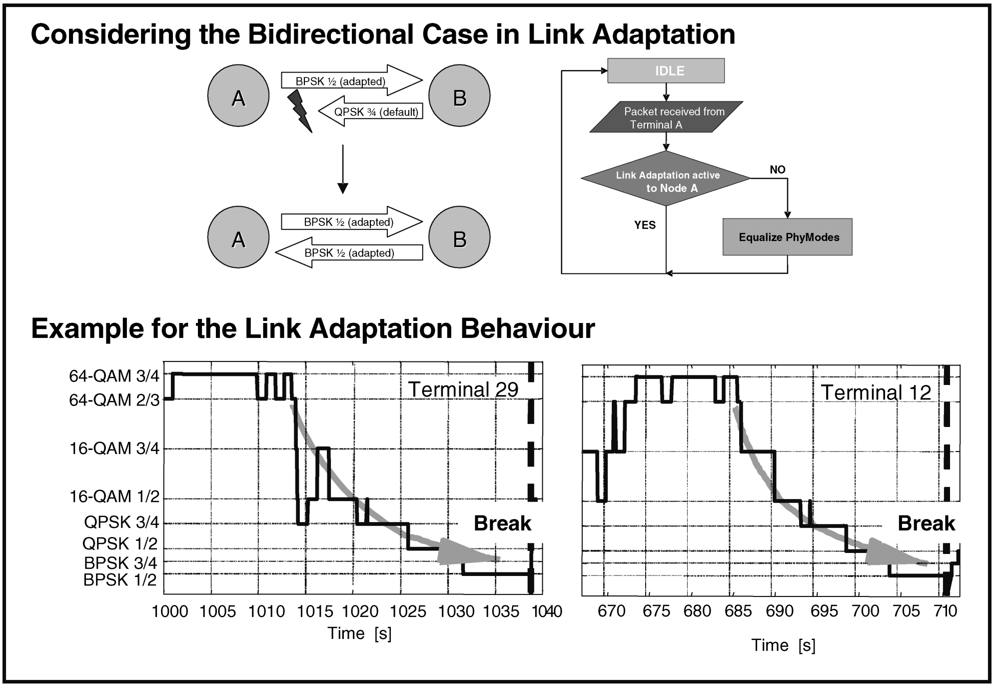


Figure 4.13 Consideration of a bidirectional case and example behavior of link adaptation.

has been idle for a given time duration, the related buffers are emptied. Consequently, a device must use the initial PHY mode if no information on the link state is available.

This LA algorithm is built for operation in an ad hoc network where an LA supervises each link separately. A pure ad hoc scenario has been investigated by simulation containing 40 devices all moving according to the Random Waypoint Model (RWP) (Bettstetter *et al.*, 2003) except the source and destination devices that remain fixed, each source offering 100 kb/s constant bit-rate traffic with 512 bytes per packet. The lower part in Figure 4.13 presents histories of LA decisions for two terminals 29 and 12 that are at some time the next hop to the source node. The typical LA behavior ("step down") prior to a link break is observable. LA usually adapts the PHY mode to avoid a link break and consequently, network connectivity information for proactive routing update can derived from this.

4.6.3 Link Breakage Prediction

Link quality degrades when devices depart from each other. Sequential step-down of LA may serve to hint to the network layer that a link will break soon, triggering route adaptation to prepare for an alternate route. Step-up of LA is of no concern here since the respective link becomes more stable.

A transformation function $RF(t)$ to formally relate MSC step-down and expected route interruption for automatic detection needs to be established as shown in Figure 4.14. The rating function ranks the importance of a PHY-mode change to avoid link breakage, e.g., a change from BPSK 3/4 to BPSK 1/2 (value 7) with high probability hints to a potential future link breakage

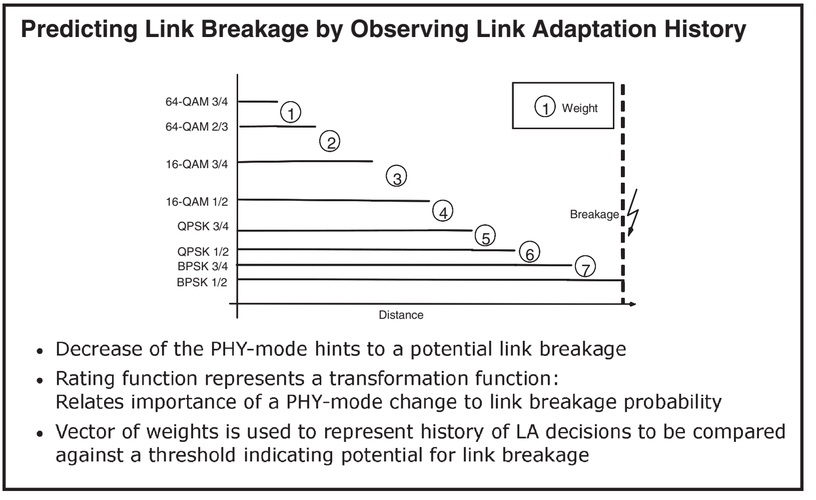


Figure 4.14 Prediction of link breakage by observation of link adaptation algorithm history. A weighting vector weights the importance of each PHY-mode change and relates it to potential link breakage.

than just switching from MSC 16 QAM 3/4 to 16 QAM 1/2 (value 2). Therefore a vector of weights may relate the importance of each PHY-mode step-down to the probability of breakage (cf. Figure 4.14).

During link operation, step-up and step-down might happen multiple times before a link will break. By summing all weights related to LA steps within a certain time period, where step-down is counted as negative, step-up as positive, an indicator for link reliability is gained. Whenever the LA algorithm advises the MAC to use the most robust PHY mode (BPSK 1/2), the slope of $RF(t)$ is checked against a certain threshold. When exceeding the threshold, the link is considered suspicious and might break, and the routing layer is triggered to start routing update (Weiss *et al.*, 2006).

4.6.4 Actions for Expected Link Break

Assuming LA provides useful information about the link state, it may trigger actions such as trying to rescue a link, thereby avoiding a route rediscovery process, or guaranteeing a certain required link quality by establishing a new route. Devices from monitoring incoming and outgoing links may know whether none, one or both directions of a link pair are being up- or downgraded by LA. A device may distinguish three cases:

- The outgoing link on a route is switched to lower PHY mode but the incoming links remain unchanged. This observation is typical for the link to a next device fades away since it moves away.
- Both the incoming as well as the outgoing links are switched to lower PHY mode, indicating movement of the observing device itself. Thus, the device should regard itself as an instable intermediate device in a route.
- The incoming link is switched to a lower PHY mode, but the outgoing links remain unchanged. This indicates that the device at the other end is moving away.

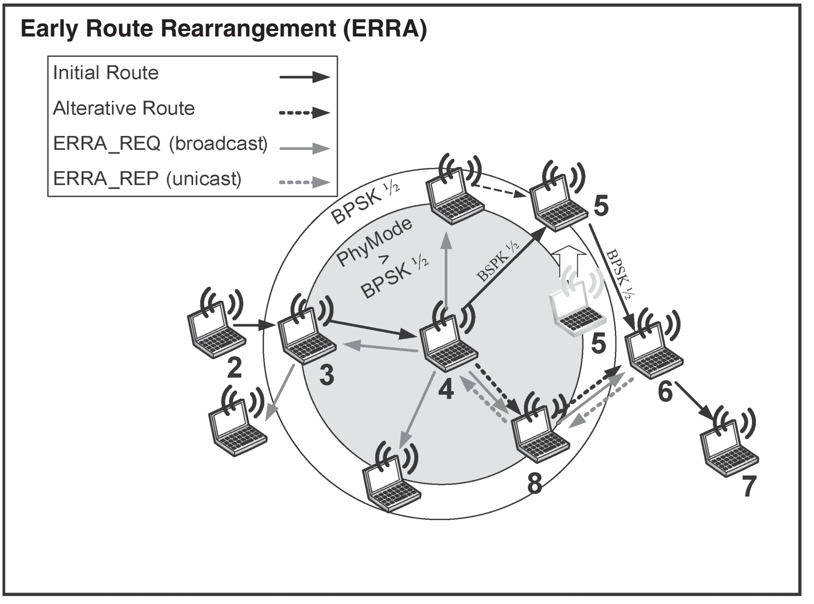


Figure 4.15 Signaling for early route rearrangement (ERRA).

The three cases are considered in the scenario depicted in Figure 4.15. Device 4 experiences case 1, device 5 experiences case 2 and device 6 experiences case 3 as described in the next section. The route rearrangement ERRA and route update ERU protocols described in the following are able to cooperate with the AODV routing protocol. AODV uses sequence numbers to distinguish new messages from older ones. Every time a device emits a route request for a certain destination, the device increments the route request sequence number for that destination. ERRA and ERU maintain these sequence numbers, too.

4.6.5 Early Route Rearrangement (ERRA)

The Early Route Rearrangement (ERRA) protocol is derived from the local repair idea that is part of the AODV routing protocol (Perkins *et al.*, 2003).

Unlike local repair in AODV, the ERRA protocol does not wait until the link is broken but prior to breakage rearranges the route to avoid disruption, see Figure 4.15 for an example. The initial route starts from source device 2 to destination device 7. Intermediate device (5) is moving away. Device 4 first detects the movement, since it steps down the PHY mode of its link outgoing to device 5. When device 4 switches to PHY-mode BPSK 1/2 it checks for a step-down pattern and finds an indication for device 5 is moving away from itself, so that it triggers the ERRA procedure to rearrange the route.

Device 4 locally broadcasts a route to device 7 rearrangement request (ERR_REQ) with a pre-determined PHY mode that is higher than the last one used for packet transmission. In Figure 4.15 the request is sent with QPSK 1/2. This ensures avoiding selection of the former link between device 4 and 5. Device 8 forwards the request to device 6, which is aware of a route to destination device

7. Accordingly, device 6 responds (ERRA_REP) and provides via device 8 an alternative route for device 4 to 7. Device 4 compares the hop count of the old and new route. If the new hop count is less or equal, the route is established immediately. Otherwise device 4 may decide to store the alternate route in a temporary routing table and use the old route unless it breaks and then switches to the route contained in the temporary routing table (Weiss *et al.*, 2006).

ERRA proactively, by rearrangement, prepares for an alternate route to avoid interruption of service. ERRA misses an opportunity as visible from the example presented: When device 4 discovers degradation of a link (to device 5), the route still exists and may be used during routing update for providing alternate routes. Therefore, ERRA is extended resulting in the early route update protocol described in the following.

4.6.6 Early Route Update (ERU)

Like ERRA, the ERU protocol proactively updates the routing table, and takes MSC step-down to the lowest PHY mode as a trigger. The current stepped-down link is used to establish alternate routes. Figure 4.16 introduces ERU signaling sequences: Device 5 is assumed to leave a route. Device 4 is trigged by its LA procedure when switching to BPSK 1/2 and transmits its neighborhood table (ERU_PATCH_INFO) piggy-backed to some data packet sent to device 5 that is also aware that the current link tends to break. Device 5 forwards the information to

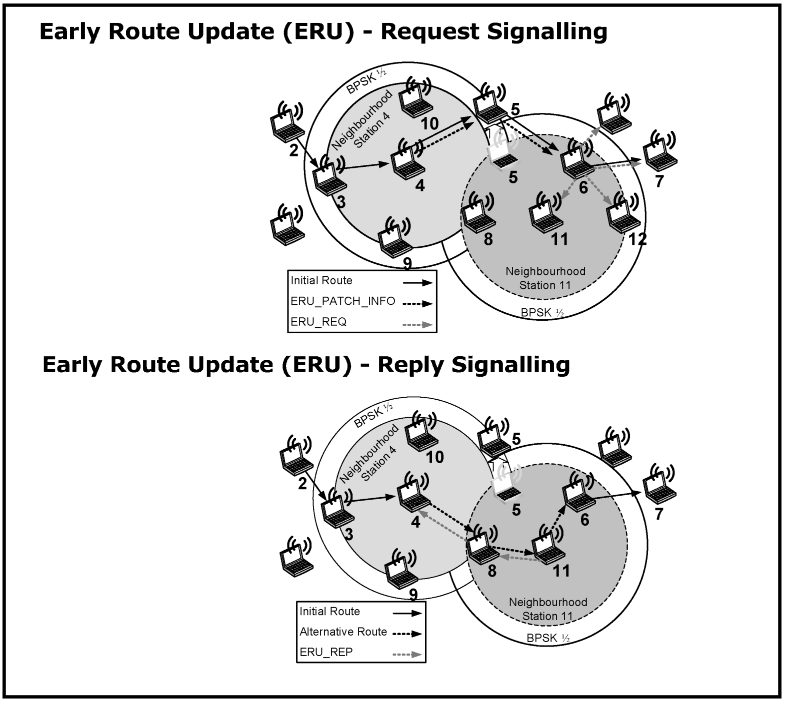


Figure 4.16 Request and reply signaling in early route update (ERU).

device 6 that has a steady outgoing link to device 7. The ERU_PATCH_INFO carries a counter (Breakage Hop Counter, BHC) to count the number of hops transmitted on a route, representing the size of the unstable part of the route. First, device 6 searches but does not find an intersection between the neighborhood received from device 5 and its own, Figure 4.16. Hence, device 6 and its neighbor devices further re-broadcast the list.

The number of broadcasts correspond to the BHC (in the example BHC = 2), greatly reducing traffic compared to flooding the vicinity device 4. According to Figure 4.16, device 11 has device 8 in its neighbor list and device 8 is included in the neighbor list of device 4. Therefore, a route 6–11–8–4 bypassing the weak link 4–5 is found. Device 11 responds in its reply message (ERU_REP) to device 4 via device 8.

4.6.7 Simulation Results

The LA algorithm and its cross-layer interaction to IP routing has been studied by simulations where 40 802.11 devices, except the source and destination devices, move according to RWP (Bettstetter *et al.*, 2003) in a (100 m × 100 m) area. Three routes are loaded with constant bit-rate traffic with a packet size of 512 bytes, and RTS/CTS handshake enabled. Figure 4.17 (left) shows the relative throughput averaged over the three routes for ERRA, ERU and AODV with local repair enabled. No throughput degradation occurs when using ERRA or ERU and these algorithms reach higher throughput than AODV at high offered load. Figure 4.17 (right) shows the mean hop count of routes versus offered traffic and reveals that for light traffic load ERRA and ERU are able to shorten the route compared to AODV.

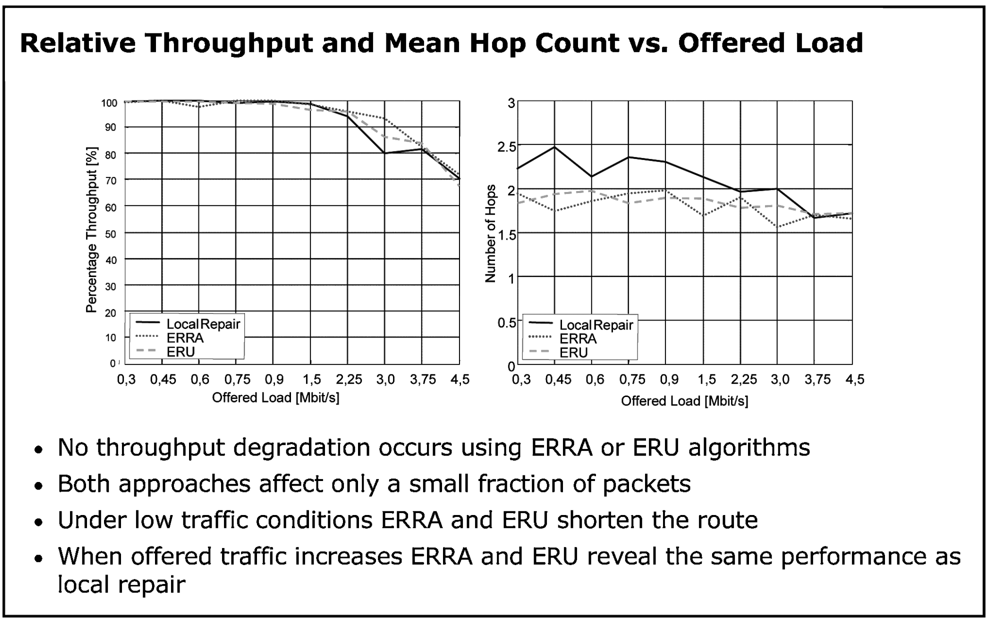


Figure 4.17 Comparison of the ERRA, ERU and AODV local repair approaches. Throughput percentage and average hop count are evaluated.

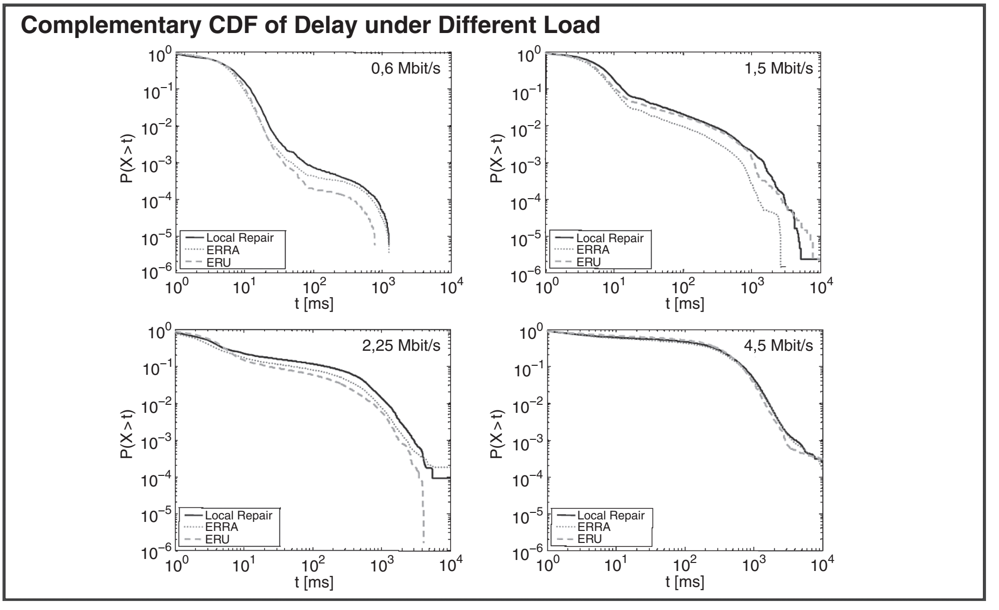


Figure 4.18 Complementary CDF for delay under different offered loads for ERRA, ERU and AODV local repair algorithms.

Most interesting is the packet delay gained by ERRA and ERU that avoid route breaks. Figure 4.18 shows the complementary Cumulative Distribution Function (CDF) of delay for different loads. It can be seen AODV with local repair is outperformed throughout by ERRA and ERU with ERU resulting in lower delay under low load.

4.6.8 Conclusions

Wireless mesh networks need high efficient routing protocols to cope with abruptly appearing obstacles, short transmission ranges and link breakage owing to mobile devices. Fast rerouting and reestablishment of routes is a main requirement. State-of-the-art solutions apply routing protocols providing multiple routes between each source–destination device pair to reduce impact of link breakage through fast and seamless switching to an alternate route. Cross-layer information exchange between MAC and IP layer may exploit current link state information in terms of MCS and the history of the link adaptation algorithm to proactively react on link degradation.

4.7 Summary

The fundamentals of meshing in the context of wireless communication have been discussed in this chapter. Arising problems and advantages of wireless mesh networks have been introduced in general. Additionally, different approaches to routing have been outlined and evaluated.

After introducing the basics of 802 WLANs, WPANs and WMANs in the subsequent chapters, the concrete approaches to meshing as part of these standards are highlighted in Chapter 8 where the realization of meshing in 802.11, 802.15 and 802.16 is described.

Note

1. Reproduced by permission of © 2006 IEEE. Source: G. R. Hiertz, S. Max, E. Weiß, L. Berlemann, D. Denteneer, S. Mangold, “Mesh Technology enabling Ubiquitous Wireless Networks – Invited paper,” in Proc. of the 2nd Annual International Internet Conference, Boston, USA, 2006. (Section 4.4 (in parts)).

5

IEEE 802.11 Wireless Local Area Networks

Stefan Mangold, Lars Berlemann, Matthias Siebert and Bernhard H. Walke

Undeniably, the IEEE 802.11 protocols and transmission schemes are truly one of the most remarkable standardization achievements. An uncountable number of devices is today based on this standard. It started with a wireless extension for local area networks in 1997, and since then has been gradually improved and extended towards a very flexible, well-understood technology. Because 802.11 was built for radio systems in unlicensed spectrum, there is virtually no limitation to the use of 802.11: unlicensed spectrum is often harmonized throughout the world, which means that such radio systems can be used at any location and time. Because of its inherent simplicity, 802.11 is the dominant standard for commercial wireless communication systems, and the research community often refers to this standard during experimentation, and when developing future wireless systems.

The IEEE published the original IEEE 802.11 standard in 1997 as a specification for the transmission scheme and medium access control protocol for Wireless Local Area Networks (WLANs). A revised version of improved accuracy followed in 1999. At the same time, 802.11a and 802.11b, which were the first substandards to extend 802.11, were published in parallel in 1999. The *IEEE Wireless LAN Edition* is a compilation of 802.11, 802.11a and 802.11b (IEEE, 2003a). Today's 802.11 is divided into many more substandards each addressing particular extensions. With this diverse set of incremental improvements, 802.11 continues to evolve into different directions that are demanded by commercial, scientific, medical, public safety, and military needs. As a result, because of this flexibility, 802.11 emerges towards an ever-present technology.

IEEE 802.11 is described and analyzed in detail in this chapter, where focus is given to the layer 2 protocols for spectrum management and Quality of Service (QoS).

5.1 Scope of 802.11

Like IEEE 802.3 (Ethernet) and IEEE 802.5 (Token Ring), the 802.11 standard focuses on the two lower layers (1 and 2) of the Open System Interconnection (OSI) reference model. This is

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

indicated in the reference model of 802.11 illustrated in Figure 5.1. This reference model divides the Data Link Control (DLC) layer (i.e., OSI layer 2) into Logical Link Control (LLC) and Medium Access Control (MAC) sublayers. 802.11 defines Physical layer (PHY) transmission schemes (OSI layer 1), and the MAC protocol, but no LLC functionality.

For LLC, the 802.11 system may rely on general protocols that are usable with all 802 standards. This LLC layer is independently specified for all 802 LANs, wireless or wired.

Whereas this LLC protocol can be applied for wired and wireless systems, the management and control functions to address the characteristic implications of wireless communication systems need to be specified by 802.11 in particular. To consider for example radio range aspects, 802.11 includes functions for the management and maintenance of the radio network, which exceed the usual MAC objectives.

5.2 Reference Model, Architecture, Services, Frame Formats

In the following, we present the 802.11 reference model, together with the architecture, the services that 802.11 provides, and the frame formats of what is exchanged between 802.11 radio stations.

5.2.1 Reference Model

A reference model is used as orientation. It has no meaning other than helping developers to discuss and understand the technology. During the standardization process, such a reference model is widely used to assist the definition of the standard's objectives, and of the modules that are needed to realize the radio system. The reference model is an abstraction of the real-life hardware and software modules of a radio standard. It often describes the part that requires standardization together with the proprietary part that does not require a standardized behavior.

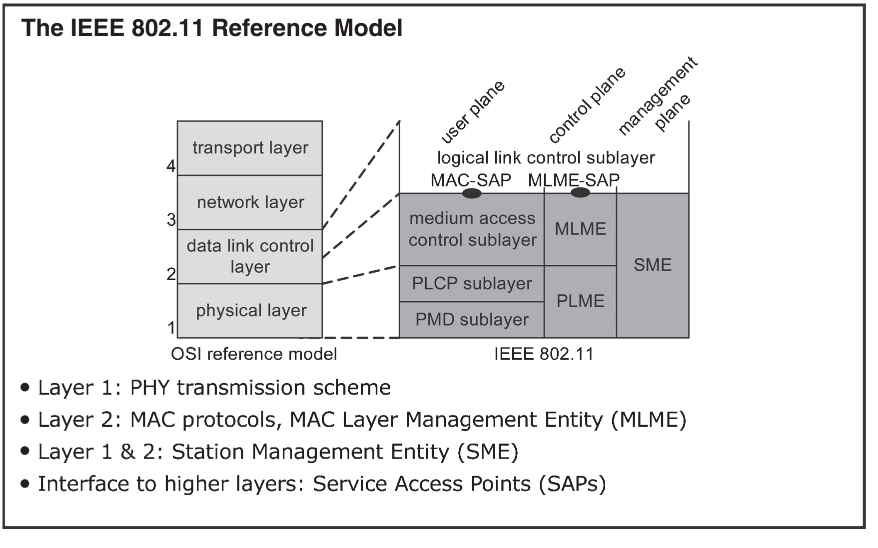


Figure 5.1 The IEEE 802.11 reference model (right) and the OSI reference model (left).

Figure 5.1 illustrates the 802.11 reference model. The Physical Medium Dependent (PMD) sublayer is responsible for sending and receiving data via the wireless channel and defines the transmission scheme, which is different for the different PHYs, whereas the Physical Layer Convergence Protocol (PLCP) sublayer adapts the requests of the common MAC to the different PHYs into a format specific to the applied PMD. The MAC user plane is fed with data frames via the MAC Service Access Point (MAC-SAP) at the MAC/LLC boundary.

The control plane incorporates the MAC Layer Management Entity (MLME) and PHY Layer Management Entity (PLME). The management plane is represented by the Station Management Entity (SME). The definition of these entities is very vague in the standard. The reason is that they are implementation dependent and do not need to be standardized to achieve interoperability of different implementations.

All these parts of the reference model will assist us when we discuss the standard or describe its various functions. However, 802.11 is a network of stations, and 802.11 services are not necessarily provided by one single station: groups of stations are often responsible for a task. These responsibilities are defined in the 802.11 network architecture.

5.2.2 Architecture

The 802.11 network architecture, illustrated in Figure 5.2, is hierarchical. Its basic element is the Basic Service Set (BSS), which is a group of stations controlled by the so-called Coordination Function (CF). The CF manages the access to the wireless medium. The Distributed Coordination Function (DCF) (described in detail in Section 5.4.1) is used by all stations in the BSS, whereas the Point Coordination Function (PCF) (described in Section 5.5.1) is an optional extension for the support of QoS, described in older versions of the standard.

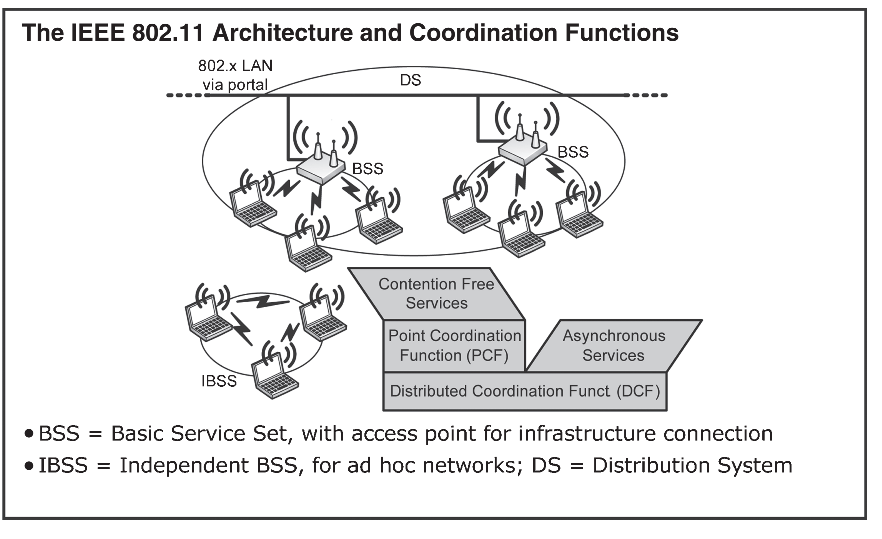


Figure 5.2 The IEEE 802.11 architecture with typical scenarios of the different service sets (BSS, IBSS, DS). The DCF/PCF coordination functions are concepts for spectrum management and medium access. The PCF uses the DCF coordination function to support QoS.

An Independent Basic Service Set (IBSS) is the simplest 802.11 network type. It is a network consisting of a minimum of two stations, where each station operates with exactly the same protocol. No station has priority over another, the responsibility of coordinating the medium access is distributed among all stations.

An infrastructure-based BSS includes one station that has access to the wired network and is therefore referred to as an Access Point (AP). The abbreviation “BSS” in the following is used to refer to both types of service sets, if not stated otherwise. Stations, including the AP, are simply referred to as stations, which in the original standardization documents is abbreviated as STA.

A BSS may also be part of a larger network, the so-called Extended Service Set (ESS). This ESS consists of on or more BSSs connected over the Distribution System (DS). Originally, without the mesh network extension, BSS and DS operate independently on different media. The BSS operates on wireless channels whereas the DS typically uses the Distribution System Medium (DSM). As the 802.11 architecture is specified independently of any specific media, the DSM may use different variants of IEEE 802 networks for its service, for example Ethernet.

5.2.3 Services

The DS provides the service to transport MAC Service Data Units (MSDUs) between stations that are not in direct communication.

An AP provides the Distribution System Services (DSS). The DSSs enable the MAC to transport MSDUs between stations that cannot communicate over a single instance of radio channel.

There are two categories of services in 802.11, the Station Services (SS) and the Distribution System Service (DSS). DSSs are not available in an IBSS. The main SS of a BSS is the MAC Service Data Unit (MSDU) delivery. Other SSs include (de-)authentication and privacy.

DSSs include (re-)association, (dis-)association, and integration. The integration service enables the delivery of MSDUs between non-802.11 LANs and the DS via the so-called *portal*. A portal is the logical point where a non-802.11 LAN is connected to the DS, for communication across the different types of LANs.

5.2.4 802.11 Frame Formats

The OFDM-based 5 GHz physical layer 802.11a (IEEE, 1999a) is used here to describe the 802.11 frames. The alternative PHYs use similar frame formats with some minor modifications and are not discussed in detail.

The 802.11a OFDM PHY consists of two sublayers, the PLCP sublayer and the PMD sublayer, as it is common for all PHYs. The PLCP sublayer maps the 802.11 MAC Protocol Data Units (MPDUs) into a frame suitable for transmitting and receiving user data, control and management information between the associated PMD entities. The PMD determines the actual transmission scheme, i.e., the way of transmitting and receiving data through the wireless medium. The PLCP maps MPDUs into PHY Service Data Units (PSDUs) within the PHY PLCP layer. The resulting frame that is finally transmitted over the channel is called the OFDM PLCP frame and starts with a PLCP preamble followed by a SIGNAL field, and ends with the DATA field.

The PLCP header consists of the following elements, see Figure 5.3: (i) a RATE field; (ii) a LENGTH field; (iii) one reserved and parity bit; (iv) six tail bits; and (v) a SERVICE field. All of these except for the SERVICE field constitute a separate single OFDM symbol, denoted as SIGNAL, which is transmitted with the most robust combination of modulation and coding rate BPSK 1/2. The *12 bit* within the LENGTH field of the PLCP Header indicate the length of the

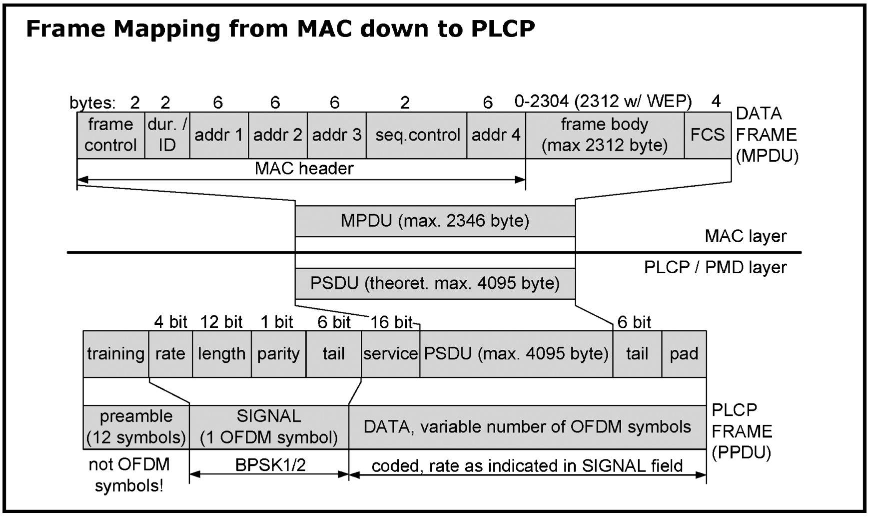


Figure 5.3 MAC DATA frame to PHY PLCP frame mapping for 802.11a.

PSDU, i.e. the number of bytes inside the PSDU. These 2 bits can theoretically represent up to *4095 byte* for the PSDU. Since *11 bit* cannot represent up to *2346 byte*, which is the maximum length of an MPDU, *12 bit* are required for this field.

Thus, the current 802.11a PHY accepts MPDUs with a frame length of up to *4095 byte*. However, the 802.11 MAC will not feed such a long MPDU to the PHY, MPDUs are limited to *2346 byte* in length. The tail bits in the SIGNAL symbol allow for decoding the RATE and the LENGTH fields immediately after their reception. The knowledge of RATE and LENGTH is required for decoding the DATA part of the frame.

The resulting OFDM PLCP frame consists of three parts, the PLCP preamble followed by the SIGNAL field and the DATA field. Figure 5.3 shows the structure of the frames for all the sublayers discussed.

The frame bodies of the previously described frames are filled with the PHY Service Data Units (PSDU) that correspond to MAC Protocol Data Units (MPDU), which have to be delivered across the BSS, or DSS. Figure 5.4 illustrates the formats of MPDUs that carry in their frame body MSDUs, if applicable.

There are four addresses in the MAC header of an MPDU. These addresses are Source Address (SA), Destination Address (DA), Transmitting station Address (TA), and Receiving station Address (RA), respectively denoted as *addr 1 . . . addr 4*. The DA (addr 1) identifies the MAC entity or entities (hence, the station/stations) intended as the final recipient(s) of the MSDU contained in the frame body. The SA (addr 2) identifies the MAC entity (the station) from which the transfer of the MSDU contained in the frame body was initiated. The TA (addr 3) identifies the MAC entity (the station) that is transmitting the MSDU contained in the frame body. Finally, the RA (addr 4) identifies the intended immediate recipient MAC entity (station) on the wireless medium. This address 4 is not necessary if MSDUs are delivered within a BSS.

There are three types of MPDUs: management, data, and control frames, as illustrated in Figure 5.4. When these frames are used and for what purpose will be explained in Section 5.4 where the 802.11 MAC protocol is discussed in more detail.

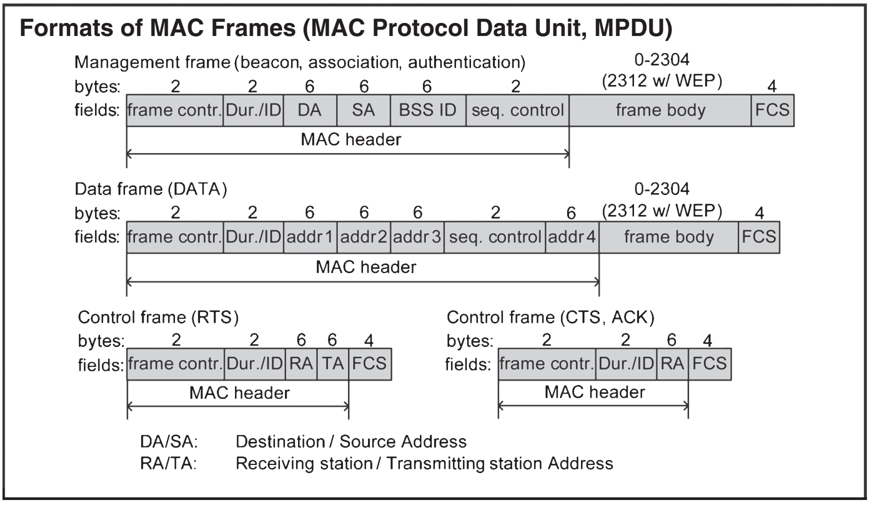


Figure 5.4 802.11 MAC frame format. There are management, data, and control frames. The frame body size of data and management frames are variable.

5.3 Physical Layer

802.11 in its original form defines three different types of PHYs, namely 2.4 GHz Frequency Hopping Spread Spectrum (FHSS), Direct Sequence Spread Spectrum (DSSS), and Infrared (IR). There are four more PHYs defined in the supplement standards 802.11a, 802.11b, 802.11g, and 802.11n, which are described later in this chapter. Figure 5.5 shows the frequency channel allocation in the 2.4 GHz ISM band. There are 13 channels available in Europe, see Section 3.2.3.1.

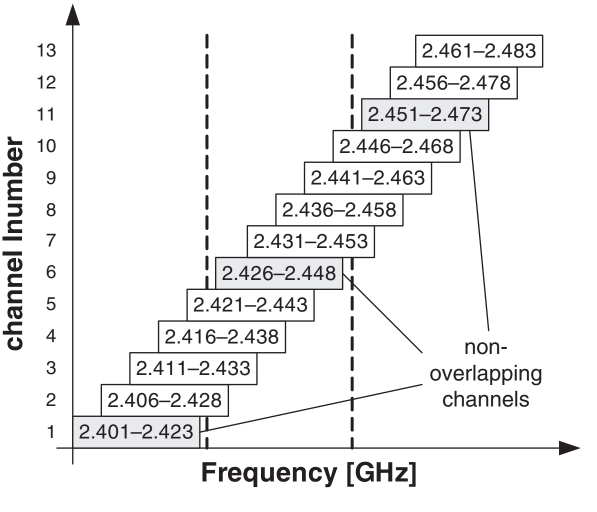


Figure 5.5 The spectrum use of different IEEE 802.11 channels in the 2.4 GHz ISM band.

5.3.1 *Frequency Hopping, Direct Sequence Spread Spectrum, and Infrared*

With FHSS, groups of communicating stations operate on a common frequency channel for short times only, before selecting another frequency channel ("hopping") to continue the communication. Which frequency channel will be used is known to all stations of the group, and follows a pseudo-random list of frequencies. Different sets of communicating stations use different lists of frequencies, which reduces the probability that they operate with the same radio resources, i.e., at the same center frequency.

In contrast to FHSS, in DSSS all stations operate at the same center frequency. In DSSS, different spreading codes allow different sets of communicating stations to reduce the mutual interference. There are three non-overlapping channels available in IEEE 802.11b (the frequency spectrum used by the different 802.11b and 802.11g channels is shown in Figure 5.4).

FHSS and DSSS in 802.11 are not CDMA, because all stations operate with the same code sequence. The objective of a spread spectrum is to evenly distribute radio emissions over some broader bandwidths in the spectrum, in order to facilitate spectral coexistence with dissimilar radio systems such as Bluetooth. This is a regulatory requirement for the ISM spectrum. Other newer physical layers, for example the attractive 802.11 g OFDM (IEEE, 2003e), do not have to apply spread spectrum, because of the flat shape of their emitted signals.

The original 802.11 standard also specifies an IR physical layer, which has so far not been commercially successful, but may be attractive for future residential in-house applications, where users often prefer light over radio emissions.

5.3.2 *802.11B Complementary Code Keying, CCK*

This is used in the IEEE 802.11 supplement standard "Higher-Speed Physical Layer Extension in the 2.4 GHz Band", known as IEEE 802.11b-1999 (IEEE, 1999b) which is now part of the *IEEE Wireless LAN Edition* (IEEE, 2003a), here referred to as 802.11b. The 802.11b supplement standard defines the High Rate Direct Sequence Spread Spectrum (HR/DSSS) transmission mode with a chip rate of 11 Mchip/s, providing the same occupied channel bandwidth and channelization scheme as DSSS. The higher data rate is achieved through a transmission mode based on eight-chip Complementary Code Keying (CCK) modulation. The code set of complementary codes is richer than the set of Walsh codes that are used in DSSS. At 11 Mb/s, the spreading code length is 8 and the symbol duration is 8 instead of 11 chips (11 chips are used for DSSS). Data bits encode the symbols with Quaternary Phase Shift Keying (QPSK) and Differential QPSK (DQPSK).

The 802.11b standard extension was first known under the acronym Wireless Fidelity (Wi-Fi) and was the first to achieve significant commercial success.

5.3.3 *802.11A/G Orthogonal Frequency Division Multiplexing*

Orthogonal Frequency Division Multiplexing (OFDM) is the transmission scheme more and more used by wireless systems. 802.11a, 802.11 g, and 802.11n are based on OFDM.

802.11a as a supplement of 802.11 specifies the PHY for transmission at 5 GHz OFDM (IEEE, 2003a). In IEEE 802.11a, one 52-subcarrier OFDM symbol occupying 16.6 MHz has a duration of 4 μs. It operates with a 52-carrier OFDM system with convolution coding and linear modulation schemes that can be adaptively chosen based on current QoS requirements and radio channel conditions – 48 carriers are used for user data transport and four subcarriers of the OFDM symbols are used for pilot symbols.

OFDM Parameters of IEEE 802.11a

Parameter	Value
Sampling rate $1/T$	20 MHz
OFDM block duration T_b	$64^*T = 3.2$ us
Guard interval duration T_g	$16^*T = 0.8$ us
OFDM symbol duration $T'_b = T_g + T_b$	$80^*T = 4$ us
Number of data sub-carriers	48
Number of pilot sub-carriers	4
Sub-carrier spacing D_f	$1/T_b = 0.3125$ MHz
Spacing betw. the outmost sub-carriers	$(N_{total} - 1)^*D_f = 15.9375$ MHz

PHY mode R	$\mathbf{P}_{\min}(R)$(dBm)	$\mathbf{SINR}_{\min}(R)$(dBm)
BPSK 1/2	−82	18
BPSK 3/4	−81	21
QPSK 1/2	−79	22
QPSK 3/4	−77	25
16QAM 1/2	−72	25
16QAM 3/4	−70	32
64QAM 2/3	−66	34
64QAM 3/4	−65	35

Figure 5.6 Values for the OFDM parameters of IEEE 802.11a (and HiperLAN/2).

The upper table in Figure 5.6 summarizes numerical values for the main parameters of the 802.11a OFDM transmission system, which are identical for most of the 802.11 g, operating in the 2.4 GHz ISM band. Additionally, the standard defines the minimum receiver input level sensitivity and SINR to receive a *1000 byte* frame with more than 90 % success probability, see lower table in Figure 5.6.

5.4 Medium Access Control Protocol

The 802.11 MAC protocol is built with the help of two coordination functions, i.e., the Distributed Coordination Function (DCF) for traffic without QoS, referred to as asynchronous services, and the Point Coordination Function (PCF) for traffic with QoS requirements, referred to as synchronous services. The coordination functions are discussed in the following.

5.4.1 Distributed Coordination Function

In the following, an *infrastructure-based* BSS of 802.11 is considered, which is composed of an AP and a number of stations associated to the AP. The AP connects the stations with the infrastructure. The basic 802.11 MAC protocol is the DCF that works as a listen-before-talk scheme, based on the Carrier Sense Multiple Access (CSMA) introduced in Section 2.6.2. Stations deliver MSDUs of arbitrary lengths (up to *2304 byte*), after detecting that there is no other transmission in progress on the radio channel.

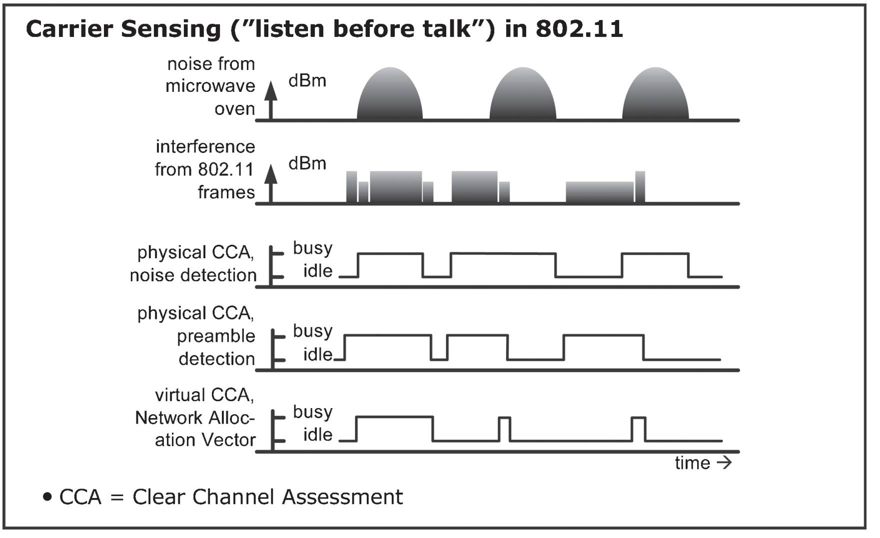


Figure 5.7 802.11 MAC performs a clear channel assessment before each frame exchange either based on noise detection, preamble detection, or through NAV reservation.

5.4.1.1 Listen Before Talk

The channel sensing function is called Clear Channel Assessment (CCA). It uses a single fixed-power threshold, which is −82 dBm according to 802.11, but may be implementation dependent. If a station detects a signal with power larger than this threshold, the radio channel is assumed to be busy and thus unavailable for transmission. Otherwise, the radio channel is assumed to be idle. There are different variants of CCA. Either all signals are evaluated (ordinary noise detection), or only signals from other 802.11 stations are evaluated (noise plus preamble detection). See Figure 5.7 for an illustration of the CCA.

The Network Allocation Vector (NAV) is illustrated in Figure 5.7. It is an addition to the physical sensing of the radio channel. It is referred to as virtual carrier sensing and in fact has the function of reserving the channel for some time. The NAV is a timer, which decrements irrespectively of the status of the medium, which can be busy or idle.

The NAV is set when a frame is received, which includes a duration field that defines how long the following frame exchange may take. As long as the NAV is set or the CCA sensed the radio channel as being busy, a station is not allowed to initiate transmissions. Thus, upon frame reception, the NAV can be eventually set for a duration that is longer than the transmission duration of this frame, and subsequent frame transmissions are protected.

5.4.1.2 Timing and Interframe Spaces

Each successful reception is acknowledged by the receiving station, as indicated in Figure 5.8 and Figure 5.9. The addressed station transmits an Acknowledgment (ACK) immediately after receiving a frame. The time between two MAC frames is called the Interframe Space (IFS). 802.11 defines four different IFSs.

Short Interframe Space (SIFS), Point Coordination Function Interframe Space (PIFS), and Distributed Coordination Function Interframe Space (DIFS) are used under normal conditions

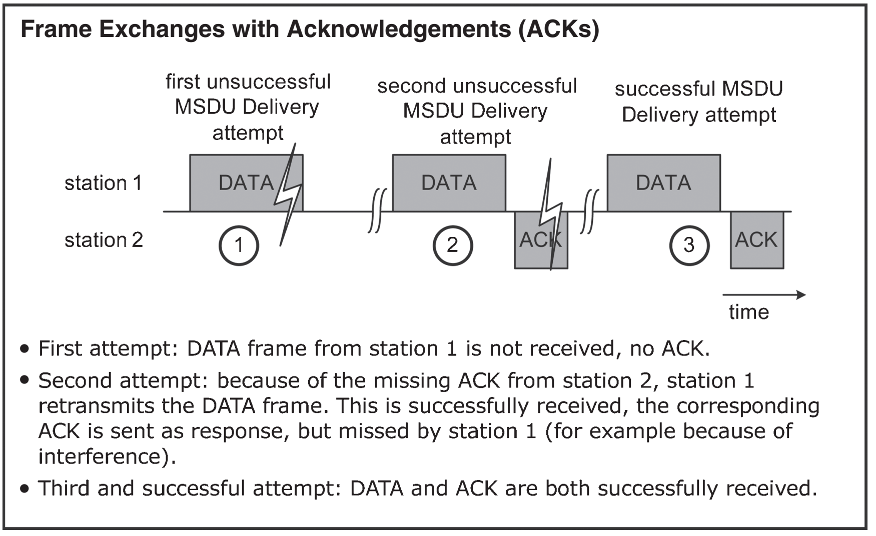


Figure 5.8 Frame exchanges and acknowledgments. Successful receptions are acknowledged by the receiving station.

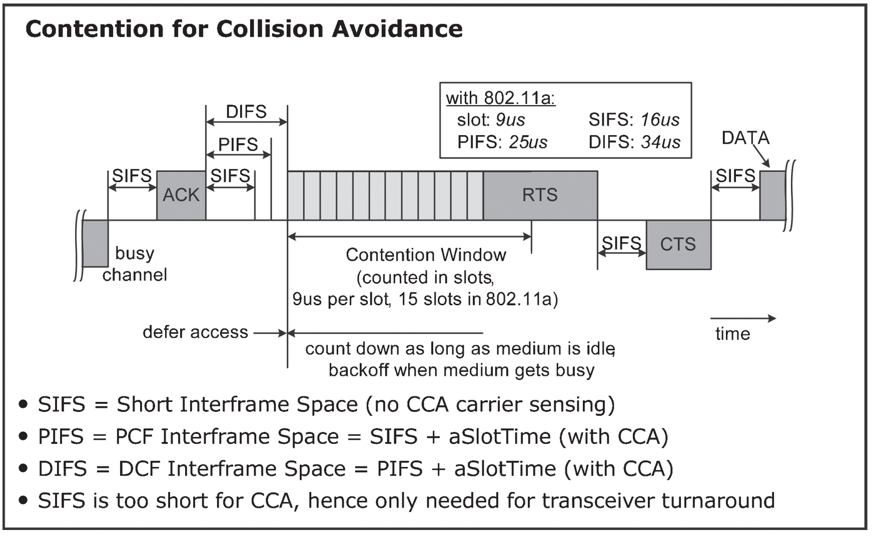


Figure 5.9 Before each transmission, a backoff is applied for a number of slots duration within the limits of the contention window.

and represent three different priority levels for medium access. The shorter the IFS, the higher the priority in medium access. The fourth IFS, called Extended Interframe Space (EIFS), is used when a station detects an ongoing transmission as being interfered, assuming that there are some stations that cannot detect each other. A hidden station scenario is then assumed, and the station has to defer from channel access for a longer time duration.

All interframe spaces are independent of the channel data rate. Due to the different characteristics of the different PHY specifications, the durations of the interframe spaces depend on the used transmission scheme. The relations between the IFS and the duration aSlotTime (also referred to as slot time) are shown in Figure 5.9. The IFSs are listed in order, from shortest to longest as follows:

- **aSlotTime:** The duration aSlotTime is used to calculate the IFSs. SIFS and aSlotTime are the basis of all other durations. In 802.11a, aSlotTime is 9 μs. As the name indicates, aSlotTime is used during the Collision Avoidance (CA). The CA is explained below.
- **SIFS:** The SIFS is used to prioritize the immediate Acknowledgment (ACK) frame of a data frame, the response (Clear To Send (CTS) frame) to a Request To Send (RTS) frame, a subsequent MPDU of a fragmented MSDU, response to any polling using the PCF, and any frames of the AP during the Contention Free Period (CFP). RTS and CTS are explained below. SIFS is 16 μs for 802.11a.
- **PIFS:** The PIFS is used by stations operating under the PCF to obtain channel access with highest priority. PIFS is calculated as PIFS = SIFS + aSlotTime and is 25 μs for 802.11a.
- **DIFS:** The DIFS is used by stations operating under the DCF to obtain channel access for frame exchanges. DIFS is calculated as: DIFS = SIFS + 2 · aSlotTime. DIFS is 34 μs for 802.11a.
- **EIFS:** The EIFS is used instead of DIFS whenever the PHY indicates that a frame transmission did not result in a correct Frame Check Sequence (FCS). The EIFS is therefore used when multiple stations initiated frame exchanges at different starting times. This occurs typically when these stations are hidden from each other. The EIFS is an extended interframe space resulting in a longer deferral from channel access, which gives other stations clearly a higher priority in medium access. As soon as one other frame is received correctly, DIFS is used again. EIFS is around 200 μs for 802.11a.

5.4.1.3 Collision Avoidance

As part of the DCF, it may happen that more than one station attempts to transmit at the same time. This is called a collision. In wireless communication, a transmitter cannot detect a collision at a receiver while transmitting. To account for this, 802.11 is based on Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA).

If two or more stations detect the channel as being idle at the same time, inevitably a collision occurs when these stations initiate a frame exchange at the same time. The 802.11 defines a CA mechanism to reduce the probability of such collisions. As part of CA, a station performs the so-called backoff procedure before starting a transmission. A station that has an MSDU to deliver has to keep sensing the channel for an additional random time duration after detecting the channel as being idle for the minimum duration DIFS, which is 34 μs for 802.11a. Only if the channel remains idle for this additional random time duration, is the station allowed to initiate its transmission. The duration of this random time is determined as a multiple of a slot duration (aSlotTime). Each station maintains a so-called Contention Window (CW), which is used to determine the number of slot times a station has to wait before transmission. Figure 5.9 shows an example: after a successful frame exchange, i.e., after the ACK transmission, a station starts the next frame exchange (RTS frame followed by CTS frame), because the radio channel has been idle for a duration equal to DIFS and its following backoff slots. The contention window size increases when a transmission fails, i.e., when the transmitted RTS or data frame has not been acknowledged.

After an unsuccessful transmission, the next backoff is performed with a doubled size of the contention window. This reduces the collision probability if there are multiple stations attempting to access the channel. The stations that deferred from channel access during the channel busy period do not select a new random backoff time, but continue to count down the time of

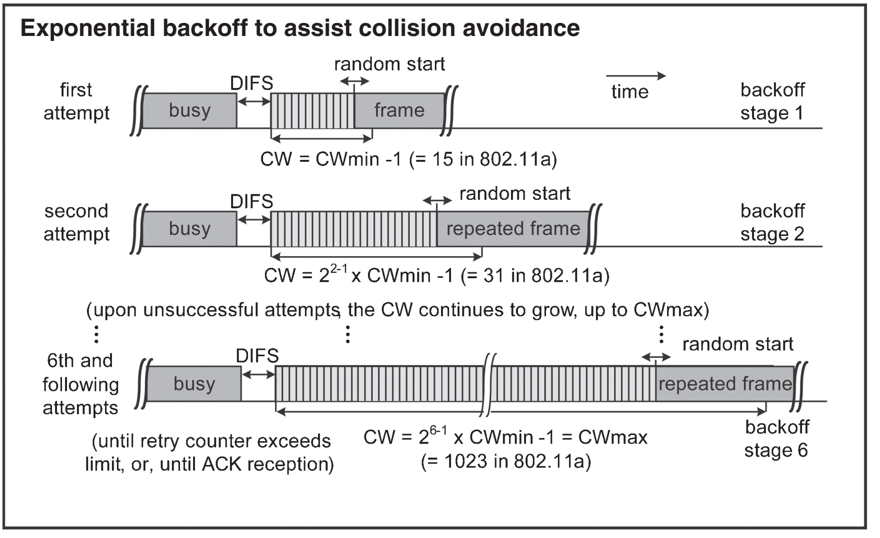


Figure 5.10 Increase of contention window size after unsuccessful frame exchanges. The size is doubled for each new attempt of collided or erroneously received MSDUs, up to a certain limit. The actual numbers vary with the PHY specifications.

the deferred backoff in progress after sensing the channel as being idle again. In this way, stations, which deferred from medium access because their random backoff time was larger than the backoff time of other stations, are given a higher priority when they resume the transmission attempt.

Figure 5.10 illustrates the increase of the contention window upon unsuccessful transmissions. Note that a station cannot differentiate between collision and failed transmission due to errors on the wireless channels. A missed ACK frame will always be interpreted as collision.

5.4.1.4 Recovery Procedure and Retransmissions

When a frame exchange is not successful, i.e., when a transmitting station does not receive an ACK frame immediately after the frame transmission, the frame size of the transmitted frame is compared against a threshold value before retransmission. All unsuccessful transmissions of frames with a frame size shorter than the threshold value, and all failed RTS transmissions, increment the Short Retry Counter (SRC). If the SRC reaches a limit (default: 7), the frame is discarded. All unsuccessful transmissions of frames with a frame size larger than the given threshold, increment the Long Retry Counter (LRC). Again, no more retransmission attempts are made when LRC is equal to a limit (default: 4). Whenever an MSDU is successfully transmitted, SRC and LRC are reset. The actual value of the threshold is implementation dependent.

5.4.1.5 Post-backoff

After each successful transmission, it is mandatory that another random backoff is performed by the transmission-completing station, even if there is no other MSDU to be delivered. This is referred to as “post-backoff,” because this backoff is performed after, not before, a transmission. This can be interpreted as the backoff for the next MSDU delivery. By using this post-backoff, it is guaranteed that any frame (with the exception of the first MSDU in a burst, arriving at an empty queue and during an idle phase) will be delivered with backoff. An MSDU arriving at the station from the higher layer may be transmitted immediately without waiting any time, if the transmission queue is empty, the latest post-backoff has been finished already, and at the same time, the channel has been idle for a minimum duration of DIFS. This helps to reduce the delivery delay in lightly loaded systems.

5.4.1.6 Fragmentation

To reduce the duration the channel is occupied when frames collide, data frames (MSDUs) can be transmitted in more than one MPDU, if their length exceeds a certain threshold. The process of partitioning an MSDU into smaller MPDUs is called fragmentation. See Figure 5.11 for an illustration of fragmentation, where the complicated protection of frames by the NAV vectors is also illustrated. An MPDU protects the subsequent transmission of its ACK within its duration field, and in addition, when fragmentation is used, transmission of the following MPDU.

Fragmentation creates MPDUs smaller than the original MSDU length to limit the probability of long MPDUs colliding and being transmitted more than once. With fragmentation, a large MSDU can be divided into several smaller data frames, i.e., fragments, which can then be transmitted sequentially as individually acknowledged frames. The benefit of fragmentation is that in the case of failed transmission, the error is detected earlier and there is less data to retransmit. It also increases the probability of successful transmission of the MSDU in scenarios where the radio channel characteristics cause higher error probabilities for longer frames than can be expected for shorter frames. Each fragment can be transmitted sequentially as an individually acknowledged data frame. The obvious drawback is the increased overhead. The process of recombining MPDUs into a single MSDU is called defragmentation, which is accomplished at each receiving station. Only MPDUs with a unicast receiver address may be fragmented. Broadcast/multicast frames may not be fragmented even if their length exceeds the implementation dependent threshold. Note the maximum length of an MSDU is limited to *2346 byte*.

Fragmentation cannot be applied to deliver multiple MSDUs in sequence pending at a transmitter to the same (or even different) receiver(s).

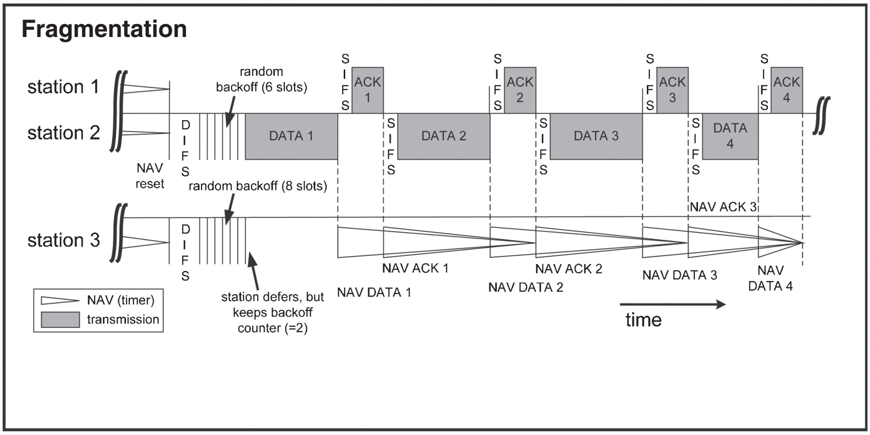


Figure 5.11 Fragmentation. Data frames protect the subsequent transmissions of their ACK responses and the following data frame with the NAV.

5.4.1.7 Hidden Stations and RTS/CTS

In wireless communication systems that use carrier sensing, the so-called hidden station problem can occur, depending on the locations of the stations. This problem arises when a station is able to successfully receive frames from two different stations but the two stations cannot detect each other.

When stations cannot detect each other, a station may sense the channel as idle even when other hidden stations are transmitting. It may initiate a transmission while the other station is already transmitting. This may result in a collision and severely interfered frames at stations that can detect coinciding transmissions of hidden stations.

To reduce throughput reduction owing to hidden stations, 802.11 specifies as an option the exchange of *Request-to-Send/Clear-to-Send* (RTS/CTS) frames. Before transmitting a data frame, a station may transmit a short RTS frame, which must be followed by a CTS frame transmitted by the receiving station. Consecutive frames in the sequence of RTS, CTS, data, and ACK are spaced by an SIFS duration (16 us for 802.11a) for transceiver turn-around. It is a decision made locally by the transmitting station, if or if not RTS/CTS is used. Upon receiving an RTS frame, the receiving station must reply with a CTS frame. The RTS and CTS frames carry in its duration fields the information, how long the sequence RTS, CTS, Data, ACK will take. Hence, other stations close to the transmitting station and hidden stations close to the receiving station will not start any transmissions; their NAV timer is set. A hidden station close to the receiving station might not receive the RTS due to the large distance, but will in most cases receive the CTS frame.

See Figure 5.12 for an example of the DCF using RTS/CTS. It is important to note that SIFS is shorter than DIFS, which always gives CTS and ACK the highest priority for access to the radio channel.

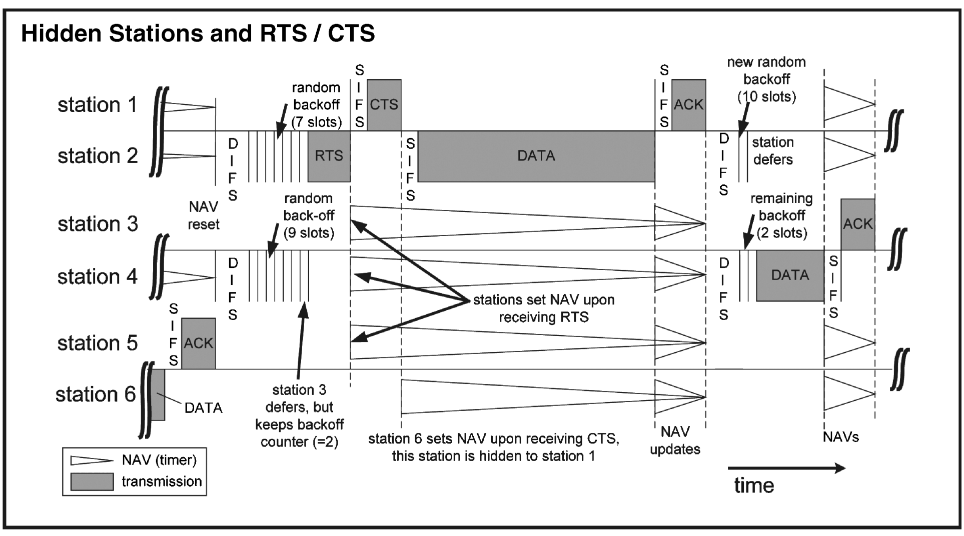


Figure 5.12 Timing of frame exchanges and NAV settings of the 802.11 DCF. Station 6 cannot detect the RTS frame of the transmitting station 2, but can detect the CTS frame of station 1. Although station 6 is hidden to station 1, it refrains from channel access because of NAV.

5.4.2 Synchronization and Cell Search

All stations within a single BSS are synchronized to a common clock by maintaining a local timer using the Timing Synchronization Function (TSF). To synchronize the stations, a management frame, the beacon, is used.

Beacons are transmitted periodically, hence, every station knows when the next beacon frame will arrive; this time is called Target Beacon Transmission Time (TBTT). The TBTT of each beacon is announced in the previous beacon.

The TSF's original function is to support various PHYs that require synchronization, and management functions such as a station joining a BSS, and saving power through sleep modes. Local timers are being updated by the information received from other stations as part of a beacon. To give beacon transmission highest priority for medium access, stations stop initiating frame exchanges upon reaching a TBTT. However, ongoing frame exchanges are completed, even if this means that beacon transmissions are delayed. Note that beacons are transmitted after the channel was idle for PIFS (25 us in 802.11a), and in a BSS without backoff. Thus, if a frame exchange is ongoing at TBTT, then the beacon is delayed. In BSS and IBSS, the synchronization is maintained by broadcasting the TSF timer in the beacon. When receiving this broadcast, the decision whether the local timer in a station has to be updated or not is different for BSS and IBSS, as described in the following.

Figure 5.13, left, illustrates the TSF in an infrastructure BSS. * Only the AP generates beacons in a BSS. At each TBTT, the AP schedules a beacon as the next frame to be transmitted. If the channel has been idle for at least PIFS before TBTT, the AP transmits the beacon at TBTT, otherwise the

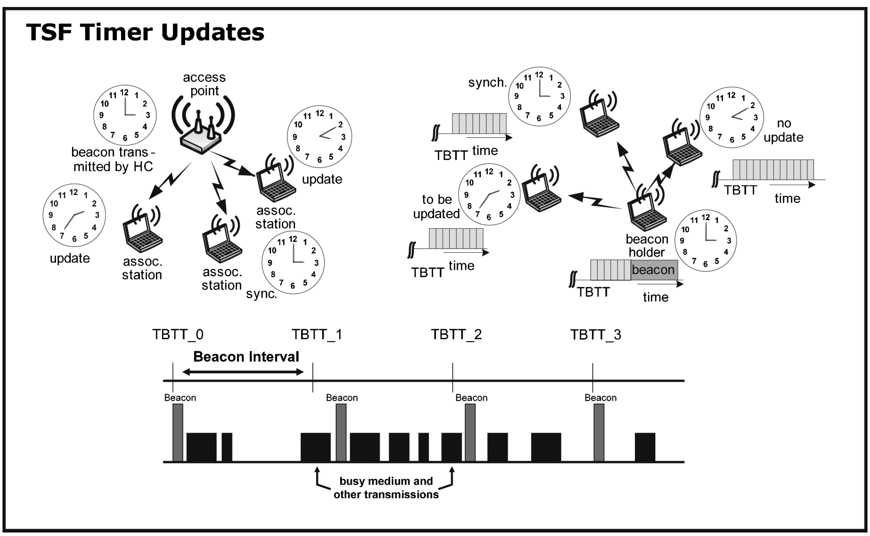


Figure 5.13 Left, top: Station 2 and 3 change their TSF timers to the value received in the beacon of the AP. Right, top: In IBSS, station 2 transmits the beacon because of the smaller backoff counter. Bottom: Target and actual beacon transmission time vary with the number of frame exchanges.

* An infrastructure BSS is referred to as BSS and an independent BSS is referred to as IBSS.

beacon is transmitted PIFS after the current transmission, immediately. All stations associated to this AP update their local timers with the information received from the beacon.

Figure 5.13, right, illustrates the TSF in an IBSS. There, the TSF is distributed over all stations. All stations take part in the generation of beacons. The beacon generation is distributed using a mechanism similar to the backoff mechanism.

At TBTT, stations that are part of an IBSS attempt to transmit a beacon after PIFS in contention, with small CWmin. Stations stop attempting to transmit a beacon when they receive a beacon from another station of the IBSS. However, beacons transmitted in contention may collide, which is allowed as part of the standard. A collided beacon is not retransmitted; the next beacon will be transmitted at the next TBTT. Note that beacons are not acknowledged by other stations. A station that transmitted a colliding beacon will not detect this collision, as it is not waiting for a subsequent ACK frame.

In BSS and IBSS, upon receiving a beacon, a station updates its local timer with the information gained from the beacon only if the received value represents an earlier time than the value currently maintained in the local timer. This is also indicated in Figure 5.13. This distributed synchronization function results in the shared information about the fastest running clock, with which the complete IBSS will synchronize.

Among the timing information needed to synchronize stations, the beacon delivers other parameters related to the protocol and to radio regulations. In addition to the timing information needed to synchronize the BSS, the beacon delivers protocol-related parameters, for example Basic Service Set Identification (BSSID), the beacon interval (next TBTT), PHY depending parameters, the duration of the Contention Free Period (CFP), and regulatory and spectrum related management information, such as the available channels and the power limits.

Depending on the type of the BSS, not all information may be contained in what is broadcasted across the BSS in the beacon. In the case of an infrastructure BSS the AP uses the beacon for instructions to its associated stations and announcements of future transmissions, e.g. delivery of multicast traffic to stations in power save mode. Stations may also use the signal strength of the received beacons to decide when to disassociate, because of channel conditions, and to which AP to associate again.

All stations within a single BSS are synchronized to a common clock. Each station maintains such a timer with modulus 2^{64} counting in increments of microseconds. This allows not only incremental but also absolute scheduling of future tasks. In an infrastructure-based network, TSF is performed by the AP that serves as timing master. Independent basic service sets with no central control instance apply a distributed TSF algorithm.

Besides the timestamp, the beacon interval field is of particular interest. It reports the number of time units between two *target* beacon transmissions. As shown in Figure 5.13, bottom, beacon transmission may be delayed due to busy medium. This is why 802.11 specifies Target Beacon Transmission Times (TBTT). If a beacon is delayed and cannot be sent at the actual TBTT_1, it will be conveyed as soon as the medium is idle. The next beacon will be scheduled at TBTT_2 = TBTT_1 + Beacon Interval. If the medium is busy at TBTT_2, a delayed beacon is sent again and the next one is scheduled for TBTT_3. If the medium is idle at that time, TBTT_3 and actual beacon transmission will coincide, regardless of any delayed previous beacon transmission. Due to this mechanism, an absolute time frame is established decoupling target beacon transmissions from current load. In principle, beacon intervals can comprise any time span between 1 TU ($\approx$1ms) and $2^{16} - 1$ TUs ($\approx$67s) with a granularity of time units. However, the specification does not prescribe any particular period. Currently, commercial IEEE 802.11a,b,g access points are shipped with a default beacon interval of around 100 ms.

Knowledge about TBTT allows associated stations to operate in power save mode. While being in *Doze* state, a station is not able to transmit or receive and consumes only very low power. Switching to the *Awake* state hence is only necessary for selected beacons. Traffic indication messages conveyed together with each beacon inform particular stations an buffered MSDUs and an upcoming multicast/broadcast messages. Depending on this information, a station may switch back to *Doze* state or stay awake for further actions.

5.4.3 Scanning Procedures in WLAN 802.11

IEEE 802.11 defines two scanning modes: *passive* and *active* scanning. The decision on which scanning mode is to be applied is taken by the *Station Management Entity* (SME). The SME is a layer-independent entity that may be viewed as residing in a separate management plane. The exact functions of the SME are not specified by the standard. In general, this entity is responsible for tasks such as the gathering of layer-dependent status information from the various layer management entities and setting the values of layer-specific parameters. Hence, the SME executes actions related to general system management. It directs the MAC Layer Management Entity (MLME) to perform either passive or active scanning with scanning directives such as a dedicated (B)SSID, a channel list or timing constraints for scanning (MinChannelTime, MaxChannelTime) can be requested. Similarly to the beacon interval, any ChannelTime parameter with respect to scanning periods is not standardized.

5.4.3.1 Passive Scanning

If passive scanning is requested, a station monitors each specified channel for a time span of at most *MaxChannelTime*. If no specific (B)SSID or channel list was requested by the SME, the station needs to survey each channel from the valid channel range for the appropriate PHY and carrier set. During the scanning, the station adds any received 802.11 beacon or probe response to its cached BSSID scan list.

Depending on the number of channels to be scanned, the beacon interval of a single BSS, the beacon interval synchronization of different BSSs, and the load of the system, the overall duration for passive scanning (scanning delay) may vary.

5.4.3.2 Active Scanning

Active scanning requires a station to generate specific request messages, so-called *probe frame*s, with subsequent processing of incoming *probe response* frames. The probe mechanism forces an AP to convey basically the same system information as done with the periodic beacon signal. In providing this information promptly on request, unnecessary waiting times of up to one entire beacon interval can be prevented. This is how active scanning supports accelerated information provision, e.g. in the scope of supporting seamless handover. In addition, probe response frames need to be acknowledged by a station to ensure integrity of data delivery. With passive scanning, ordinary beacons are sent as broadcast transmission such that a station might not receive full system information due to interference caused e.g. by hidden stations.

5.5 Medium Access Control with Support for Quality-of-Service[1]

5.5.1 Point Coordination Function

To support time-bounded services, the IEEE 802.11 standard defines the Point Coordination Function (PCF) to let stations have priority access to the radio channel, coordinated by a station called the Point Coordinator (PC). The PC typically resides in the AP.

The PCF has higher priority than the DCF, because the period during which the PCF is used is protected from the DCF access by the NAV. The time during which 802.11 stations operate is divided into repeated periods, called superframes. A superframe starts with a beacon. With an active PCF, a Contention Free Period (CFP) and a Contention Period (CP) are alternating over time, where a CFP and the following CP form a superframe. During the CFP, the PCF is used for accessing the channel, while the DCF is used during the CP. It is mandatory that a superframe includes a CP of a minimum length that allows at least one MSDU delivery under DCF. A superframe starts with a beacon frame, regardless if the PCF is active or not. The beacon frame is a management frame that maintains the synchronization of the local timers in the stations and delivers protocol related parameters, as explained earlier.

The PC, which is typically co-located with the AP, generates beacon frames at regular beacon frame intervals, thus every station knows when the next beacon frame will arrive. During CFP, there is no contention among stations; instead, stations are polled. See Figure 5.14 for a typical frame exchange sequence during CFP. The PC polls a station asking for a pending frame. It is assumed in the example that the PC itself has data pending for this station and uses a combined data and poll frame by piggybacking the *CF-Poll* frame into the data frame. No idle period longer than PIFS occurs during CFP.

The PC continues with polling other stations until the CFP expires. A *CF-End* control frame is transmitted by the PC as the last frame within the CFP to signal the end of the CFP.

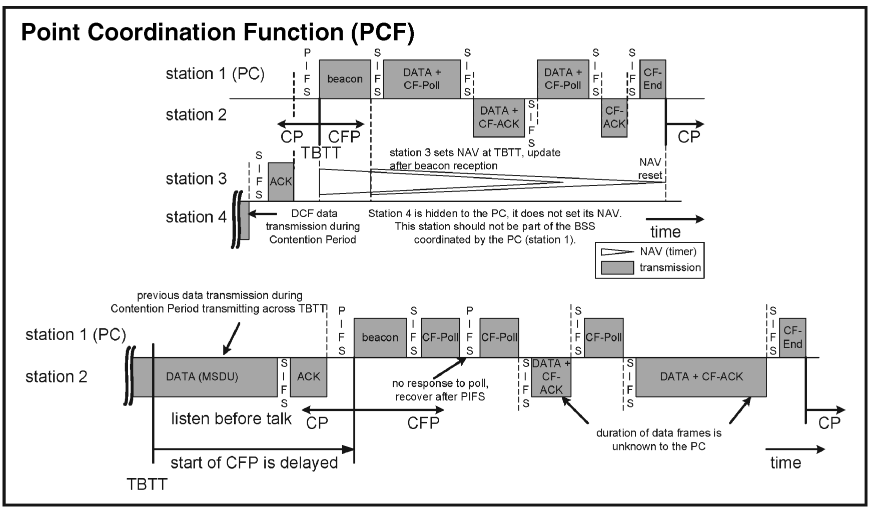


Figure 5.14 Example for the PCF operation. Station 1 is the PC and polls station 2. Station 3 detects the beacon frame and updates the NAV for the whole CFP. Start of the CFP may be delayed, see lower part of the Figure.

5.5.2 QoS Support with PCF

There are problems with the PCF that led to the current activities within the IEEE 802.11 working group to enhance the protocol. Among many others, these include the unpredictable beacon delays and unknown transmission durations of the polled stations. At TBTT, a PC schedules the beacon as the next frame to be transmitted, but the beacon can only be transmitted when the medium has been determined to be idle for at least PIFS. For the legacy 802.11 standard, stations can start their transmissions even if the MSDU delivery is not finished before the upcoming TBTT. Depending on whether the wireless medium is idle or busy at TBTT, a delay of the beacon frame may occur. The time the beacon frame is delayed from TBTT determines the delay of the transmission of time-bounded MSDUs that have to be delivered in the CFP. This may severely affect the QoS as this introduces unpredictable time delays in each CFP. Beacon frame delays of around 4.9 ms are possible in 802.11a in the worst case (longest MSDU, fragmentation, RTS/CTS, most robust modulation and coding scheme).

Another problem with the PCF is the unknown transmission time of polled stations. A station that has been polled by the PC is allowed to deliver an MSDU that may be fragmented and of arbitrary length, up to the maximum of *2304 byte* (*2312 byte* with encryption). Further, different modulation and coding schemes are specified in 802.11a. As a result, the duration of the MSDU delivery after polling is not under the control of the PC, which reduces the QoS provided to other stations that are polled during the rest of the CFP.

5.5.3 QoS Support Mechanisms of 802.11E

Enhancements to the above-described 802.11 MAC led to the 802.11e extension (IEEE, 2005b) of the 802.11 standard. The 802.11e extension introduces the Hybrid Coordination Function (HCF) for QoS support. The HCF defines two medium access mechanisms that are referred to as: (i) the contention-based channel access and (ii) the controlled channel access (which includes polling). Note that 802.11e uses "channel access" as a synonym for "medium access". The contention-based channel access is referred to as Enhanced Distributed Channel Access (EDCA) whereas the controlled channel access is referred to as HCF Controlled Channel Access (HCCA). With 802.11e, there may still be the two phases of operation within a superframe, i.e., CP and CFP. The EDCA is used only in the CP, while the HCCA is used in both phases. Figure 5.15 illustrates the main elements of the 802.11e MAC architecture in the context of 802.11. The legacy 802.11 DCF is the basis for the contention-based access of the HCF and the legacy 802.11 PCF. The PCF offers contention-free services to legacy stations and is used by the HCF for polling stations as discussed below. Both HCCA as well as EDCA use the known 802.11 MAC frames of the DCF to transmit user data on the radio channel, namely the DATA/ACK frame exchange sequence with an optional preceding RTS/CTS.

Stations operating under the 802.11e protocol are referred to as 802.11e stations in this chapter. The station that operates as the central coordinator for all other stations within the same QoS supporting BSS (QBSS) is called the Hybrid Coordinator (HC). Similar to the PC, the HC resides within an 802.11e AP. A BSS that includes an 802.11e-compliant HC is referred to as a QBSS. There are multiple backoff processes operating in parallel within one 802.11e station, which will be explained later on. Therefore, in the following we refer to backoff entities that attempt to deliver MSDUs, instead of stations.

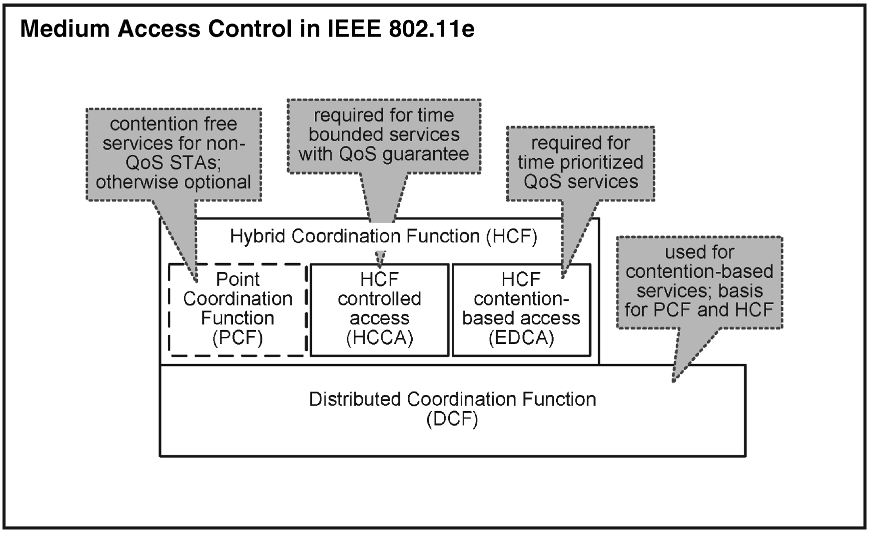


Figure 5.15 Medium Access Control architecture of IEEE 802.11e in extending 802.11.

5.5.4 Improvements of the Legacy 802.11 MAC

An 802.11e station (more precisely, a backoff entity) that obtained medium access must not utilize radio resources for a duration longer than a specified limit. This important new attribute of the 802.11e MAC is referred to as a Transmission Opportunity (TXOP). A TXOP is an interval of time during which a backoff entity has the right to deliver MSDUs. A TXOP is defined by its starting time and duration. TXOPs, which are obtained via the contention-based medium access, are referred to as EDCA-TXOPs. Alternatively, a TXOP, which is obtained by the HC via the controlled medium access, is referred to as an HCCA-TXOP or a polled TXOP. In general, a time period in which the HC has control over the wireless medium (in the CFP or gained in the CP) is referred to as the Controlled Access Phase (CAP). Such TXOPs obtained by the HC are protected with the help of the NAV, introduced in Section 5.4.1, even if no RTS/CTS is used.

The duration of an EDCA-TXOP is limited by a QBSS-wide parameter referred to as *TXOPlimit*. This *TXOPlimit* is distributed regularly by the HC within an information field of the beacon.† However, in the absence of legacy stations, the *TXOPlimit* allows the control of the maximum time a backoff entity allocates the medium for MSDU delivery, and therefore is an important means of controlling the MSDU delivery delay.

Another enhancement proposed during the standardization process is not part of the final standard: It was intended to not allow any EDCA backoff entity to transmit across the TBTT. That is, a frame exchange is initiated only if it can be completed before the upcoming TBTT. This reduces the expected beacon delay, which would give the HC a better control over the medium especially if the optional CFP is used after the beacon transmission. However, as EDCA-TXOPs

† Legacy stations will only understand the fields known from the legacy standard, whereas 802.11e backoff entities additionally will understand all new information fields. The new information fields are ignored by legacy stations. Therefore, legacy stations may transmit for longer durations than allowed by the TXOPlimit.

do not respect the upcoming TBTT, additional procedures are required to guarantee a timely initialization of the next superframe, especially in the context of coexistence and interworking of different 802 based wireless systems, see Sections 8.3 and 9.2.

Additionally, an 802.11e backoff entity is allowed to transmit frames directly to another backoff entity in a QBSS, without involving communication with the AP. In the legacy 802.11 protocol, within an infrastructure-based BSS, all data frames are either sent or received by the AP. For this purpose, an 802.11e station needs to establish a direct link with another 802.11e station using the Direct Link Protocol (DLP) before initiating direct frame transmissions.

5.5.5 Contention-based Medium Access

The QoS support in EDCA is provided by the introduction of Access Categories (ACs) and multiple independent backoff entities. MSDUs are delivered by parallel backoff entities within one 802.11e station, where backoff entities are prioritized using AC-specific contention parameters, called the EDCA parameter set. There are four ACs, thus, four backoff entities exist in every 802.11e station. The ACs are labeled according to their target application, i.e., AC_VO (voice), AC_VI (video), AC_BE (best effort), and AC_BK (background). These ACs result from a mapping of the user priorities from Annex H.2 of IEEE 802.1D (IEEE, 1998), as defined in Figure 5.16(a) from IEEE (2005b). See Figure 5.17 for an illustration of the parallel backoff

(a) Mapping of User Priorities

Priority	802.1D User Priority	802.11e Access Category (AC)	Service Type
lowest	1	AC_BK	background
	2	AC_BK	background
	0	AC_BE	best effort
	3	AC_BE	best effort
	4	AC_VI	video
	5	AC_VI	video
	6	AC_VO	voice
highest	7	AC_VO	voice

(b) Default Values of EDCA Parameter Sets

AC	CWmin	CWmax	AIFSN	AIFS*	TXOPlimit
Legacy 802.11	15	1023	2	34 us	
AC_BK	15	1023	7	79 us	0/0
AC_BE	15	1023	3	43 us	0/0
AC_VI	7	15	2	34 us	6.016/3.008 ms
AC_VO	3	7	2	34 us	3.264/1.504 ms

Figure 5.16 (a) Mapping of user priorities (IEEE 802.1D) to access categories from (IEEE, 2005b). (b) Default values of EDCA parameters based on (IEEE, 2005b). Star indicates dependency on PHY, here 802.11b/802.11a are selected.

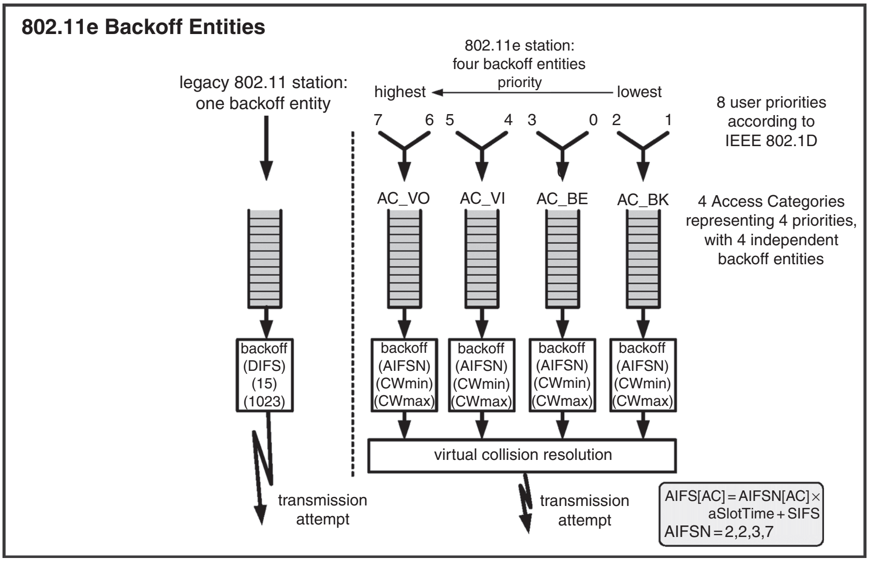


Figure 5.17 Legacy 802.11 station and 802.11e station with four ACs within one station.

entities. These EDCA parameter sets modify the backoff process with individual interframe spaces and contention windows per AC introducing a probability-based prioritization as explained below.

5.5.6 EDCA Parameters Per AC

Contention-based medium access is performed in every backoff entity by using different values for the EDCA parameter set. The EDCA parameters of each backoff entity are defined by the HC. Default values for the EDCA parameters are given in Figure 5.16(b) according to IEEE (2005b). The EDCA parameter set can be modified over time by the HC, and is announced via information fields in beacon frames. The same EDCA parameter set is used by the backoff entities of the same AC in different stations. It is essential that the same values for the parameters are used by all backoff entities.

Each backoff entity within a station independently contends for a TXOP. It starts down-counting the backoff counter after detecting the medium being idle for a duration defined by the Arbitration Interframe Space (*AIFS*[*AC*]) instead of DIFS, which is used by legacy stations. The *AIFS*[*AC*] is at least DIFS, and can be enlarged per AC with the help of the Arbitration Interframe Space Number (*AIFSN*[*AC*]). The *AIFSN*[*AC*] defines the duration of AIFS[AC] according to

$$AIFS\,[AC] = SIFS + AIFSN\,[AC] \cdot aSlotTime, \quad AIFSN\,[AC] \geq 2.$$

AIFSN[*AC*] should be selected by the HC such that the earliest access time of EDCA stations is DIFS, equivalent to legacy 802.11. The parameter *aSlotTime* defines the duration of a slot. The smaller the *AIFSN*[*AC*], the higher the medium access priority.

The minimum size of the contention window, *CWmin*[*AC*], is another parameter dependent on the AC. The initial value for the backoff counter is a random number taken from an interval defined by the CW, similar to legacy DCF. See Figure 5.18 for an illustration of the *AIFS*[*AC*] and *CWmin*[*AC*]. Four priorities are shown in the figure. The smaller the *CWmin*[*AC*], the higher

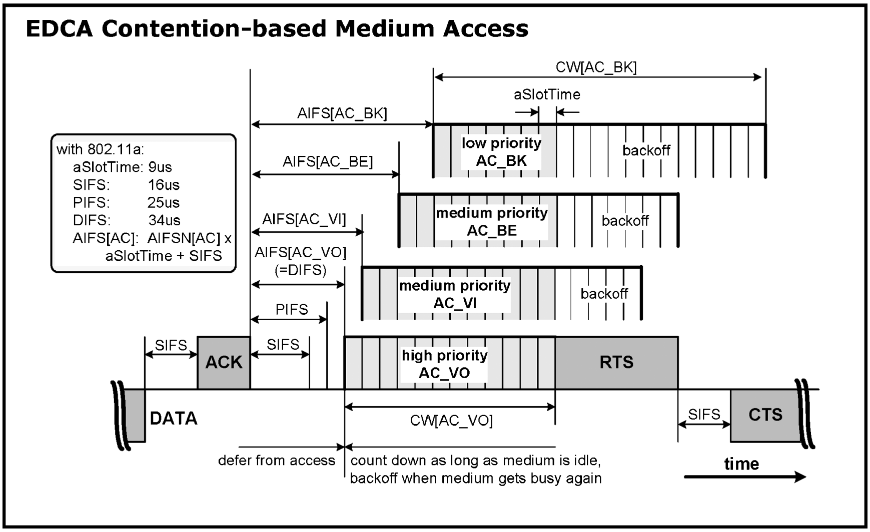


Figure 5.18 In EDCA, multiple backoff entities contend for medium access with different priorities in parallel. The earliest possible medium access time after a busy medium is DIFS.

the priority in medium access. A big difference between legacy DCF and 802.11e EDCA in terms of the backoff countdown rule is as follows: (i) the first backoff countdown occurs at the end of the *AIFSN*[*AC*] interval, and (ii) a frame transmission is initiated after a slot from the moment when the backoff counter becomes zero. However, the collision probability increases with smaller *CWmin*[*AC*] if there are more than one backoff entity of the respective AC operating in the QBSS. In the case of *AIFSN*[*AC*] being selected such that the earliest medium access time is DIFS, priority over legacy stations can be supported by setting *CWmin*[*AC*] <15 (for 802.11a).

The positions and sizes of the contention windows relative to each other, as defined in Figure 5.16(b) per AC by the EDCA parameter set, are important factors to define relative priority in medium access per AC.

The contention window increases upon unsuccessful frame exchanges: The size of the contention window $CW_i[AC]$ in backoff stage i is defined as

$$CW_i\,[AC] = min\left[2^i\,(CWmin\,[AC] + 1) - 1,\, CWmax\,[AC]\right]$$

The contention window never exceeds the value of *CWmax*[*AC*]. This parameter is defined per AC as part of the EDCA parameter set. The smaller the *CWmax*[*AC*], the higher the medium access priority. However, a small *CWmax*[*AC*] may increase the collision probability. Further, it should be highlighted that there are retry counters (similar to legacy 802.11) that limit the number of retransmissions. The 802.11e protocol also defines a maximum MSDU lifetime per AC, which specifies the maximum time a frame may remain in the MAC. Once the maximum lifetime has passed since a frame arrived at the MAC, the frame is dropped without being transmitted. This feature can be useful since transmitting a frame too late is not meaningful to many real-time applications.

In addition to the backoff parameters, the *TXOPlimit*[*AC*] is defined per AC as part of the EDCA parameter set. The larger *TXOPlimit*[*AC*] is, the larger the share of capacity for this AC. Once a TXOP is obtained using a backoff, a backoff entity may continue to deliver more than one MSDU consecutively during the same TXOP, which may take up to the duration of *TXOPlimit*[*AC*]. This important concept in 802.11e is referred to as continuation of an EDCF-TXOP.

Further, every 802.11e station has a short retry counter and a long retry counter for each AC with initial values of zero, denoted as *QSRC*[*AC*] and *QLRC*[*AC*]. These two retry counters are used to differentiate between MAC frames of different lengths and default values for these counters are not given in the standard (IEEE, 2005b). The retry counters determine how often a frame is retransmitted after a collision until it is discarded. In the case of AC_VO and AC_VI small retry counters are required while AC_BE and AC_BK are less sensitive to many retransmissions. As neither default values nor a capability for centralized management are defined in the standard, the prioritization through retry counters is difficult. Retry counters typically have a predefined default value of 7.

As described above, four backoff entities with different EDCA parameter sets may be active inside an 802.11e station. During contention, when the counters of two or more backoff entities in the same station reach zero at the same time, a virtual collision occurs. Upon access to the same slot by more than one backoff entity of one station, the backoff entity with the higher priority will transmit, whereas all other backoff entities will act as if a collision occurred on the medium. It may still occur that the transmission of the backoff entity with the higher probability collides with another transmission initiated by other stations.

5.5.7 Evaluation of Contention-based Medium Access

We use event-driven stochastic simulation to evaluate the performance of the 802.11e MAC for the 802.11a Physical layer (PHY) at 5 GHz that allows up to 54 Mbit/s. For the delay results of the MSDU delivery, we give empirical complementary Cumulative Distribution Functions (CDFs) of the resulting stochastic data.

Transmission powers and distances between stations are chosen such that stations are not hidden to each other with the selected PHY modes. If not stated otherwise, control frames are transmitted at 6 Mbit/s and data frames are transmitted at 24 Mbit/s. Each station generates the same mixture of offered traffic of four data streams, which we label according to the known ACs. A simple traffic model is used to make sure that the characteristics of the traffic sources do not influence the results, which would lead us to misinterpretations (for example, a correlated packet arrival of realistic voice traffic would have impact on throughput results). If not stated otherwise, at all backoff entities, MSDUs of *512 byte* with negative-exponentially distributed inter-arrival times are generated. Each stream carries 250 kbit/s.

We use the EDCA parameters as provided in Figure 5.19. Neither RTS/CTS nor fragmentation is used. The duration of EDCA-TXOPs allows stations to transmit one data frame after winning the contention in EDCA. Beacon frames are not generated.

Values for the EDCA Parameter Sets as Used in Simulations

	AC_VO	AC_VI	AC_BE	AC_BK	High (AC H)	Med. (AC M)	Low (AC L)
AIFSN:	*2*	*2*	*3*	*7*	*2*	*4*	*7*
CWmin:	*3*	*7*	*15*	*15*	*7*	*10*	*15*
CWmax:	*7*	*15*	*1023*	*1023*	*7*	*31*	*255*
	(used for throughput evaluation)				(used for delay evaluation of OQBSS)		

Figure 5.19 Simulation set-up for the evaluation of contention-base medium access.

5.5.7.1 Related Work

An analytic Markov model of the backoff process from the approximation analysis of the legacy 802.11 DCF by Bianchi (1998a; 1998b; 2000) is the basis for many recent publications. Bianchi uses a two-dimensional Markov chain to calculate the saturation throughput of contending legacy backoff entities. A simplification of this well-known model is done in Bianchi and Tinnirello (2005). A Z-transform based analysis of the service time in saturated 802.11 networks is performed in Zanella and De Pellegrini (2005) ending up in a linear evaluation of the service time distribution.

The EDCA saturation throughput of competing backoff entities of different ACs is analyzed and discussed in Mangold (2003). The throughput capacity per AC and the mutual influences of the ACs on each other are evaluated in terms of share of capacity per AC. This analytic model is extended in Berlemann (2006) for a saturation analysis of the mean service time and the service time distribution. Xiao (2004) also developed a model to analyze the prioritization through contention window size differentiation of the EDCA. The throughput and mean delay is evaluated by neglecting the different AIFS per AC and the virtual collision mechanism as originally specified in 802.11e. Robinson and Randhawa (2004) extended Bianchi's model to analyze the saturation throughput performance of the EDCA mechanism ending up in a complex, difficult to analyze model. A compact throughput and mean delay analysis of the EDCA in saturation under consideration of virtual collisions is given in Tantra *et al.* (2005). A similar approach, but with a more complex model, is provided in Tsai and Wu (2005), where different multidimensional Markov models per ACs are applied.

5.5.7.2 EDCA throughput Capacity in an Isolated QBSS with Four Stations

The throughput capacity (the saturation throughput) depends on a large number of parameters. The results given here are valid only for the MAC and PHY settings we use in our simulation. In the simulation discussed in this section and in the following section, all frames are transmitted at 24 Mbit/s.

Figure 5.20 is derived from a simulation with a single station (the AP) transmitting four streams to each of three stations, one stream per priority. Hence, the AP transmits 12 streams in total. The default EDCF parameters of the four ACs (AC_VI…AC_BK) are used (see Figure 5.18). One data frame per EDCF-TXOP is transmitted. The measured throughput per AC and the scenario is illustrated in Figure 5.20.

It is known that the throughput capacity in 802.11 depends on the size of the delivered MSDUs, and the PHY modes used. The larger the MSDU size, the higher the throughput capacity, as long as the number of collisions and the amount of interference is small. The higher the PHY mode, the higher the throughput capacity, as long as the channel conditions are sufficient. In Figure 5.20, we observe that with increasing offered traffic, ACs with higher priority restrain the throughput capacity of the ACs with lower priority, thanks to smaller AIFSN, CWmin and CWmax values.

5.5.7.3 EDCA throughput with Increasing Number of Stations

In this section we evaluate the QoS support with EDCA medium access in an isolated QBSS as shown in Figure 5.21. All stations rely again on prioritized medium access over the EDCA only, i.e., no polling-based access is utilized. Each station transmits using all four backoff entities ACs (250 kbit/s each); hence each station attempts to carry 1 Mbit/s in total.

To investigate the performance of 802.11e in hot spots, where it is likely that a large number of stations are associated with an AP, we increase the number of contending stations up to 16. Each

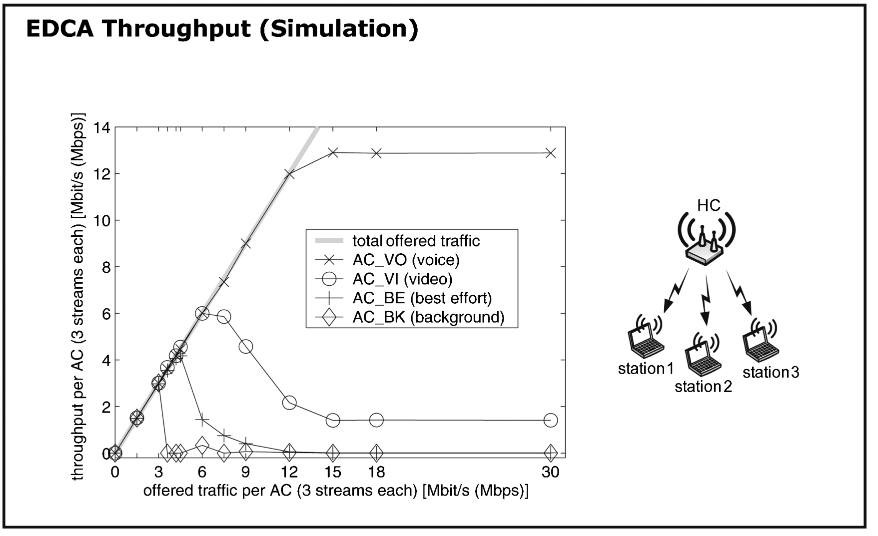


Figure 5.20 EDCA throughput capacity in an isolated QBSS with four stations. Throughput per AC with increasing offered traffic per AC, for the illustrated scenario. Reproduced by permission of © 2003 IEEE[1].

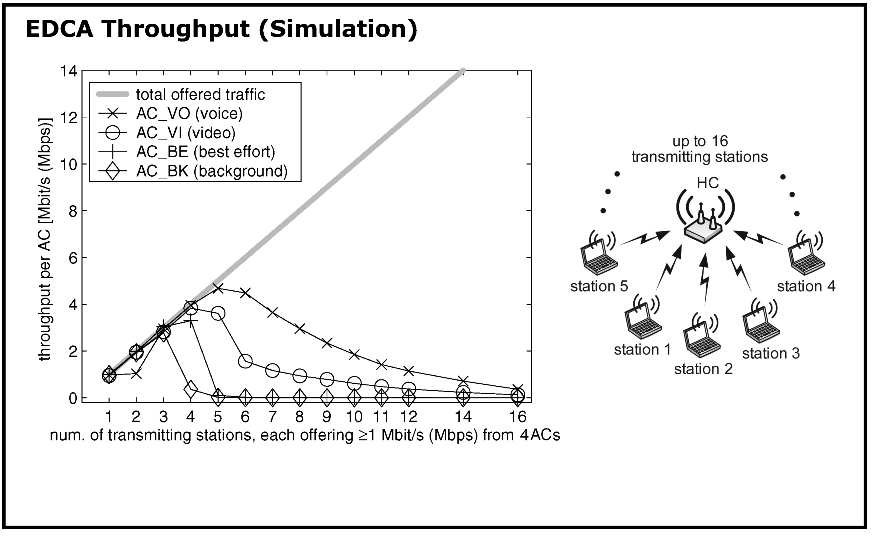


Figure 5.21 Throughput per AC with increasing number of stations, and constant offered traffic per station, for the illustrated scenario. Reproduced by permission of © 2003 IEEE[1].

station offers the same traffic. It can be observed from Figure 5.21 that with the default values for the EDCA parameter set the throughput reduces dramatically, once the number of stations becomes large. This is a result of the increased collision probability particularly in AC_VO, where *CWmin* and *CWmax* are small numbers. This is clearly not a desirable result. However, in this case, the HC should set new values for the EDCA parameters, which would then be used by all stations of the QBSS. It should be noted that the highest priority AC of EDCA is expected to be used for voice-like calls in real-life scenarios, which means less offered traffic per station at AC_VO (assuming one call per station), as is applied in this simulation.

5.5.8 Controlled Medium Access

The controlled medium access of the HCF, referred to as HCF Controlled Channel Access (HCCA) extends the EDCA access rules by allowing the highest priority medium access to the HC during both, the contention-free period and the contention period (CFP and CP). The details about the controlled medium access are summarized in this section.

A TXOP can be obtained by the HC via the controlled medium access. The HC may allocate TXOPs to itself to initiate MSDU deliveries whenever required, after detecting the medium idle for PIFS duration, and without backoff. To give the HC higher priority over legacy DCF and EDCA access, *AIFSN*[*AC*] must be selected such that the earliest medium access for EDCA stations is DIFS for any AC.

During CP, each TXOP of an 802.11e station begins either when the medium is determined to be available under the EDCA rules, i.e., after *AIFS*[*AC*] plus the random backoff time, or when a backoff entity receives a poll frame, the *QoS CF-Poll*, from the HC. The *QoS CF-Poll* from the HC can be transmitted after a PIFS idle period, without any backoff, by the HC. During CFP, the starting time and maximum duration of each TXOP is also specified by the HC, again using the *QoS CF-Poll* frame. During CFP, 802.11e backoff entities will not attempt to access the medium without being explicitly polled, hence, only the HC can allocate TXOPs by transmitting *QoS CF-Poll* frames, or by immediately transmitting downlink data. During a polled TXOP, a polled station can transmit multiple frames, which the station selects to transmit according to its scheduling algorithm, with an SIFS time gap between two consecutive frames as long as the entire frame exchange duration does not exceed the allocated maximum *TXOPlimit*.

Polled TXOP allocations may be delayed by the duration of an EDCA-TXOP, as illustrated in Figure 5.22. The HC controls the maximum duration of EDCA-TXOPs within its QBSS by announcing the *TXOPlimit*[*AC*] for every AC via the beacon. Therefore, it is able to allocate polled TXOPs at any time during the CP, and the optional CFP. When very small MSDU delivery delays are required, *QoS CF-Polls* may be transmitted a duration of *TXOPlimit*[*AC*] earlier than the optimal polled TXOP allocation time to avoid any MSDU delivery delay imposed by EDCA-TXOPs at all. However, the largest *TXOPlimit*[*AC*] of the four ACs must be considered.

5.5.8.1 QoS Guarantee with HCCA vs. EDCA

We now evaluate the effect of allowing the highest priority to the HC medium access, according to HCCA. The values for the EDCA parameter set are selected according to AC_H (high priority), AC_M (medium priority), AC_L (low priority), as given in Figure 5.19. Note that the HCCA will show the same tendencies if the values for the EDCA parameter set of the previous sections had been selected for evaluation. At the high priority AC, MSDUs of 80 bytes arrive at constant periods. The period length depends on the offered traffic, and is 5 ms for

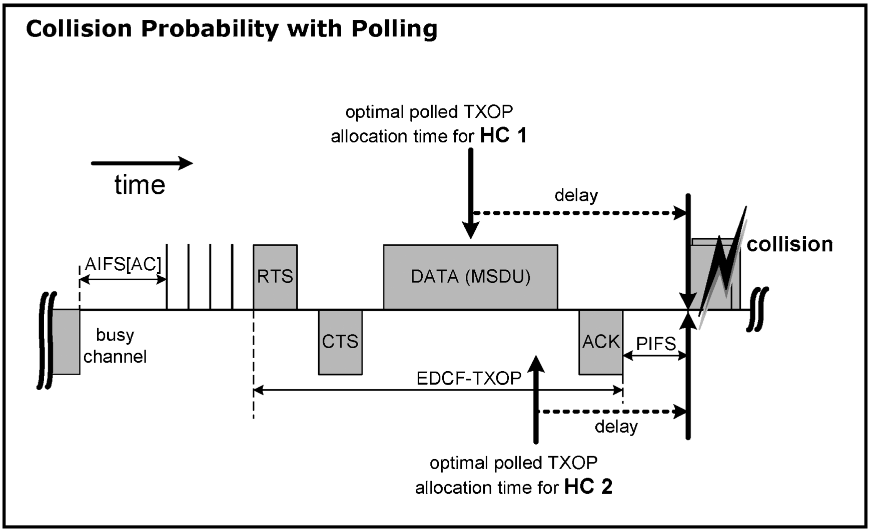


Figure 5.22 A polled TXOP allocation. Any 802.11e frame exchange will not take longer than the TXOPlimit, which is the limit for all EDCA-TXOPs and under control of the HC.

the offered traffic of 128 kbit/s. Medium and low priority ACs generate MSDUs of 200 bytes with negative-exponentially distributed interarrival times. Each stream carries 160 kbit/s. Data frames are transmitted at 24 Mbit/s, control frames (ACKs) at 6 Mbit/s. In contrast to the previous scenarios, where all stations including the AP contend for medium access via EDCA, now the AP carries an additional isochronous downstream (80 bytes per MSDU, 128 kbit/s) which is delivered with the HCCA priority, which has higher priority than AC_H. That is, the data frames of this stream are immediately transmitted after PIFS when the medium is detected as idle. Note that the HCCA achieves its strict delay requirements in the CP by setting a maximum TXOP duration for all other streams.

Figure 5.23 shows the resulting MSDU delivery delay distributions for an isolated QBSS, and for overlapping QBSSs. The overlapping QBSS scenario is illustrated in the figure. The resulting delays for the HCCA stream and for AC_M are also illustrated. It can be seen that whereas the delays of the EDCA increase unpredictably with increased offered traffic, the HCCA delays remain below a certain threshold, which is defined by the TXOPlimit. Only the HCCA stream stays within its maximum delay limit.

The MSDU delivery delay even for the HCCA stream is significantly increased, if two (or more) QBSSs are overlapping each other, so that they mutually interfere. In this scenario, even polled data frames of highest priority suffer from an unpredictable delay and throughput degradation due to uncoordinated resource sharing between HCs. One result of such a scenario is given in Figure 5.23. It can be seen that the delay of the high priority stream exceeds the TXOPlimit of 300 us defined by the HC for the HCCA stream under QBSS overlap. Note that the given results is a summary of a variety of conflicts and related delays observed in overlapping QBSSs. One example conflict is that two HCs periodically poll stations at similar times. A solution for MAC frame based access protocols like 802.15, 802.16 and HiperLAN2 that also applies to superframe based protocols like 802.11 to support coexistence of infrastructure BSSs operating on the same radio channel is known from (Kraemling 2000). A distributed control based algorithm

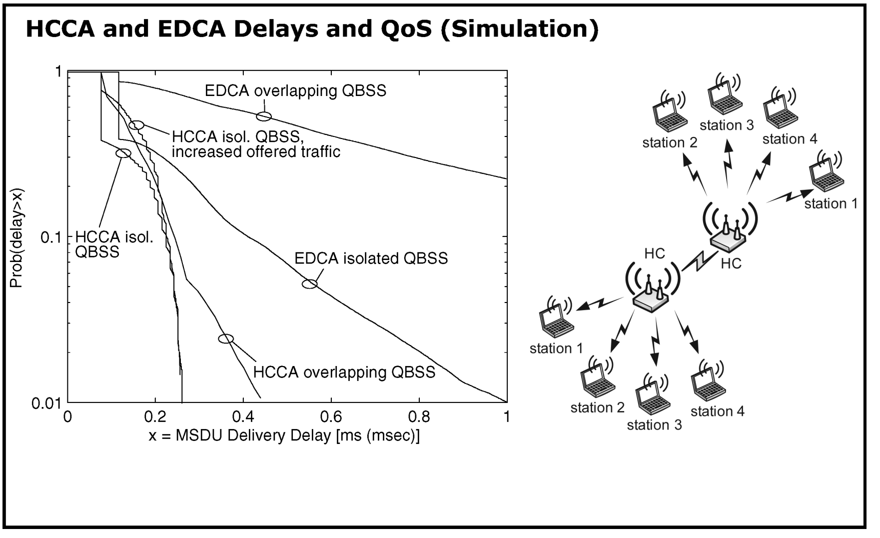


Figure 5.23 MSDU delivery delay for AC_M (EDCA) and HCCA, in isolated and overlapping QBSSs. The increase of the HCCA delay in overlapping QBSSs is not under control of a HC, and therefore undesirable. Reproduced by permission of © 2003 IEEE[1].

is introduced there to enable coexistence and its performance at the example of HiperLAN/2 is analyzed. Then all poll frames collide, and the HCCA throughput drops down to zero. For the overlapping QBSS problem, various solution concepts are proposed. One solution would be to apply dynamic frequency selection, to let a QBSS dynamically select a free medium. Other approaches are based on spectrum policies.

5.5.8.2 The Superframe

Figure 5.24 illustrates an example of a superframe that includes a CFP and a CP. The superframe starts with a beacon transmitted by the HC (indicated with (1)). During the CFP, i.e., the first part of the superframe, the backoff entities only transmit upon being polled by the HC. Indicated with (2) is the transmission of a fragmented MSDU within the CFP. The CFP ends with the *CF-End* frame transmitted by the HC as shown at (3). During the following CP, all backoff entities attempt to transmit through the contention-based medium access of the HCF, i.e., the EDCA. EDCA-TXOPs each containing an RTS/CTS handshake are shown (4). During the CP, the HC may poll a station, which is different from the PCF of the legacy 802.11. This is shown as an example where following the two EDCA-TXOPs, the HC polls a station to allocate a polled TXOP during which a fragmented MSDU is transmitted, as indicated with (5).

5.5.9 Block Acknowledgment

With the optional block acknowledgment, the throughput efficiency of the protocol is improved. Block acknowledgments allow a backoff entity to deliver a number of MSDUs being delivered consecutively during one TXOP and transmitted without individual ACK frames. The MPDUs

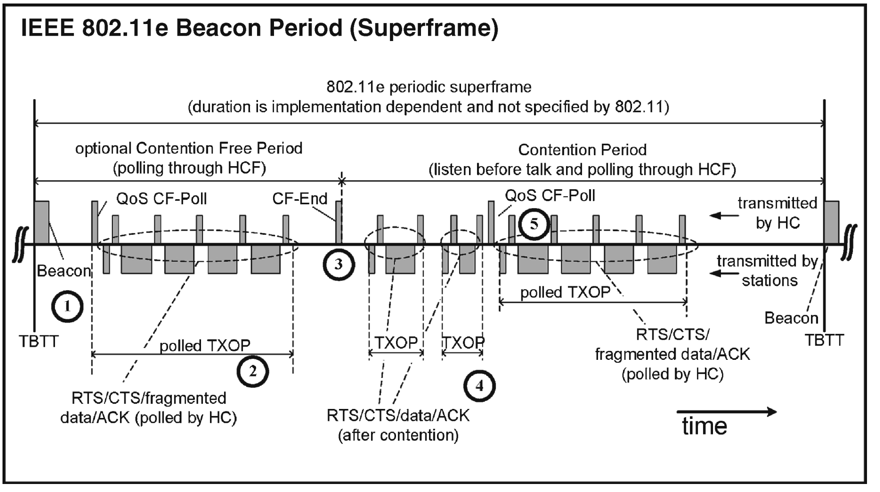


Figure 5.24 Example of an 802.11e superframe where the HC grants TXOPs in Contention-Free Period and Contention Period. The duration of the superframe is not specified in the standard.

that are transmitted during the TXOP are referred to as a block of MPDUs. At the end of the block, or in a later TXOP, all MPDUs are acknowledged by a bit pattern transmitted in the block acknowledgment frame, thus reducing the overhead of control exchange sequences to a minimum of one acknowledgment frame per number of MPDUs delivered in a block (Figure 5.25).

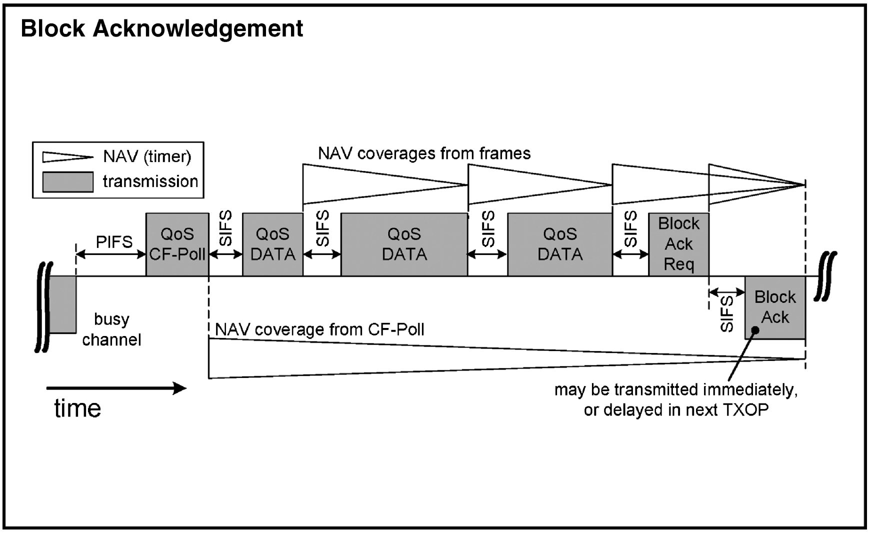


Figure 5.25 Frame exchanges with Block Acknowledgment.

5.5.10 Direct Link Protocol (DLP)

Any backoff entity can directly communicate with any other backoff entity in a QBSS, without communicating via the AP. In the legacy 802.11 protocol, within an infrastructure-based BSS (which is denoted as BSS), all data frames are sent to the AP, and received from the AP. This, however, consumes at least twice the channel capacity compared to the direct communication. Only in an independent BSS (which is denoted as IBSS), station-to-station communication is allowed in the legacy protocol, due to the absence of the AP. The direct communication in 802.11e is referred to as Direct Link (DiL). A set-up procedure, the Direct Link Protocol (DLP), is defined to establish a DiL between 802.11e backoff entities.

5.6 Radio Spectrum Management

5.6.1 Measurements in 802.11

Control of link adaptation, power control, and error correction schemes is highly dependent on detailed knowledge of up-to-date signal and interference conditions in the field. Accordingly, all cellular systems specify the exchange of related information.

While a lot of measurement procedures are defined in UMTS and GSM, legacy 802.11 does not provide any of them. Recognizing the need for specific link information, a standard amendment for spectrum and transmit power management extensions, known as 802.11h (IEEE, 2003b), has been specified and was approved in 2003. Focus was put on two services: Dynamic Frequency Selection (DFS) and Transmit Power Control (TPC). While 802.11h mainly addresses internal use of data, a further amendment is currently under preparation by IEEE 802.11 Task Group k, which will be referred to as 802.11k (IEEE, 2005k). While 802.11h targets on *spectrum management*, 802.11k specifies further measuring options, which are referred to as *radio resource measurements*. A particular aim thereby is to allow for locally taken measurements on the one hand as well as exchange of measurement information with peer stations including third parties and external sources on the other hand. In this way, new services such as location-based services can be encouraged.

Both amendments, 802.11h and k, introduce basic structures for requesting and reporting measurement information of 802 type only. No inter-RAT measurement procedures are defined and there is no standardized way of performing handover from WLAN to other cellular mobile radio networks. Interoperability between heterogeneous networks, however, is considered by the Media Independent Handoff Working Group IEEE 802.21 whose aim is to develop standards to enable handover and interoperability between heterogeneous network types including both 802 and non-802 networks.

5.6.1.1 Information Transfer

Stations may provide measurement information autonomously and/or on a regular basis. However, common practice is to generate respective reports as a response to previously received measurement requests. Legacy IEEE 802.11 defines three different MAC frame types: *management* frames, *control* frames and *data* frames. Data frames are reserved for transporting layer 3 data, control frames handle the access to the medium and management frames serve for the administration of the BSS and terminals. All spectrum management actions (802.11h) and radio resource management actions (802.11k) therefore use the management format.

The general MAC frame format comprises a set of fields that occur in a fixed order in all frame types, see Figure 5.4. The first two octets serve as *Frame Control* to identify the type of frame to be transmitted. A subsequent two-octet *Duration/ID* field informs about the duration of each frame. Depending on the frame format, one to four *Address* fields are specified to announce sender and receiver of a particular message. The address field usage thereby is specified by the relative position of the address field within the MAC header. Further on, a *Sequence Control* field is present in some frame types, which is needed for de-/fragmentation and serialization of messages. A *Frame Body* field of variable length and a 32-bit Frame Check Sequence (FCS) field with a 32-bit Cyclic Redundancy Code (CRC) complete the general MAC frame format.

The initial *Frame Control* field provides a Type description field (data, control, or management) and a Subtype field for more specific information. Depending on the chosen (sub-) type, the variable *Frame Body* field is assorted. For management frames of legacy 802.11, the subtype defines Association and Authentication fields, Probe directives, Beacons and Traffic Indication Announcements. To transmit spectrum management (802.11h) or radio measurement (802.11k) directives, a special management frame of subtype *Action* is used, which was newly defined within 802.11h.

The Action field entry type *Category* in a management frame specifies whether spectrum management or radio measurement tasks are to be fulfilled. Depending on the category, the corresponding Action field defines several actions to be executed.

The Frame Body Format is the same for 802.11h and k. The Measurement *Request* Frame Body Format, however, has been slightly enhanced by 802.11k. Two additional fields, *Number of Repetitions* and *Frame Restart Delay*, have been added. The first allows requesting for repeated execution of the indicated measurement while the second specifies the intermediate time span to be waited.

Central elements of the Measurement Request frames and the Measurement Report frames are the Measurement Request Element and the Measurement Report Element, respectively.

Both include a *Measurement Request/Report Mode* field and a *Measurement Type* field. The *Measurement Request Mode* Field informs about the disposition of a station to accept autonomously sent requests or reports from other stations. An additional field as used by 802.11k allows specifying whether the measurement duration included in any request shall be considered as mandatory or aspired. The associated *Measurement Report Mode* Field gives feedback whether a respective measurement request was received in time or can (not) be replied.

The *Measurement Request/Report Type* fields are the actual information carriers. Depending on the type, specific measurement actions are performed by the receiving station. A detailed description of each of these Measurement Types is given in the following sections.

5.6.1.2 Specific Measurements in 802.11h

Spectrum management as introduced by 802.11h defines a couple of new *Information Elements* (= optional components or variable length fields transmitted in the frame body of *Management* frames).

The newly introduced information elements allow for exchange of capability information and commentatorship on specific link conditions. Important properties of the channel may be derived by applying these IEs: Reception strength of the *TPC report* that entails the transmit power with which it was conveyed allows for direct derivation of the pathloss. Application of the *Quiet* element may be used to assist in making channel measurements without interference from other STAs in the (I)BSS.

IEEE 802.11h Measurements Reports (MR) are generated autonomously or on request. The request may be sent to an individual (AP → STA or STA → AP) or group destination address (AP → STAs; STA → STAs for IBSS only), whereby the latter is recommended to be used with care to avoid reply storms (IEEE, 2003b). Except for IBSS, STAs are not allowed to request other STAs for measurement reports. Basically, all measurement requests and reports are enabled by default but an entity may announce that it will not accept any requests or autonomously sent reports, cp. *Measurement Request Mode* field, an STA shall honor all other requests while an AP may ignore a request to disable a mandatory measurement request.

IEEE 802.11h measurement requests/reports always address one out of three types to specify further actions. The three types defined are Basic request/report, Clear Channel Assessment (CCA) request/report and Receive Power Indication (RPI) histogram request/report. While support of Basic Report generation is mandatory for STAs, support for CCA reports and RPI histogram reports is optional.

Parameters included in the request field for either of these three measurement request types define the *Channel Number* which specifies the channel to be scanned, a *Measurement Start Time* and *Measurement Duration*. The latter has a length of 2 bytes reflecting a number of Time Units (TUs) that correspond to the designated measuring period $T_{Measure}$. Accordingly, monitoring is performed for a time span of $1 \leq T_{Measure} \leq 2^{16} - 1$ time units corresponding to $1024\,\mu s \leq T_{Measure} \leq 67.12\,s$. As a rule of thumb, one can keep in mind that the upper bound for measurement durations in 802.11h and k comprises approximately one minute.

5.6.1.3 Basic Report

On reception of a measurement request of type *Basic Request*, an STA is obliged to perform required measurements and to reply with a *Basic Report*. The initial three fields echo the channel number, the measurement start time and duration as included in a preceding request (if there was one). An additional field, the *Map* field, is added to convey measuring results.

5.6.1.4 Clear Channel Assessment (CCA) Report

On reception of a measurement request of type *CCA Request*, an STA optionally generates a corresponding CCA Report as reply. The initial three fields echo the channel number, the measurement start time and duration as included in a preceding request (if there was one). An additional field, the *CCA Busy Fraction* field, is added to convey measuring results.

To compile a CCA report the terminal surveys the channel status applying a resolution on the time scale of microseconds. The fractional duration over which the CCA indicated the channel was busy during the measurement duration is calculated hereafter by

$$CCA\_Busy\_Fraction = Ceiling\left[255^{*}\frac{Duration\ CCA\ indicated\ channel\ was\ busy(\mu s)}{1024^{*}Measurement\ duration(TUs)}\right].$$

5.6.1.5 Receive Power Indication (RPI) Histogram Report

On reception of a measurement request of type *RPI Histogram Request*, an STA optionally generates a corresponding RPI Histogram Report as reply. The initial three fields echo the channel number, the measurement start time and duration as included in a preceding request (if there was one). The actual outcome of the measurement is conveyed as an RPI histogram. Each of the RPI levels thereby is coded with an eight-bit field.

Determination of the fractional part of each RPI level is similar to the approach for CCA Busy Fraction. For the whole measurement duration the received power level on the specified channel is recorded and a classification in one of eight RPI levels is performed.

In the end, the fraction of time with which each RPI level was received is calculated by the following equation and the result is written to the respective density field.

$$RPI\_X\_density = Ceiling\left[255^{*}\frac{Duration\ receiving\ at\ RPI\ value(\mu s)}{1024^{*}Measurement\ duration(TUs)}\right].$$

The sum of the densities at the receiving end will be approximately 255, but can end up to 262 due to rounding effects. The density resolution of each single field is limited by a number of eight bits used for transmission. Hence, data integrity is bounded to 1/256, which means a maximum deviation of 0.4 % has to be accepted.

5.6.2 Specific Measurements in 802.11K

In addition to *spectrum management* support by 802.11h, additional procedures are currently specified by 802.11k to support *radio measurements*. The main target thereby is to provide means for measurement information exchange between different communication partners and third parties. Thus, sophisticated controlling mechanisms and new services will be supported or enabled.

Radio resource measurement as introduced by 802.11k defines some further new *Information Elements* (= optional components or variable length fields transmitted in the frame body of *Management* frames, see Section 5.6.1.2).

Some IEs such as AP Channel Report, Neighbor Report and RCPI have been newly introduced in 802.11k. Their main objective is to provide support in fast and reliable neighbor BSS transition (handover). Being periodically transmitted together with the beacon, the *AP Channel Report* allows for easy recognition of other APs. More detailed information may be inquired by an STA using a *Neighbor Report Request*. Specific elements included in the response give information whether a new candidate transmission AP supports pre-authentication for fast roaming as specified in 802.11i (IEEE, 2004b) and indicate whether the new AP supports the same security level as the current one. Further selected capability information depletes the necessity for extra Probe Requests with capability enquiry to each of the reported APs. Once a handover to another AP is triggered, the procedure can be optimized by taking advantage of the Time Synchronization Function (TSF) parameters included in the Neighbor Report. Since the Neighbor Report can also be conveyed in a piggy-backed way with an *Association Response*, all important information on neighbor APs are delivered straight away as soon as an STA associates with an AP. In such a manner, fast proceeding handover as required when traversing adjacent hot spots may be facilitated.

However, similarly to 802.11h, 802.11k stresses one particular information element, which is the Measurement Request/Report (MR). In addition to the three 802.11h measurement types (Basic, CCA, RPI Request/Report), nine further measurement request types and eight respective responses have been specified: *Channel Load Request/Report, Noise Histogram Request/Report, Beacon Request/Report, Frame Request/Report, Hidden Node Request/Report, Medium Sensing Time Histogram Request/Report, STA Statistics Request/Report, LCI Request/Report* and *Measurement Pause Request*.

Similar to 802.11h, most of the different measurement type requests in 802.11k apply a generic request field. The fields *Channel Number* and *Measurement Duration* have the same meaning as explained before. A new field, the *Regulatory Class* allows request/report measurements for different frequency bands. The *Measurement Start Time* field that allows 802.11h to schedule measurement procedures at a dedicated point of time in the future ($\pm 32\,\mu s$) is replaced by a *Randomization Interval* field in 802.11k. Prior to making any measurements, an STA will calculate a random delay distributed uniformly in the range of 0 and the randomization interval. The maximum possible value corresponds to $D_{max} = 2^{16} - 1\,\text{TUs} \approx 67\,s$. The new approach for directing measurements by a randomization interval rather than a given time point is beneficial due to several reasons: First, traffic storms that could arise from synchronized broadcast and multicast measurements can be avoided. Second, reliable measurements can only be derived if based on statistical independence among probes. Third, it is intuitively senseless to provide means for requesting measures that are to be taken too far in the future (the address room of 64 bits for the Measurement Start Time in 802.11h theoretically allows for scheduling of measurement procedures up to $T_{MST} = (2^{64} - 1)\mu s > 500$ years (!) in advance. Thus, replacement of the eight-octet *Measurement Start Time* field by a two-octet *Randomization Interval* field improves straightforwardness and reduces the overhead.

Anyway, while the Measurement Start Time field in 802.11k was simplified compared to 802.11h, the corresponding *response* frames maintain a similar eight-octet date field to allow for absolute indication of the *Actual Measurement Start Time* at which the reported measurements were taken.

One further innovation of 802.11k is the definition of a *Duration Mandatory* bit field. If set to 1, the specified *Measurement Duration* will be interpreted as mandatory otherwise it will be interpreted as target duration.

The following subsections introduce all new measurement types. Most of them apply the generic request format, hence, only the corresponding responses are presented. If necessary, additional information on requests formats different to the generic one are described. All requests are transmitted in the *Measurement Request* field. Corresponding responses are transmitted in the respective *Measurement Report* fields.

5.6.2.1 Channel Load Report

A *Channel Load Report* is generated as a reply to a previous *Channel Load Request*. The first four fields, *Channel Number*, *Regulatory Class*, *Actual Measurement Start Time* and *Measurement Duration* include similar information as explained for the request. In fact, if an STA was able to fulfill all requirements as specified, these fields will entail a copy of the preceding request. Similarly, possible differences from requested parameters can be reported.

The actual payload of interest is the *Channel Load* field. It contains the proportion of measurement duration for which the measuring STA determined the channel to be busy. Following the definition for 802.11h's CCA Report, the channel load value is defined as:

$$Channel\_Load = Ceiling\left[255^{*}\frac{Channel\ busy\ time(\mu s)}{1024^{*}Measurement\ duration(TUs)}\right]$$

The *Channel busy time* in the equation is the time during which either the physical carrier sense or the Network Allocation Vector (NAV) indicated channel busy.

The difference between the *CCA Report* in 802.11h, see Section 5.6.4, and the *Channel Load Report* in 802.11k is that the latter comprises both *physical* carrier sense mechanisms (Clear

Channel Assessment) as well as *virtual* carrier sense mechanisms (NAV). In other words, the CCA Report is inherently included in the Channel Load Report.

The need for the definition of the Channel Load Report is easier to understand with a little example: A scheduling algorithm that solely exploits 802.11h CCA Reports could easily degrade the overall performance within a scenario. Assuming hidden stations, a CCA Report would report an idle medium. However, a Channel Load Report further considers NAVs as set by the MAC due to RTS/CTS messages. The scheduling algorithms thus would backoff the own medium access such that other transmissions are not disturbed.

5.6.2.2 Noise Histogram Report

On accepting a *Noise Histogram Request*, an STA will reply with a *Noise Histogram Report*. The actual payload is reported as *Noise Histogram* reflected by the known eight density fields. Compilation of the density histogram resembles the procedure of the RPI Histogram as defined in 802.11h, see Section 5.6.1.5. Similarly, the same classification as before is applied for the quantization into eight density levels. However, while for the RPI Histogram *any* received power during the *whole* measurement duration is recorded, the Noise Histogram is based on measurements taken exclusively when NAV is equal to 0. Thus, measurements are only taken when the virtual carrier sense mechanism indicated idle channel. This is how the Noise Histogram Report includes only non-802.11 energy in its result.

5.6.2.3 Beacon Report

On accepting a *Beacon Request*, an STA will respond with a *Beacon Report* for each requested BSSID. The format of the Beacon Request is slightly enhanced compared to the generic measurement frame format. While the first four fields are the same, five additional fields have been specified: *Measurement Mode*, *BSSID*, *Reporting Condition*, *Threshold/Offset* and *Hysteresis*.

The aim of a Beacon Request/Report is to gather information on other BSSs in the reception range of a station. The *Measurement Mode* field in the request indicates the mode to be used for the measurements:

- Passive Mode: An STA is ordered to compile a report based on all Beacon or Probe Response management frames with the requested BSSID being received within a given time span.
- Active Mode: The STA is ordered to transmit a Probe Request to the broadcast destination address. The rest of the procedure resembles the passive mode above. Similarities to the active scanning procedure as described in Section 5.4.3.2 are obvious. The difference is that Beacon Request/Reports evaluate both Beacon and Probe Responses, while active scanning is restricted to Probe Responses.
- Beacon Table: The STA is advised to return a Beacon Report containing the current contents of any stored beacon information for any channel with the requested BSSID. In particular, no extra measurements will be performed.

The last three fields, *Reporting Condition*, *Threshold/Offset* and *Hysteresis*, are used if conditional reporting is to be supported. In this case, no time-bounded measurement durations determine the time for feedback signaling. Instead, upper and lower limits for RSSI and/or RCPI are defined. Exceeding or falling below these thresholds initiates the report.

For the Beacon Report, besides well-known fields as included in all other reports further information fields are included summarizing results of measurements: The *PHY Type* field indicates the physical medium type of the Beacon/Probe Response frame being reported and *RCPI* indicates the channel power with which it was received. *Parent TSF* contains the lower four octets of the measuring STA's TSF timer value at the time the Beacon/Response frame being reported was received. *Target TSF*, in turn, contains the timestamp field from the reported Beacon/Probe Response frame. Further information, reported within the received Beacon/Probe Response frame, is mapped to the *Beacon Interval* and *Capability Information* fields. Any other elements that were received may be reported by the *Received Elements* field.

5.6.2.4 Frame Report

A *Frame Request* frame conveyed with the generic format yields to the compilation of a *Frame Report* frame. The central element here is the *Frame Report Entry* with its four subfields *Transmit Address*, *BSSID*, *RCPI* and *Number of Frames*. A Frame Report Entry thus is a summary of the traffic from one specific transmit address, whereby the RCPI indicates the received channel power in dBm. Its value corresponds either to the most recently received frame, respectively is calculated as average of the values of the individual frames received. The *Number of Frames* field is a count of the individual frames transmitted by one specific station and gives information about the activity of this station within the measurement duration.

5.6.2.5 Hidden Station Report

Similar to most other radio measurement requests in 802.11k, the *Hidden Node Request* applies the generic measurement request format. If a station accepts a respective request, it will respond with a Measurement Report frame containing a variable number of *Hidden Station Entries*. Each Hidden Station Entry comprises one doublet reporting the MAC address of a hidden station and the number of associated detected frames.

The algorithm that detects a possible hidden station relies on reception of a particular frame transmitted from any Station A to any other Station B. If Station B is supposed to acknowledge the correct reception to Station A, but this acknowledgment cannot be detected by the measuring station, Station B is likely to be a hidden station with respect to the measuring station. Accordingly, the MAC address of Station B, which is known from the initial transmission from Station A → Station B, is recorded as *Hidden Station Address*. To minimize false alarms, the measuring station only exploits initial transmissions. Retransmissions are not evaluated.

5.6.2.6 Medium Sensing Time Histogram Report

Similar to the Beacon Request, the Medium Sensing Time Histogram Request (MSTH Request) specifies further request elements in addition to the generic request format.

The new elements defined are the *Medium Sensing Measurement Subtype*, *RPI Threshold*, *Bin Offset*, *Bin Duration* and *Number of Bins*. The *Medium Sensing Measurement Subtype* field is used to distinguish further subtypes.

The general purpose of all time histograms is to provide additional information with respect to channel usage rather than sole percentage values of busy or idle time. The CCA report (see Section 5.6.1.5) and the RPI histogram report (Section 5.6.1.4) as specified in 802.11h, for instance, provide feedback on the proportional appearance of busy time or RPI levels within the measurement period. A value of e.g. 50% busy time, however, entails different possibilities: e.g., a channel was completely busy during the first half of scanning and completely idle during the second half, or a channel is alternately busy and idle with intervals much smaller than the measurement duration. Obviously, information on the respective distribution is lost if only percentage values are provided.

Medium Sensing Time Histograms take this into account by providing information about specific durations, e.g. busy and idle times, represented as probability densities. Information transfer thereby is accomplished with the help of so called *Bins*. A Bin represents a certain amount of time. The minimum medium sensing interval, Bin 0, is specified by the *Bin Offset*. Medium sensing intervals smaller than Bin Offset are ignored and not reported. With increasing ordinal number, the time span represented by Bin 1, Bin 2, Bin 3,…, Bin $(N-1)$ increases by *Bin Duration*, see the following equation, whereby N denotes the overall *Number of Bins*.

$$\begin{aligned}
Bin\ 0 &= Bin\ Offset;\\
Bin\ 1 &= Bin\ 0 + Bin\ Duration;\\
Bin\ 2 &= Bin\ 1 + Bin\ Duration = Bin\ 0 + 2 \cdot Bin\ Duration\\
&\vdots\\
Bin\ N-1 &= Bin\ 0 + (N-1) \cdot BinDuration
\end{aligned}$$

On accepting an MSTH Request, a station is supposed to reply with an MSTH Report. A new field, the *Total Number of Medium Sensing Intervals* is included for the purpose of assessing the reported data.

Reliable information on the statistical distribution of channel allocation is a powerful means to support QoS aspects. The specification of 802.11k proposes to use MSTHs to estimate traffic load priorities or detect non-802.11 radio activities. Knowing the characteristics of another interference source may be exploited e.g. by choosing optimal frame lengths when operating in the same channel.

5.6.2.7 STA Statistics Report

An *STA Statistics Report* is passed back as reply to a preceding *STA Statistics Request*. Each AP/STA in 802.11 administers a Management Information Base (MIB) that comprises the managed objects, attributes, actions and notifications required to manage a station. Specific MAC counters administered by the MIB provide the necessary support for access control, generation and verification of frame check sequences, and proper delivery of valid data to upper layers.

The STA Statistics Report allows provision of information of dedicated so-called *Group Identities* to other stations. The current 802.11k draft (IEEE, 2005k) specifies only one Group Identity that basically subsumes all MAC counters. The STA Statistics Report thereby reports incrementally, listing the change in STA counters within the Measurement Duration. A Measurement Duration equal to 0 is used to report instantaneous values of indicated STA counters.

5.6.2.8 LCI Report

The Location Configuration Information (LCI) Report is used to convey information on specific positions of stations. Its format has been adopted from IETF RFC 3825 (Polk *et al.*, 2004) that specifies a DHCP option for coordinate-based location configuration information. The report includes an LCI element to provide information on latitude, longitude and altitude. Each measure is further characterized by a resolution field. The *Datum* field comprises an eight-bit value encoding the horizontal and vertical references used for the coordinates given in the LCI. A value of 1 relates to the World Geodetic System 1984 (WGS84).

The mechanism based on which location information to be included in the LCI report is collected is not specified by 802.11k to allow the accuracy of the reported location to be "best effort". A requesting Station A has two options when asking for a position: It may prompt another Station B to provide either the *local* position (= position of Station A) or it may prompt Station B to provide the *remote* position (= position of Station B).

In such a manner, a station may derive information on its own position, without any further capability to perform positioning. If the associated AP supports network-based foreign positioning, the requesting station simply needs to send a *local* LCI request to the AP.

5.6.2.9 Measurement Pause Request

The *Measurement Pause Request* is the latest measurement request element that has been specified by 802.11k so far. It is used to provide time delays between the execution times of measurement request elements in a Measurement Request Frame. Unlike other requests, there is no associated response defined. The only element conveyed is a two-octet *Pause Time* field. Bit 0 serves as switch for the time span defined by the remaining 15 bits. A scale factor of 1 or 1000 corresponding to a minimum/maximum measurement pause duration of $1\,\text{ms} \leq\approx T_{P1} \leq\approx 33{,}6\,\text{s}$ or $1\,\text{s} \leq\approx T_{P2} \leq\approx 9\,\text{h}19\,\text{m}$ is possible.

5.7 History and Selected Sub-standards, i.e., Amendments

5.7.1 IEEE 802.11

The 802.11 standard is considered as the root standard, defining operation and interfaces at MAC and PHY for data networks such as the popular TCP/IP. Three PHY layer interfaces are defined that are not compatible with each other. One is based on Infrared (IR) communications, and the other two use the 2.4 GHz unlicensed band, which is in general considered to be harmonized over all regulatory regions in the world. One is based on FHSS and the other uses DSSS.

802.11 was published in 1997, and an updated version has been available since 1999. In 2003, the first substandards 802.11a and 802.11b were merged into the 1999 document (IEEE, 2003a).

5.7.2 IEEE 802.11a

This extension defines the PHY that allows up to 54 Mb/s by operating in the 5 GHz unlicensed band, making use of the OFDM (IEEE, 1999a, 2003a). This PHY is mainly considered here. IEEE 802.11a is also sometimes referred to as Wi-Fi5, to highlight that 802.11a networks operate in the 5 GHz band, and to reduce the apparent confusions of the many abbreviations.

5.7.3 IEEE 802.11b

IEEE 802.11b defines the HR/DSSS transmission mode with a chip rate of 11 Mchip/s, providing the same occupied channel bandwidth and channelization scheme as DSSS (IEEE, 1999b). The higher data rate is achieved through a transmission mode based on eight-chip CCK modulation. The code set of complementary codes is richer than the set of Walsh codes. At 11 Mbit/s, the spreading code length is 8 and the symbol duration is 8 instead of 11 chips, as it was with the DSSS. Data bits encode the symbols with QPSK and DQPSK.

5.7.4 IEEE 802.11c

This task group of the 802.11 working group finished its work by not developing an additional supplement standard, but by providing information for changes in other standards. The results of 802.11c were modifications of other standards, not a separate document. The 802.11c task group defined protocols for what is referred to as AP bridging. 802.11 APs can communicate with each other across networks within relatively short distances.

5.7.5 IEEE 802.11d

This standard is related to radio regulation in an international context. The use of the frequency spectrum is regulated by nations and is different from one nation to another. 802.11d provides procedures and protocols to let 802.11 networks operate compliantly to what is regulated, by introducing regulatory domains. If a station does not comply with the rules defined for a specific regulatory domain, it will not initiate transmissions, and not associate with a network. The domains are identified by information elements that are broadcast by the AP.

5.7.6 IEEE 802.11e

The 802.11e task group is defining enhancements to 802.11 to allow QoS support (IEEE, 2005b). It is described in detail in Section 5.5 of this chapter. 802.11e will work with any PHY extension.

5.7.7 IEEE 802.11f

AP handovers are supported by the Inter AP Protocol (IAPP), defined by 802.11f. A constant operation while the station is actually moving is supported when the IAPP is used. The concept of handover is familiar from cellular networks, and will need to be standardized, as APs and stations will be provided by different vendors.

5.7.8 IEEE 802.11g

IEEE 802.11g (IEEE, 2003e) combines the advantages of 802.11b (relatively large coverage) and 802.11a (higher throughput) by defining the application of the multi-carrier 802.11a OFDM transmission scheme in the 2.4 GHz band, in which originally 802.11b stations are operating. Therefore, 802.11g does provide up to 54 Mbit/s at the air interface. There is also an extended rate PHY mode based on DSSS single carrier (802.11b), which allows up to 33 Mbit/s. Further, because 802.11g and 802.11b stations are likely to operate at the same time in many scenarios, it is possible to use the DSSS-based preambles and headers together with the remainder of a frame being transmitted with extended rate PHY modes (single carrier or multi-carrier). With this multi-mode operation, 802.11g stations will be able to interwork and coexist with 802.11b networks, which makes 802.11g attractive for increasing the capacity of already rolled-out 802.11b networks.

5.7.9 IEEE 802.11h

Dynamic Frequency Selection (DFS) and Transmitter Power Control (TPC) are defined by this group, with a focus on 802.11a and the 5 GHz band (IEEE, 2003b). The reasons for applying these schemes are spectrum sharing and efficiency, QoS support and energy consumption.

To select the frequency channel to operate its BSS, an AP needs to know the status of all frequency channels. While the status of the current channel is available to the AP, the AP needs to collect the information about other channels as well, in order to initiate a channel selection. This will be performed via the standardized channel measurements by other stations and the AP itself. The channel measurement by the AP does not need to be standardized, as it does not need to report the measurement results to other stations. However, the AP measurement should be performed in such a way that the service disruption is minimized. Any other measurements must be standardized in the context of 802.11h. The channel measurements by stations will be (i) detection of other BSSs; (ii) measurement of Clear Channel Assessment (CCA) busy periods; and (iii) measurement of received signal strength statistics.

TPC is a difficult task in 802.11 networks, since as part of the DCF; every station needs to detect all transmissions of frames within its BSS. Thus, there are no peer links between two stations that are subject to TPC. However, to meet future regulatory requirements, and for increased spectrum efficiency, and in order to reduce interference imposed on other networks, TPC is standardized in 802.11h.

5.7.10 IEEE 802.11i

Security and privacy become increasingly important with the growing popularity of 802.11. There are problems in the algorithms for providing security defined in the legacy 802.11 protocol. 802.11i is tasked with improving the security by enhancing the Wired Equivalent Privacy (WEP) protocol (IEEE, 2004b).

5.7.11 IEEE 802.11k

The objective of 802.11k (IEEE, 2005k) is to provide measurements and frame formats by which a radio station can initiate, measure and assess the radio environment. Note that 802.11k is referring to radio resource measurements, and not radio resource management. Therefore, actions that make use of the new information are not defined, only the part of the entire management process is defined by 802.11k that involves the measurement, including requesting and reporting.

Note

1. Reproduced by permission of © 2003 IEEE. Source: S. Mangold, S. Choi, G. R. Hiertz, O. Klein, and B. Walke, "Analysis of IEEE 802.11e for QoS Support in Wireless LANs," *IEEE Wireless Communications*, vol. 10, no. 6, pp. 40–50, 2003. (Section 5.5 (in parts), Fig. 5.20, 5.21, 5.23.)

6

IEEE 802.15 Wireless Personal Area Networks

Guido Hiertz, Yunpeng Zang and Bernhard H. Walke

In recent years, wireless technology has increased the available data rate constantly. Especially 802.11 of the IEEE (IEEE, 2003a, 2005c) has gained worldwide market success. Data rates ranging from 1 Mb/s up to 54 Mb/s are available. Current development in Task Group n of IEEE 802.11 aims at 100 Mb/s, measured on top of the Medium Access Control (MAC) layer. However, new technologies such as Multiple Input Multiple Output (MIMO) antenna arrays promise higher data rates in the same channel width. Other technologies such as Ultra-Wide Band (UWB) or Multiband Orthogonal Frequency Division Multiplex (MB-OFDM) aim at even higher data rates to support up to 1 Gb/s. To allow for an efficient usage of the limited spectrum, new concepts for medium access are needed. This chapter describes today's centralized and decentralized MAC protocols for wireless personal area networks.

All of these technologies have a limited coverage range and the achievable throughput decreases rapidly with increasing distance. Therefore, the latest research and all major standardization bodies are developing highly efficient packet-relaying strategies to increase the coverage range by means of range extenders. IEEE 802.16 (WMAN, WiMax (WiMax-Forum, 2005)) has foreseen mesh clusters in the current standard, Task Group s (IEEE, 2005f) of IEEE 802.11 (WLAN, Wi-Fi (WiFi-Alliance, 2005)) has developed mesh networking among IEEE 802.11 access points, while TG5 (IEEE, 2005h) of IEEE 802.15 (WPAN) (IEEE, 2005g) has established a mesh task group for low- and high-rate applications.

This chapter is outlined as follows. First, we give an insight to the activities of the IEEE and the market for WPAN devices in Section 1. In Section 2 we explain the high-speed WPAN standard IEEE 802.15.3. Explanation of the PHY (802.15.3a (IEEE, 2005i)) and MAC enhancements (802.15.3b (IEEE, 2005j)) follow. We then give the latest insight to the work of WiMedia (and the Multiband OFDM Alliance (MBOA)), which have competed with the DS-UWB alliance for leadership in 802.15.3a and have been successful in standardizing their proposal as an international standard for UWB-based WPANs (ECMA, 2005a, 2005b) by the European Computer

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

Manufacturers Association (ECMA), see Section 6.7. Section 6.8 summarizes this chapter. The IEEE 802.15 activities on WPAN mesh technology are described in Section 8.1.3.

6.1 Scope of 802.15

6.1.1 Objectives

The WPAN Working Group (WG) 802.15 began to standardize Bluetooth as an IEEE standard in 1999. This work has been accomplished by IEEE 802.15 task group 1 (TG1) with IEEE Std 802.15.1-2002 (IEEE, 2002c), which is based on the Bluetooth v1.1 specification. Further work on updating this standard to the Bluetooth v1.2 specification was taken by the 802.15 task group 1a (TG1a). Meanwhile, 802.15 WG also undertook to develop a recommended practice for the coexistence of IEEE 802.11b and Bluetooth. The results of this work became IEEE Std 802.15.2-2003 (IEEE, 2002d). In addition to Bluetooth-related standardization, 802.15 WG developed two WPAN technologies covering both high-rate WPAN (IEEE Std 802.15.3-2003 (IEEE, 2003f)) and low-rate WPAN (IEEE Std 802.15.4-2003 (IEEE, 2003d)). An overview on the organization of 802.15 WG is presented in Figure 6.1.

6.1.2 Different Subgroups

IEEE 802.15.1 (Bluetooth) foresees a basic mesh network architecture. In scatternet mode of operation, data can be relayed among several wireless links. To the authors' best knowledge no devices on the market currently support scatternet operation. A typical Bluetooth network has a master and one or more slave devices. A PDA could form a master in a network consisting of a headset, an MP3 player and a cell phone. In a scatternet, one of the slave devices becomes a master too. Depending on its communication partner, it works in either master or slave mode. Additionally, Slave devices may associate with multiple masters. Hence, complex network

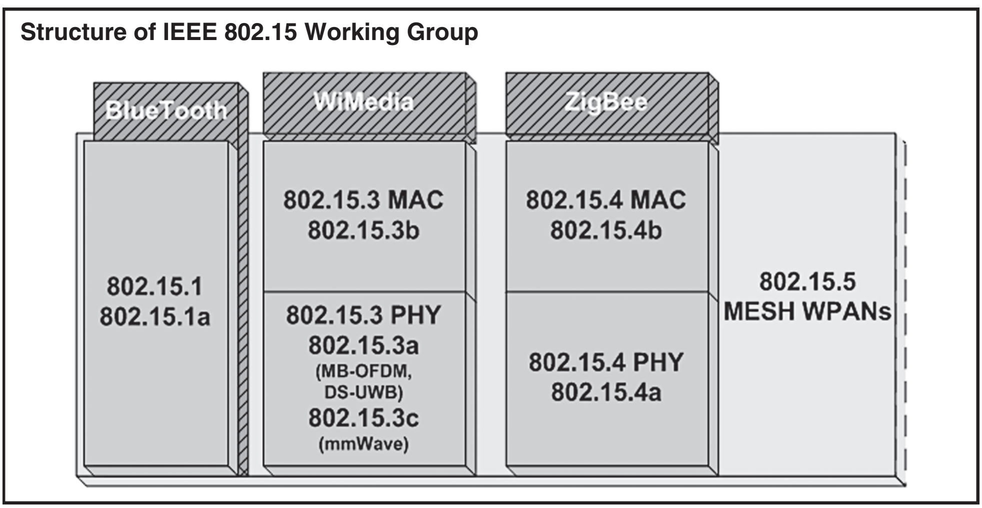


Figure 6.1 IEEE 802.15 Working Group organization chart.

topologies can be formed with Bluetooth. TG 802.15.1a implements revision 1.2 of Bluetooth and updates the IEEE 802.15.1 standard.

Task group 802.15.3 (TG2) describes means for support of coexistence of IEEE 802.15 WPAN and IEEE 802.11 WLAN that both share unlicensed frequency bands. TG2 is especially important for the 2.4 GHz ISM band that is used by both frequency-hopping Bluetooth (IEEE 802.15.1) and direct sequence spread spectrum based (IEEE 802.11b) and OFDM based IEEE 802.11g devices. High-rate WPAN is designed for portable consumer digital imaging and multimedia applications that require high bit rate and good QoS support. Besides the original task group 802.15.3 (TG3) that has now disbanded, there are still several groups actively contributing to the high-rate WPAN. They are: task group 802.15.3a (TG3a) working on an ultra-wide band based high-rate alternative PHY; task group 802.15.3b (TG3b) working on maintaining and enhancing the IEEE Std 802.15.3-2003 Medium Access Control (MAC); task group 802.15.3c (TG3c) working on another high-rate alternative PHY in millimeter Wave (mmW) band. WiMedia, an independent industrial alliance, is cooperating closely with 802.15.3 groups, focussing on the establishment of a common convergence platform for various wireless applications using high-rate WPANs. A low-rate WPAN has been developed mainly for applications such as wireless sensor networks that are more restricted in power consumption, reliability and complexity, besides data rate. At present, there are also two task groups working on the further developments of low-rate WPAN. One is task group 802.15.4a (TG4a) investigating an alternative PHY based on UWB technology for low-rate WPAN, and the other is task group 802.15.4b (TG4b) working on enhancements and clarifications to the completed 802.15.4 MAC. Zigbee is the well-known industrial association that is working on the basis of the IEEE low-rate WPAN specification; its topics are mainly the higher layer issues of low-rate WPANs. As IEEE 802.15.4 does not define any path selection methods, IEEE 802.15.5 (TG5), also known as the mesh WPAN task group, is chartered to develop a recommended practice for the necessary mechanisms that must be present in the PHY and MAC layers of WPANs to enable mesh networking. With respect to the MAC the mesh WPAN TG may amend additional procedures. The work of this group covers both high-rate and low-rate WPANs. So far, no proposal submission on low-rate mesh WPAN has been received by IEEE 802.15.5. Details of IEEE high-rate WPAN standardization and working groups, including IEEE Std 802.15.3-2003, TG3a, TG3b and SG3c, are introduced in the next section. A discussion on mesh WPANs can be found in Chapter 8.

6.2 802.15.3 – High-speed Wireless Personal Area Networks

Task group 802.15.3 (TG3) is chartered to standardize the high-rate WPAN. The standard was completed in 2003 (IEEE, 2003f), specifying a centralized MAC and a trellis-coded modulation-based PHY offering a data rate up to 55 Mb/s. TG3 has now been put into hibernation, however, there are three active groups still working on high-rate WPANs: TG3a, TG3b and SG3c. TG3a is responsible for the standardization of the UWB-based high-rate alternative PHY for 802.15 WPAN and for specifying a higher speed PHY amendment to IEEE Std 802.15.3-2003. This PHY will operate on the UWB frequency band and provide data rates ranging up to 480 Mb/s. TG3b takes the maintenance and enhancement of the original 802.15.3 MAC as its tasks. The newly finished 802.15.3 MAC specification that was designed for a maximum data rate of 55 Mb/s, is again under revision by TG3b to support higher data rates and better QoS. SG3c is working on another high-rate alternative PHY running on the 60 GHz frequency band that is also known as mmW band. An overview of the active groups for high rate WPAN is presented in Figure 6.2. The subsequent sections are organized as follows: the original 802.15.3 (TG3) WPAN MAC and PHY are reviewed in Section 6.3, followed by an introduction of the work in TG3a. TG3b and TG3c are discussed in Section 6.5 and Section 6.6, respectively.

Overview of IEEE 802.15.3 High-Rate WPAN Groups

	TG3	TG3a	TG3b	SG3c
MAC	Original TG3 MAC	–	Enhancement for original TG3 MAC	–
PHY	Original TG3 PHY	Amendment on UWB PHY	–	Amendment on mmWave PHY

TG3 WPAN systems have the following features:

- Ad-hoc network topology
- High data rate
- QoS support
- Power saving
- Low cost

Figure 6.2 Scope of the IEEE 802.15.3 task groups.

6.3 Task Group 3

Due to the insufficient ability of Bluetooth in supporting multimedia and high-rate applications, IEEE 802.15 WG started the standardization of high-rate WPAN (TG3) according to the motivation from several companies in 1999. Intended for high-rate multimedia applications, the TG3 WPAN system is distinguished from other WPAN/WLAN systems by the following features:

- Ad-hoc network topology
- High data rate
- Power saving
- Low cost

The single-carrier PHY specified in the TG3 standard is based on trellis-coded modulation that can provide a data rate up to 55 Mb/s. A centralized MAC protocol is specified in the TG3 standard with the capability of supporting QoS for both isochronous and asynchronous applications, as introduced in the following subsections.

6.3.1 802.15.3 Medium Access Control

The basic functionalities provided by TG3 MAC are:

- Fast connection time
- Ad-hoc networking
- Transmission with QoS support
- Security
- Dynamic membership
- Efficient data transfer

6.3.1.1 802.15.3 Network Topology

Being an ad-hoc wireless network, the 802.15.3 WPAN is basically organized into piconets, each comprising a number of independent data devices (DEV) and one piconet coordinator (PNC), see Figure 6.3. The PNC duty is to provide the basic timing information through beaconing for the whole piconet, and to manage the QoS requirements, power save mode and access control to the piconet. A typical 802.15.3 piconet topology is shown in Figure 6.4.

At starting, a device first performs a passive scanning of all channels trying to find an existing piconet. If any piconet is found, i.e. a beacon signal at a certain channel is received from a PNC, the device will try to associate with it. If no PNC is found or association to existing PNCs fails, the device may decide to start its own piconet by sending its own beacon. It may happen that no available channel can be found for creating a new piconet. Then, a device may start a dependent instead of an independent piconet.

Two kinds of dependent piconets are defined for 802.15.3 systems, namely the child piconet and the neighbor piconet. Similar to the scatternet topology in 802.15.1, the child piconet and neighbor piconet architectures are shown in Figure 6.4.

As the 802.15.3 WPAN is a synchronized system with central control, a dependent piconet depends on the time allocation from the parent PNC and is synchronized with the parent piconet. The motivation of having child piconets is to extend the coverage area of the parent piconet or shift some computational load or memory consumption to another PNC-capable DEV. A neighbor piconet is used for sharing the frequency spectrum between different piconets if there is no vacant PHY channel available. The most important distinction between these two kinds of dependent piconets is that the child PNC is a member of the parent piconet, thus is able to exchange data with any DEV in the parent piconet, while the neighbor piconet is an autonomous piconet and its PNC is not a member of the parent piconet. Therefore, the neighbor PNC cannot exchange data with any DEV in the parent piconet. Both child and neighbor piconets depend on the Channel Time Allocation (CTA) from the parent piconet and have their own Piconet ID (PNID). The

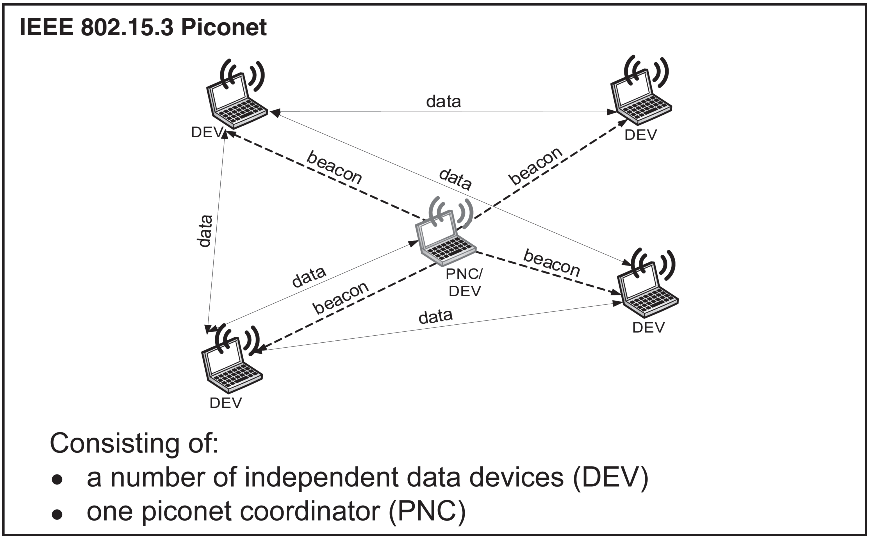


Figure 6.3 802.15.3 piconet elements.

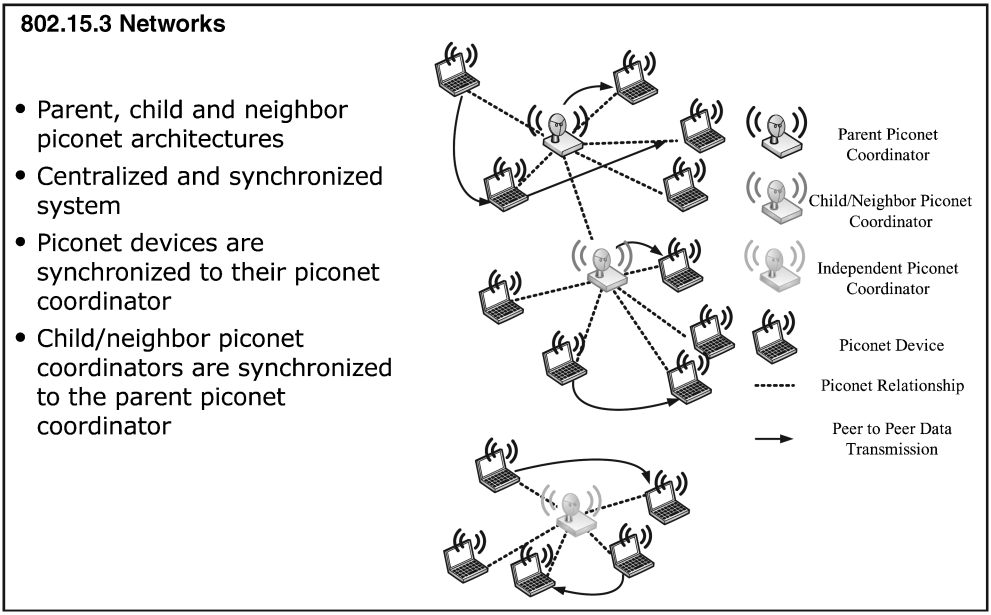


Figure 6.4 IEEE 802.15.3 network topology.

parent piconet may have more than one child or neighbor piconets and dependent piconets may have their own child/neighbor piconets within the CTA they get from their parent PNCs.

6.3.1.2 802.15.3 Medium Access Control

In principle, 802.15.3 uses a medium access control protocol that is a combination of CSMA/CA and TDMA, which is based on a synchronized superframe structure, as shown in Figure 6.5.

An 802.15.3 MAC superframe consists of three parts: beacon, Contention Access Period (CAP), and Channel Time Allocation Period (CTAP), also known as Contention-Free Period (CFP). The beacon is generated by a PNC and is used to announce piconet-related information such as superframe length, CAP and CTAP duration, network identity and timing synchronization stamp. The duration of a superframe may vary from one superframe to the other. The minimal length of a superframe is 512 μs and the maximal length is 65.535 μs. All devices (DEV) in a piconet must synchronize their local clocks to the PNC reference clock delivered in the beacon of every superframe. The synchronization applies for child and neighbor piconets as well. The CFP is reserved for transmission of QoS sensitive traffic that uses TDMA as the channel access mechanism, while CAP is mainly used for CSMA/CA based channel access to transmit asynchronous data and commands. During CFP a varying number of Channel Time Allocations (CTA) and Management Channel Time Allocations (MCTA) are scheduled, used for data and command transmission, respectively.

6.3.1.3 Contention Access Period (CAP)

Similar to IEEE 802.11, the 802.15.3 MAC protocol also employs CSMA/CA in CAP with the exception that no transmission may exceed the duration of the CAP.

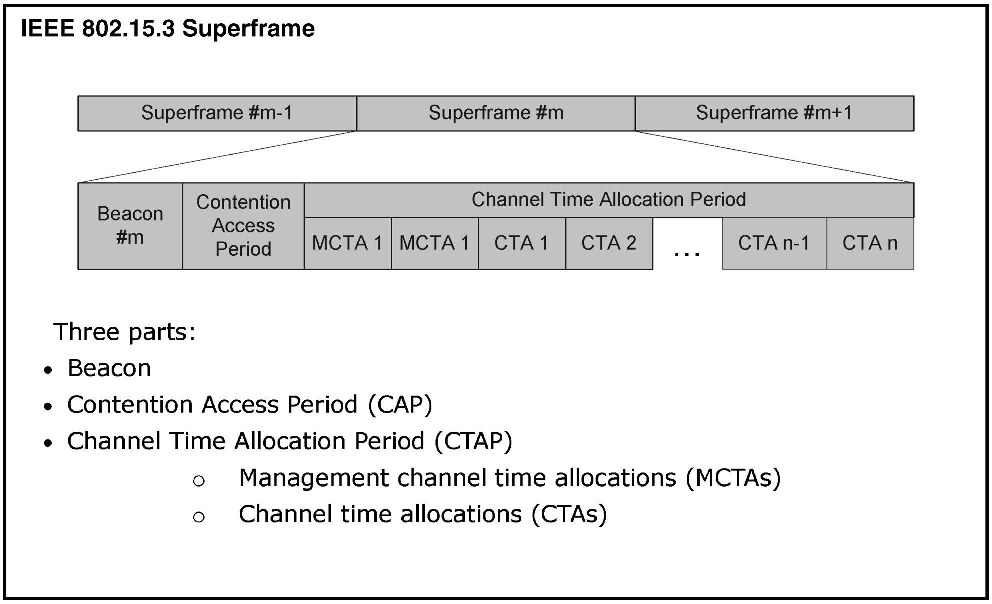


Figure 6.5 802.15.3 MAC superframe structure.

As shown in Figure 6.6, a backoff algorithm must be applied before sending any frame other than the immediate acknowledgment frame, and the channel has to be sensed idle for a Backoff Interframe Space (BIFS) before any backoff may be carried out. The backoff slot duration is specified as the time duration needed for sensing the channel state. The amount of backoff slots is randomly generated according to the backoff_window size that depends on the retry_count of the currently processed frame. The backoff counter decrements by one each time

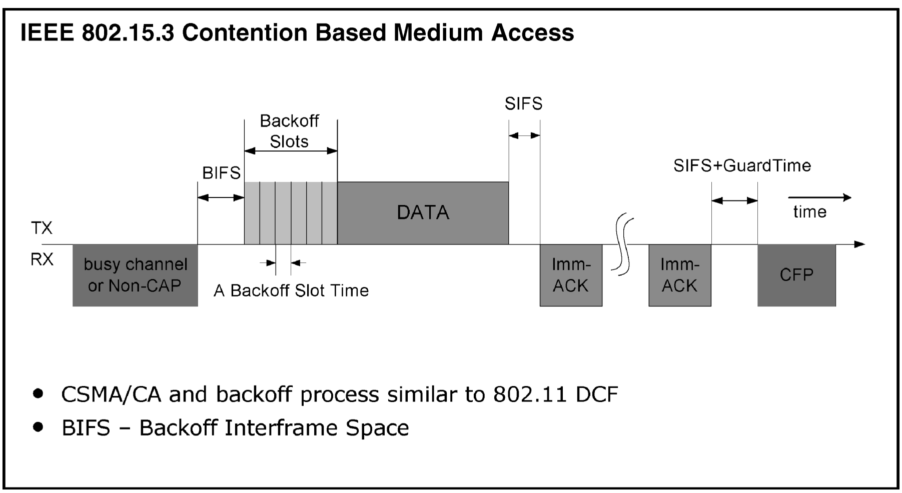


Figure 6.6 IEEE 802.15.3 CAP medium access control protocol.

the channel is sensed idle for a backoff slot time duration. When the backoff counter reaches zero the frame is transmitted.

Each time a transmission fails, e.g., no immediate acknowledgment is received by the transmitter, the retry_count of the current frame is incremented until its maximal value is reached. The retry_count value also directly affects the backoff_window size.

The channel state busy, and all superframe states other than CAP will suspend the backoff counter.

In this way, the DEV may use the CAP for transmitting its asynchronous data. A drawback of the 802.15.3 CSMA/CA protocol is that it lacks QoS that is provided by other MAC protocols like Enhanced Distributed Channel Access (EDCA) of 802.11e or Prioritized Channel Access (PCA) of the WiMedia MAC.

6.3.1.4 Channel Time Allocation Period (CTAP)

Different from the CSMA/CA scheme used in CAP, strict TDMA is used during CTAP, the CFP. The time during CFP is divided into CTAs that are managed by the PNC of the piconet. A CTA stands for a guaranteed start time and reserved transmission duration, where no other DEVs will compete for the channel time resource. CTAs may be allocated by the PNC for both isochronous flows and asynchronous data packets. Rather than for a periodical CTA in each superframe, asynchronous data transfer is requested for a total amount of time in one superframe only.

For isochronous data flows, there are two types of CTAs dependent on how the PNC manages the CTAs: dynamic CTA and pseudo-static CTA. The position of dynamic CTAs in a superframe can be arranged by the PNC independently from superframe to superframe. While the pseudo-static CTAs, which are only used for isochronous streams, are relatively fixed positioned in every superframe in order to allow the destination DEV to be able to receive the data even if it has missed the beacon for several superframes. The position of pseudo-static CTAs in a superframe may be reallocated by the PNC if there is a need, with a frequency of change that is much lower than for dynamic CTAs.

Requests for channel time are generated by DEVs that have data to transmit during the CFP, and sent during the CAP. The duration of the channel time requested must cover the data transmission time itself, the Short Interframe Space (SIFS) and the acknowledgment frame so that the PNC is able to transmit an acknowledgment immediately, if the packet was received successfully. Upon receiving a request, the PNC evaluates the current channel usage and all pending requests for CTAs. The PNC allocates CTAs to DEVs by announcing the CTA allocation in the beacon of the next superframe.

The allocated CTAs are used by the owner DEV like a TDMA channel, since no competition by other DEVs is possible. However, a DEV must not extend its transmission beyond the end of the CTA assigned to it by the PNC. CTA allocation also includes the SIFS and acknowledgment frame duration required to finish a complete cycle, data-SIFS-ACK. Between adjacent CTAs, a guard time plus an SIFS duration are always required to avoid overlap of use of adjacent CTAs due to drift of a DEV's local clock.

In addition to the CTA used for data transmission, there is another one used for management (M) called MCTA that has PNCID as either the source ID or the destination ID in the CTA. The MCTA is used by the PNC to transmit or receive the command frame.

6.3.1.5 802.15.3 Data Transmission

To reduce the packet error ratio (PER) for frames of large size, 802.15.3 MAC supports fragmentation and de-fragmentation of data frames. Any data or command frame with a size greater than a certain threshold should be fragmented. All fragments have equal size, which is no less than pMinFragmentSize, except for the last one in a sequence.

A DEV can choose one of three Acknowledgment (ACK) policies when transmitting frames: no-ACK, immediate-ACK (Imm-ACK) and delayed-ACK (Dly-ACK). No-ACK policy is used for frames that do not require guaranteed delivery, and the transmitting DEV just assumes that the transmission is successful without expecting any acknowledgment frame from the receiving DEV. If the Imm-ACK is used, an acknowledgment is always expected after an SIFS following the data frame transmission. Different from the other two ACK policies, the Dly-ACK applies only to the directed flow data frames, i.e. isochronous flows in CFP. With the ACK policy set to Dly-ACK, multiple frames can be acknowledged in a burst by one acknowledgment frame.

If an expected Imm-ACK frame or Dly-ACK frame is not received within a Retransmission Interframe Space (RIFS), the source DEV will start the retransmission of the frame if there is enough channel time remaining and the failed frame's retransmission limit has not been reached. Otherwise, a new frame should be sent instead of retrying the failed one.

The ACK policies in 802.15.3 MAC are quite similar to the ones used in IEEE 802.11e and WiMedia MAC, except that the Block-ACK (B-ACK) in WiMedia MAC can be used for both isochronous and asynchronous transmissions.

6.3.1.6 802.15.3 Network Security and Robustness

The 802.15.3 MAC protocol provides two security modes: mode 0, the open mode for no security membership and payload protection; and mode 1 with secure membership and payload protection.

The following methods for maintaining network robustness and improving coexistence performance are supported in 802.15.3 MAC: Transmit Power Control (TPC), Dynamic Channel Selection (DCS), and network coordinator handover.

TPC is carried out in two ways: Either in CAP every DEV uses the same transmit power to prevent unfair channel contention by using high transmission power, or, i.e. during CFP, the optimal transmission power is decided by the transmission pair (source and destination DEVs) in order to reduce the interference to other piconets.

DCS is another way to avoid interference from other piconets also running on the same channel. After the PNC has evaluated the status of the current channel and other channels, it may determine to change the current channel. Channel switching may be performed by the PCN without interrupting service provided by the piconet. If the current PNC is about to power down or leave the piconet, it may handover the control of the piconet to a DEV that is PNC capable. After successful handover of the coordinator function to a new PNC, the old PNC may leave the piconet without affecting the association relationship and service provided by the piconet to its DEVs.

6.3.1.7 802.15.3 Power Management

To enable a long operation time of battery-powered DEVs, 802.15.3 defines three power save modes. In addition to the normal ACTIVE mode: DEVs may run in Device Synchronized Power Save, Piconet Synchronized Power Save and Asynchronous Power Save mode.

6.3.2 802.15.3 Physical Layer

The 2.4 GHz PHY of the TG3 standard operates in the frequency band from 2.4 to 2.4835 GHz, which is available for unlicensed used in most regions of the world, see Section 3.2.3. The 802.15.3

IEEE 802.15.3 Channel Plan at 2.4 GHz

Channel ID	Center frequency	High-density	Low-density
1	2.412 GHz	X	X
2	2.428 GHz	X	
3	2.437 GHz		X
4	2.445 GHz	X	
5	2.462 GHz	X	X

IEEE 802.15.3 PHY Modes at 2.4 GHZ

Modulation type	Coding	Data Rate
QPSK	8-state TCM	11 Mb/s
DQPSK	None	22 Mb/s
16-QAM	8-state TCM	33 Mb/s
32-QAM	8-state TCM	44 Mb/s
64-QAM	8-state TCM	55 Mb/s

Figure 6.7 IEEE 802.15.3 physical layer overview.

physical layer specifies two channel plans with four channels and three channels, respectively. The four-channel plan is intended for high density applications, while the three-channel plan provides a better coexistence with IEEE Std. 802.11b-1999 networks as shown in Figure 6.7.

The 2.4 GHz PHY supports five data rates ranging from 11 to 55 Mb/s. The base rate of 22 Mb/s is uncoded, while all other PHY modes use Trellis-Coded Modulation (TCM), see Figure 6.7.

The base rate, 22 Mb/s, is used for MAC and PHY header transmission to ensure traffic detection at all DEVs. In the 2.4 GHz PHY, the header checksum appended to the MAC header is calculated over both the MAC and PHY header, in favor of efficiency. In order to increase the probability of correctly receiving the MAC and PHY header in the 11 Mb/s PHY mode, the MAC and PHY header is repeated twice.

The PHY preamble uses a constant-amplitude, zero-autocorrelation sequence due to its ability of obtaining synchronization, timing information and frequency offset.

The 2.4 GHz PHY requires an on-air bandwidth of only 15 MHz which allows operation of more channels and reduces interference to other systems as well as reduces susceptibility to interference from other systems.

6.4 Task Group 3a

As applications such as wireless video and HDTV are becoming more prevalent, higher data rates are required to support the need. TG3a of IEEE 802.15 WG was founded to develop an alternate PHY layer for WPANs targeting a higher speed wireless link for applications that involve imaging and multimedia. It should be based on UWB transmission. According to the fundamental requirements of TG3a (IEEE, 2002b), the intended PHY should be able to provide data rates of 110 Mb/s at 10 m, 200 Mb/s at 4 m and 480 Mb/s at even reduced range, at a PER below 8 % for 1024-byte frame body size.

The IEEE 802.15 3a task group was approved for its project authorization request in December 2002 and issued a call for proposal in early 2003. However, due to the unsolvable deadlock between two main competing proposals during the down selection process TG3a failed in getting a common baseline draft standard. The task group finally withdrew its project authorization request and was canceled in 2006. The two main proposals, although not being standardized by IEEE, represent the state of the art on the PHY of UWB-based PWAN. One of them is the Direct Sequence Ultra-Wide Band (DS-UWB) proposal from Motorola and Xtreme. It is based on Direct Sequence Spread Spectrum communication. The other uses a Multiband Orthogonal Frequency Division Multiplex (MB-OFDM) based solution supported by the MBOA. MBOA is an industrial alliance formed by leading semiconductor and consumer electronic companies such as Intel, TI, Philips and Sony to push the MB-OFDM proposal for 802.15.3a. In March 2005, MBOA and the WiMedia alliance combined into a single alliance towards the UWB-based WPAN market. The MAC layer developed on top of the MB-OFDM PHY by MBOA is also known as WiMedia MAC or MBOA MAC, see Section 6.7. Both PHY proposals are reviewed in Sections 6.4.1 and 6.4.2.

6.4.1 DS-UWB Proposal

The DS-UWB system (IEEE, 2004d) is based on direct sequence spreading of Binary Phase Shift Keying (BPSK) and Quaternary Bi-orthogonal Keying (4BOK) UWB pulses. Using convolutional coding with the coding rate of 1/2 and 3/4, the DS-UWB system can provide data rates of 26, 55, 110, 220, 500, 660, 1000 and 1320 Mb/s. The UWB spectrum is divided into two different bands that can be used independently by the DS-UWB system. As shown in Figure 6.8, the lower band occupies the spectrum from 3.1 to 4.85 GHz, and the upper band occupies the spectrum from 6.2 to 9.7 GHz.

Channelization: Up to six logical physical channels, which are identified through unique operating frequencies and acquisition codes, are designed in each band, i.e., in total 12 piconet channels are supported within the UWB spectrum. For a DS-UWB compliant device, the support

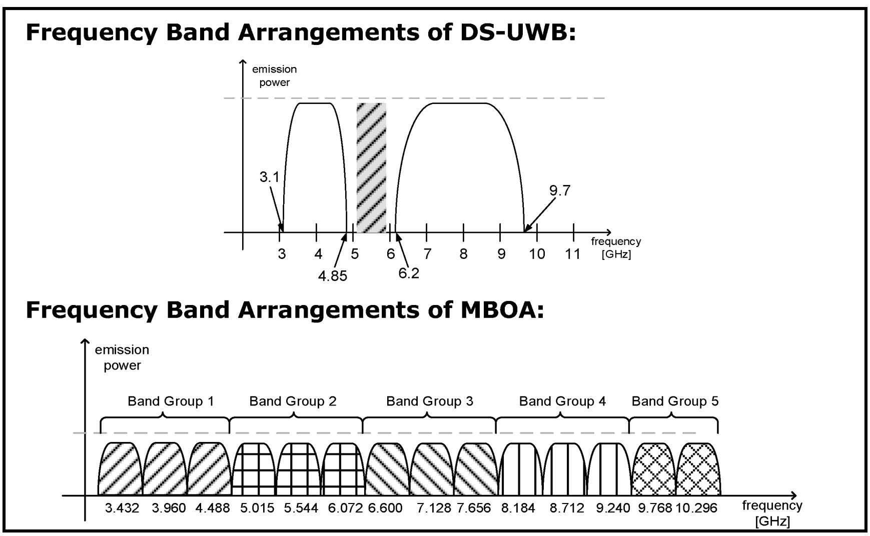


Figure 6.8 Frequency band arrangements of the DS-UWB and MBOA proposals.

of channels 1–4 (in the lower band) is mandatory and the support of channels 5–12 (in the upper band) is optional.

Modulation: The DS-UWB system uses BPSK and 4-BOK to modulate the data symbols while each transmitted symbol is composed out of a sequence of UWB pulses. Variable-length spreading code sequences with sequence lengths ranging from 1 to 24 pulses are used for achieving various data rates.

Forward Error Control: Convolutional coding with constraint lengths of 4 and 6 are used to correct transmission errors. The basic coding rate is 1/2 and the coding rate of 3/4 can be achieved through puncturing. To protect the coded data from burst error, convolutional interleaving is applied right after convolutional coding.

PHY header and preambles: The PHY header carriers information about the symbol rate, the number of bits per symbol and the used FEC scheme, from which the receiving DEVs can calculate the resulting bit rate. The PHY preamble is used in the DS-UWB system for clock/carrier acquisition and receiver training. During piconet establishment, the PNC selects one of six available Piconet Acquisition Codewords (PAC) for its PHY preamble used for acquisition (corresponding to the piconet channel being used.) Depending on the application bit rate, there are three preamble lengths: short preamble (5 μs) for high bit rate, nominal preamble (15 μs) as default preamble choice and long preamble (30 μs) for applications with extended range.

The parameters of PHY modes using BPSK and 4-BOK supported in DS-UWB lower band are listed in Figure 6.9.

6.4.2 MB-OFDM Proposal

Multiband-Orthogonal Frequency Division Multiplex (MB-OFDM) is the basis for this PHY of a WPAN system operating in the unlicensed 3.1–10.6 GHz UWB frequency band. To provide a data rate up to 480 Mb/s out of each 528 MHz band (see Figure 6.8), the MB-OFDM system employs

BPSK in Lower Band

Data Rate	FEC Rate	Code Length	Bits per Symbol	Symbol Rate
28 Mb/s	1/2	L = 24	1	Fchip/24
55 Mb/s	1/2	L = 12	1	Fchip/12
110 Mb/s	1/2	L = 6	1	Fchip/6
220 Mb/s	1/2	L = 3	1	Fchip/3
500 Mb/s	3/4	L = 2	1	Fchip/2
660 Mb/s	1	L = 2	1	Fchip/2
1000 Mb/s	3/4	L = 1	1	Fchip
1320 Mb/s	1	L = 1	1	Fchip

4-BOK in Lower Band

Data Rate	FEC Rate	Code Length	Bits per Symbol	Symbol Rate
110 Mb/s	1/2	L = 12	2	Fchip/12
220 Mb/s	1/2	L = 6	2	Fchip/6
500 Mb/s	3/4	L = 4	2	Fchip/4
660 Mb/s	1	L = 4	2	Fchip/4
1000 Mb/s	3/4	L = 2	2	Fchip/2
1320 Mb/s	1	L = 2	2	Fchip/2

Figure 6.9 DS-UWB PHY modes.

128-subcarrier OFDM. Convolutional coding with a coding rate of 1/3, 1/2, 5/8 and 3/4 is used for forward error correction. Multiple logical channels are supported via the Time-Frequency Code (TFC) that interleaves coded data over three frequency bands (called a band group). A total number of 18 logical channels are supported by applying the TFC over four three-band groups and one two-band group as shown in Figure 6.12. Support of band group 1 is mandatory.

In the MB-OFDM proposal, a channel spacing of 528 MHz is specified. Within each channel, each OFDM transmission consists of 128 separate subcarriers, 100 of which are used for data transmission, 12 as predefined pilot signals and another 10 as guard subcarriers. Each subcarrier is spaced 4.125 MHz from adjacent subcarriers and is modulated independently, using either QPSK or Dual-Carrier Modulation (DCM). In conjunction with a specific coding rate and spreading factor, MB-OFDM can provide a data rate from 53.3 Mb/s up to 480 Mb/s. The MB-OFDM PHY functions are presented as follows.

Scrambling: The data stream is scrambled with a pseudo-random sequence to prevent long sequences of 0s or 1s. This way, the peak to average power ratio is reduced.

Channel coding: FEC is used based on convolutional coding with rate 1/3 and a constraint length of 7. Additional code rates of 1/2, 5/8 and 3/4 are obtained by puncturing.

Data interleaving: Given a convolutional code, interleaving serves to optimize the FEC process. Three steps of interleaving are carried out in MB-OFDM PHY: (i) symbol interleaving across the OFDM symbols; (ii) intra-symbol tone interleaving; and (iii) intra-symbol cyclic shifts. Each interleaving block corresponds to six OFDM symbols over the air, which is the minimal transmission unit for MAC layer.

Subcarrier constellation mapping: Modulation schemes QPSK and DCM both have the same mapping efficiency of two bits coded data per subcarrier.

OFDM modulation is applied through a 128-point IFFT. 122 inputs of the 128-point IFFT are mapped to 100 data subcarriers, 12 pilot subcarriers and another 10 guard subcarriers, see

PHY Parameters

Parameter	Value
Number of data subcarriers	100
Number of pilot carriers	12
Number of guard carriers	10
Number of total subcarriers	122
Subcarrier frequency spacing	4.125 MHz
Symbol interval (TSYM)	312.5ns

PHY Modes

Modulation	Code Rate	Date Rate [Mb/s]	Coded bits per OFDM symbol (N_{CBPS})	Conjugate Symmetric Input to IFFT	Time Spreading Factor (TSF)
QPSK	1/3	53.3	100	Yes	2
QPSK	1/2	80	100	Yes	2
QPSK	1/3	106.7	200	No	2
QPSK	1/2	160	200	No	2
QPSK	5/8	200	200	No	2
DCM	1/2	320	200	No	1
DCM	5/8	400	200	No	1
DCM	3/4	480	200	No	1

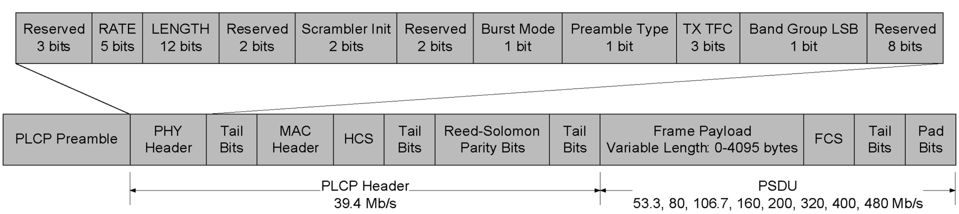
Figure 6.10 MB-OFDM PHY mode parameters.

Figure 6.10. For data subcarriers, two OFDM modulation schemes are employed depending on whether the conjugate symmetric input to IFFT is used or not. For the information data rates 53.3 Mb/s and 80 Mb/s, the conjugate symmetric input is used to IFFT, which maps 50 modulated data subcarriers symmetrically to 100 IFFT inputs of one OFDM symbol. This way, a frequency domain spreading with a factor of 2 is achieved. For information data rates other than 53.3 Mb/s and 80 Mb/s, no symmetric input is used. Twelve Pilot subcarriers are inserted at particular locations in order to make coherent detection robust against frequency offsets and phase noise. Another five guard subcarriers are located at each edge of the occupied frequency band, totaling 10, to fulfill the requirement on the frequency occupancy from regulatory bodies.

Time-domain spreading: To improve frequency diversity and the performance of simultaneously operating piconets, MB-OFDM defines time-domain spreading for data rates of 53.3, 80, 106.7, 160 and 200 Mb/s. It is realized by transmitting the same information over two OFDM symbols. Time-domain spreading combined with frequency-domain spreading (conjugate symmetric input to IFFT) and different coding rates enable the MB-OFDM PHY to provide eight transmission modes, Figure 6.10.

PLCP frame structure: The frame coming from the MAC layer is encapsulated in a PLCP Protocol Data Unit (PPDU), as shown in Figure 6.11. The frame begins with the Physical Layer Convergence Protocol (PLCP) preamble. Two modes of preamble are used in the MB-OFDM system: the standard PLCP preamble and the shortened streaming mode PLCP preamble. Both preambles comprise packet synchronization sequence, frame synchronization sequence and a six-symbol length channel estimation sequence. In addition to providing synchronization, carrier-offset recovery and channel estimation, the MB-OFDM preamble also provides the channelization information of the current channel, out of 18 logical channels. The shortened streaming mode PLCP preamble is used for frames with a data rate of greater than 200 Mb/s and as a part of a

PLCP Frame Structure



Interframe Spaces Defined for MB-OFDM

PHY Parameter	Value
pMIFSTime	6 * TSYM = 1.875μs
pSIFSTime	32 * TSYM = 10μs
pCCADetectTime	15 * TSYM = 5.625μs
pBandSwitchTime	9.47 ns

Figure 6.11 PLCP frame format and interframe spaces defined for MB-OFDM.

	DS-UWB	MB-OFDM
PHY technique	Single-carrier direct sequence spreading	Multiband-OFDM
Frequency management	Two bands (lower band: 3.1 ~ 4.85 GHz Higher band: 6.2 ~ 9.7 GHz)	Four frequency band groups. (three bands of each in group 1, 2, 3 and 4, two band in group 5) Each band covers 528 MHz.
Data rates	28, 55, 110, 220, 500, 660, 1000 and 1320 Mb/s	53.3, 80, 106.7, 160, 200, 320, 400 and 480 Mb/s
Modulation	BPSK or 4BOK	QPSK or DCM
Channel coding	Convolutional coding with interleaving	Convolutional coding with interleaving
Logical Channels	Total 12 channels (6 in lower band and 6 in higher band)	Total 18 channels through TFC

Figure 6.12 Comparison between MB-OFDM and DS-UWB PHY proposals.

burst transmission, other than the first frame. Due to the fact that the streaming mode preamble is much shorter than the standard preamble, using the streaming mode preamble can improve system throughput of burst transmission, significantly.

The preamble is followed by the PLCP header that contains both the PHY header and MAC header. According to the WiMedia MAC protocol, the length of the MAC header is fixed to 10 bytes. The maximal frame payload length is 4095 bytes. The PLCP header is always sent at a data rate of 39.4 Mb/s and has a fixed duration of 3.75 μs, while the remainder of the frame is transmitted at the operational speed of the link decided by the upper layer. The PHY-related MAC parameters for the MAC study in Section 6.7.3 can be found in Figure 6.11.

A short comparison between DS-UWB and MB-OFDM PHY proposals is presented in Figure 6.12.

6.5 Task Group 3b

The purpose of TG3b is to improve the ability of the original TG3 MAC to support emerging wireless multimedia applications, e.g. multimedia streaming, time synchronization, low latency data transfer and peripheral connectivity. An amendment for the 802.15.3 MAC standard is being developed by TG3b. This amendment contains changes to standard IEEE 802.15.3-2003 required to improve implementation and interoperability. It includes minor optimizations while preserving backward compatibility. The main contribution of this amendment is its enhancement to TG3 MAC with better QoS support. In addition, it corrects errors, clarifies ambiguities and adds editorial clarifications for the IEEE Std 802.15.3-2003.

6.6 Task Group 3c

Early in 1993, the Federal Communication Commission (FCC) of the United States established a general unlicensed band at 59–64 GHz, also known as mmW band, and launched the effort for minimizing the interference between the systems working on this band. The purpose of IEEE

Bit Rate	2 Gb/s @ 10m mandatory; 3 Gb/s @ 10m desired
Channelization	a minimum of 4 channels out of the bandwidth available in the U.S., Canada and Japan regulatory domains
Coexistence and Interference Resistance	Easy coexistence with other wireless devices including IEEE 802.11, IEEE 802.15.1, IEEE 802.15.3, IEEE 802.15.4, IEEE 802.16, ISM devices, mmW analog broadcasting systems and mmW collision avoidance radars
Form factor	Capable of fitting into a form factor consistent with a camera, PDA, etc.
Complexity	Minimal to enable mass commercial adaptation for cost sensitive products
Supplements to 802.15.3 functionality	PHY operates with the 802.15.3 MAC without fundamental changes in operation. (Three may be MAC modifications necessary to support the mmW PHY.)

Figure 6.13 Proposed mmW WPAN system PHY requirements.

802.15.3c TG is to transmit using an mmW-based PHY enhancement amendment to IEEE Std 802.15.3-2003.

Stimulated by an evolutionary market development for a group of applications, e.g. multiple HDTV video streams and wireless display requiring high data rates of more than 2 Gb/s with real-time dependence that will not be addressed in TG3 PHY, the TG3c aimed to define at 24 GHz and above, an alternative PHY clause for a higher data rate amendment to Std IEEE 802.15.3-2003. The mmW-based alternative PHY will achieve higher data rates, higher frequency reusage and superior coexistence than the existing TG3 wireless system standards. Multiple PHY modes and related data rates will be offered. At least one mandatory mode with a data rate of at least 2 Gb/s is required, and at least one optional mode with a data rate of 3 Gb/s or more is desired, in order to satisfy an evolutionary set of consumer multimedia industry needs for WPAN communications. The parameters listed in Figure 6.13 give an overview on the technical system requirements targeted by the TG3c mmW WPAN PHY.

Additionally, the mmW WPAN PHY developed in TG3c also distinguishes itself from all other WPAN systems by the expected coexistence performance and the higher spatial spectrum efficiency owing to the use of directional antenna.

6.7 WiMedia (Multiband OFDM) Alliance MAC Layer[1]

Next-generation WPANs are intended for a variety of applications. The WiMedia alliance has developed a new PHY and MAC protocol fitting the needs of a mass market. The WiMedia PHY and MAC have been accepted as an international ECMA standard (ECMA, 2005b), providing wireless data rates of up to 480 Mb/s using UWB technology. Besides its ability to support QoS, ease of use and reliability are major characteristics of this standard. The WiMedia standard provides seamless isochronous and packet-based asynchronous services. Details on the WiMedia protocol and its performance evaluation are discussed in the following subsections and can also be found in Hiertz *et al.* (2005a) and Zang *et al.* (2005).

6.7.1 Overview

The market for WPANs was dominated by the Bluetooth technology. Its ease of use and its energy-conserving procedures as well as the wide scope of applications led to a worldwide market success. The WPAN Working Group (WG) 802.15 began to standardize Bluetooth as an IEEE standard (IEEE, 2002c) in 1999.

Similar to the Bluetooth market the Universal Serial Bus (USB) (USB, 2000) is a tremendous success for wired short-range devices. With the introduction of version 2.0 of USB the maximum throughput is increased from 12 Mb/s to 480 Mb/s. Its easy and cheap deployment led to a wide acceptance. Besides the use of USB for Human Interface Devices for personal computers, a variety of QoS sensitive applications such as video, audio and high-speed device access, similar to IEEE 1394 (Firewire, (IEEE, 1996)) are available.

As an evolutional path towards replacement of wireline communications and as a successor to Bluetooth, WG IEEE 802.15 has developed a high-speed WPAN standard. 802.15.3 defined a new MAC layer and 802.15.3a tried to standardize a new PHY layer (IEEE, 2004e). However, 802.15.3a failed agreement on a baseline proposal. Besides that, the centralized MAC of 802.15.3 has shortcomings in ad-hoc network operation. This is why major industrial companies established the MBOA, which later merged with the WiMedia alliance, in order to develop new PHY and MAC specifications based on UWB OFDM technology to fit the needs of a high volume mass market. The WiMedia alliance comprises most of the world's biggest semiconductors, consumer electronics, PC and cell phone manufacturers. Because of its huge market support, it is expected that WiMedia will set the de facto standard for the next-generation WPAN. Applications of this standard include (but are not limited to) wireless USB, wireless IP (high-speed file transfer, use in handheld devices owing to the low power consumption, voice and video streaming, etc.), and wireless IEEE 1394.

6.7.2 Next Generation WPAN – WiMedia MAC

In this section, we describe the WiMedia MAC specified in standard ECMA-368. It is designed for high-speed, short-range communication infrastructureless networks.

Support of QoS is a crucial requirement for multimedia applications in WPANs. Thus, eight traffic classes may be differentiated by mapping to four Access Categories (ACs). Furthermore, WiMedia MAC supports asynchronous and isochronous traffic based on packets of arbitrary length of up to 4095 bytes.

6.7.2.1 Medium Access

To combine the efficiency of TDMA-based systems with packet-based technology WiMedia introduces the Prioritized Contention Access (PCA) and the Distributed Reservation Protocol (DRP) (Hiertz *et al.*, 2003; Hiertz and Habetha, 2003). While the first one is known from the EDCA introduced in Section 5.5.6, the latter one is based on an advanced reservation scheme providing collision-free access to the channel.

6.7.2.2 Prioritized Contention Access

Prioritized Contention Access (PCA) is a contention-based CSMA/CA MAC protocol relying on a prioritized backoff procedure (Figure 6.14).

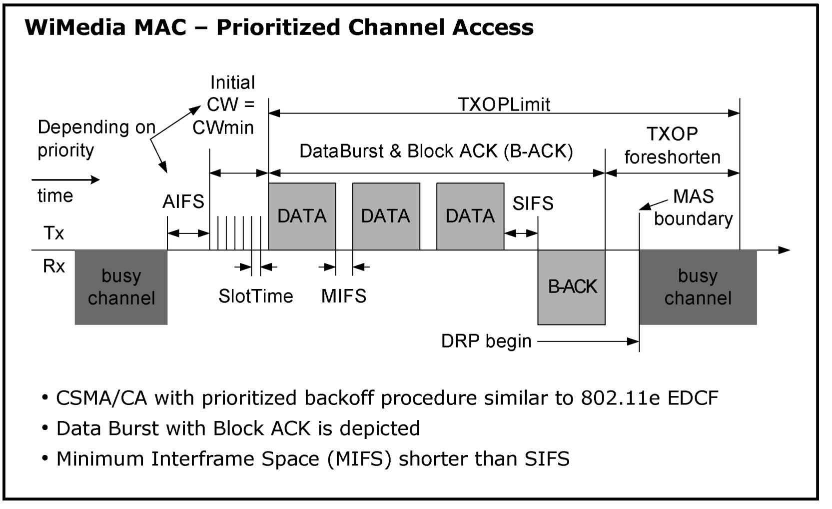


Figure 6.14 PCA is based on CSMA/CA as known from EDCA of Std 802.11e. Stations start transmitting after a random backoff period.

Virtual stations of different priority residing inside every station compete for channel access. Prior to a transmission attempt a station must sense the channel idle for a time duration of Arbitration Interframe Space (AIFS). Afterwards, it has to keep on sensing the channel for a backoff time duration of multiples of a SlotTime. The duration of a SlotTime is PHY dependent determined by the underlying WiMedia PHY. The backoff time duration is a random number of SlotTimes drawn from a uniform distribution over the interval (0, CW). The initial value of the contention window (CW) is CWmin. The duration of AIFS and CWmin depend on the priority of the backoff class. Whenever the station senses the channel idle during backoff slots, it decrements its slot counter, until it reaches zero permitting the station to transmit a data packet. If the station senses the channel as busy, it freezes its slot counter. After the channel is sensed idle for duration AIFS again, the backoff procedure continues counting down the remaining slots. With every failed transmission a station doubles its CW to reduce the probability of a collision with other stations.

6.7.2.3 Distributed Reservation Protocol

The Distributed Reservation Protocol (DRP) provides collision-free channel access. It announces future transmissions and thus allows devices to coordinate their channel access (Hiertz *et al.*, 2003; Hiertz and Habetha, 2003). Time is divided into superframes that comprise 256 Medium Access Slots (MAS) of length 256 μs each, see Figure 6.15. At the start of a superframe a Beacon Period (BP) comprising n beacon slots ($1 \leq n \leq$mMaxBPLength) enables transmission of n beacons, see Section 6.7.2.8. Each device (DEV) sends a single beacon per BP. At the end of the BP the Data Transfer Period (DTP) starts which lasts until the end of the superframe.

Devices must transmit beacons for DEV discovery, sleep-mode operation and other purposes, but especially for announcing reservations. All active devices must listen to all beacons in the BP and thereby learn from their neighbor DEVs about blocked MASs. Hence, a DEV reserving

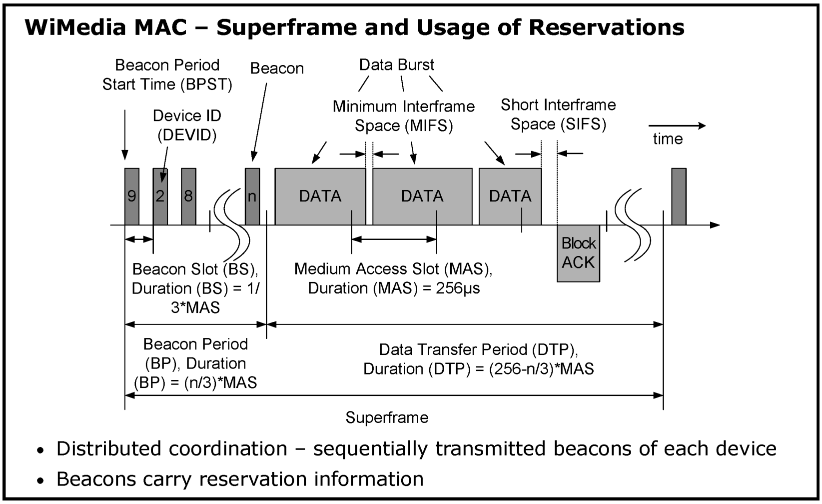


Figure 6.15 Each WiMedia MAC Superframe consists of 256 MASs. *n* MASs are used for the Beacon Period (BP). The rest is used for the Data Transfer Period (DTP).

MASs provides information about the start MAS and the number of subsequent MASs to be reserved in a superframe. These slots may be used for hard or soft reservations.

A hard reservation enables a station to start its transmission immediately at the start of a reserved MASs, since all other stations must complete their transmissions at SIFS plus a guard interval before the reserved MAS starts. The reserved MAS itself may be used solely by the reserving station and its communication partner DEVs. No other transmission is permitted there. Hence, isochronous real-time traffic via DRP is supported.

For service classes with less stringent QoS requirements, DRP provides soft reservation. During MAS that is soft reserved, the PCA is used. Only the owner of a reserved MAS may access the medium at the start of the first MAS of a sequence of soft reserved slots. All other DEVs must wait for access according to PCA. The purpose of soft reservation in DRP is to allow other DEVs access, if the owner that had soft reserved some sequential MASs does not fully use it.

A reservation can be negotiated explicitly or implicitly between sender and receiver(s) of a data flow. The sender may initiate an explicit DRP negotiation with a DRP Request frame send to the intended receiver(s) of the data. A unicast DRP Request frame is acknowledged immediately by the receiver, a multicast frame is not acknowledged. Afterwards, the intended receivers respond with a DRP Response frame, establishing the reservation.

An implicit DRP negotiation is performed by means of the beacon. The DEV that starts a DRP reservation includes a DRP Information Element (DRPIE) in its beacon frame. The intended receiver responds in its next own beacon with a corresponding Information Element.

In both, implicit and explicit DRP negotiation, the receiver must inform the sender whether the reservation must be shifted in time, is acceptable or not. Once a reservation has been established, both sender and receiver of a DRP reservation inform their neighbor DEVs about the reservation by including the reservation information in a DRPIE broadcast in their beacon frames.

If there is any not used capacity during hard reserved MASs, a frame exchange of Unused DRP Announcement (UDA) and Unused DRP Response (UDR) enables other stations to use it in the reserved channel.

6.7.2.4 Transmission Opportunities

Regardless whether a DEV accesses the channel via PCA or DRP, the duration of every frame exchange is bounded by a TXOPLimit. For Transmission Opportunities (TXOP) gained via DRP the TXOPLimit equals the duration of the reserved MASs. For PCA channel access the TXOPLimit is given per priority. However, the duration of a TXOP gained under PCA is further restricted by the closest DRP reservation, since no PCA transmission may delay or foreshorten any reserved MAS. When accessing the medium with PCA or making a new DRP reservation, a device has to respect all existing reservations. Besides these limitations, all decisions regarding the data exchange are solely up to the transmitting station.

6.7.2.5 Acknowledgement Policies

The WiMedia MAC defines three Acknowledgment (ACK) policies:

- No-ACK
- Immediate ACK (Imm-ACK)
- Block-ACK

To allow the receiver to distinguish between the desired ACK policies, each directed frame carries an “ACK policy” field in the frame control field inside the MAC header.

Immediate ACK policy works similar to standard 802.11. Each successfully received MAC Protocol Data Unit (MPDU) is acknowledged after a Short Interframe Space (SIFS) period by the receiver. The SIFS period is needed for transceiver turnaround and frame checking. It is used in between every frame exchange. With No-ACK policy no acknowledgment is generated at all. Block-ACK policy increases the efficiency, since a group of MPDUs is acknowledged with a single frame by the receiver (Hiertz *et al.*, 2005b).

6.7.2.6 Minimum Interframe Space and Frame Aggregation

In between consecutive frames pending for transmission, a station may improve the throughput efficiency by transmitting after Minimum Interframe Space (MIFS) instead of SIFS. Further, any station may benefit from frame aggregation, i.e. the concatenation of subsequent frames into a single data stream. However, the aggregated stream is subject to the same maximum size as any data frame payload. An exemplary aggregated frame is depicted in Figure 6.16.

6.7.2.7 Fragmentation and RTS/CTS Handshake

To reduce the Packet Error Ratio (PER) under bad channel conditions a device may choose to split any MSDU into a maximum of eight fragments.

To cope with hidden stations inherent in every wireless network, the Request To Send (RTS)/Clear To Send (CTS) handshake known from Std 802.11 may be used. However, the overhead of an RTS/CTS handshake is wasted if DRP is used.

6.7.2.8 Beacon Period and Beacon Frames

Each superframe starts with a BP. The maximum length of the BP is defined as mMaxBPLength, which is a multiple of beacon slots. During the BP devices sequentially broadcast beacons at the

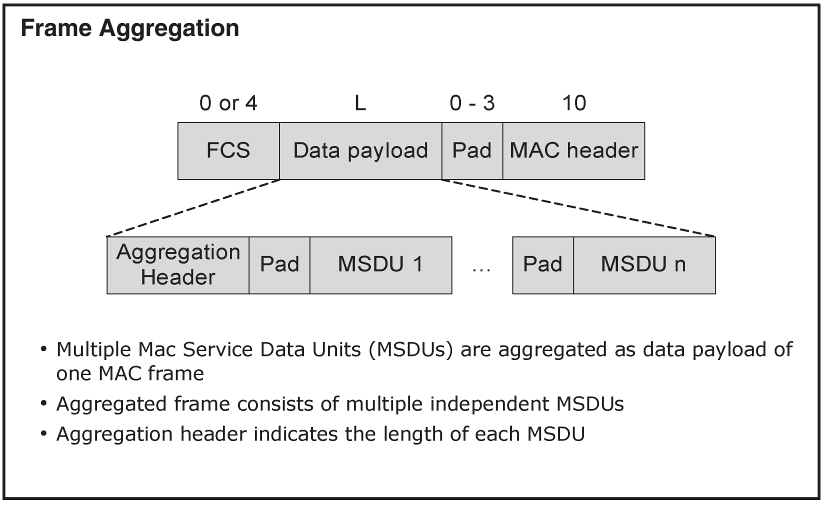


Figure 6.16 Frame aggregation.

base rate (currently 55 Mb/s). Each beacon will not exceed a length of mMaxBeaconLength which is equal to mBeaconSlotLength – SIFS – mBeaconGuardTime. The length of mBeaconSlotLength is one-third of an MAS.

From every received beacon a station learns about its direct neighborhood. In a beacon a DEV broadcasts, which beacon slots in its view are occupied by which DEVID. Thus, each DEV learns about its neighbor DEVs neighborhood, see Figure 6.17. Therefore, if during the most recent (mMaxLostBeacons +1) superframes a DEV neither receives a beacon in a specific beacon slot nor learns from neighbor DEVs' beacons that this slot is occupied, it treats the slot as empty.

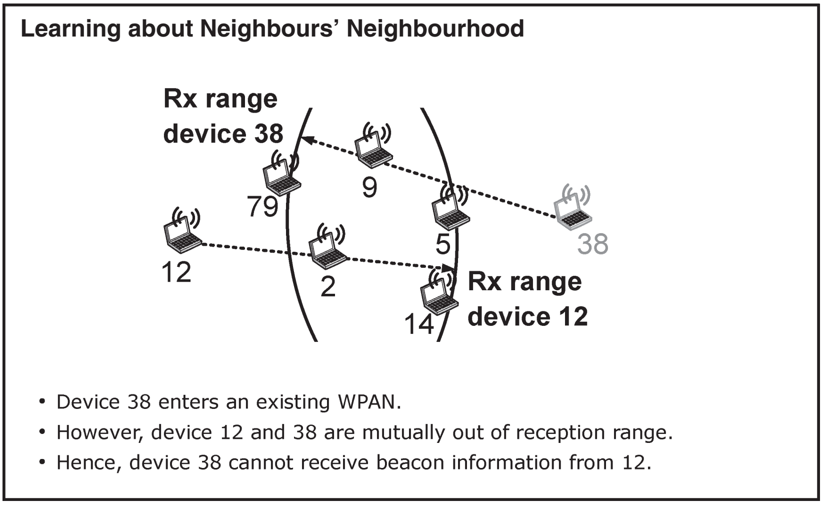


Figure 6.17 Broadcast of beacons in WPANs.

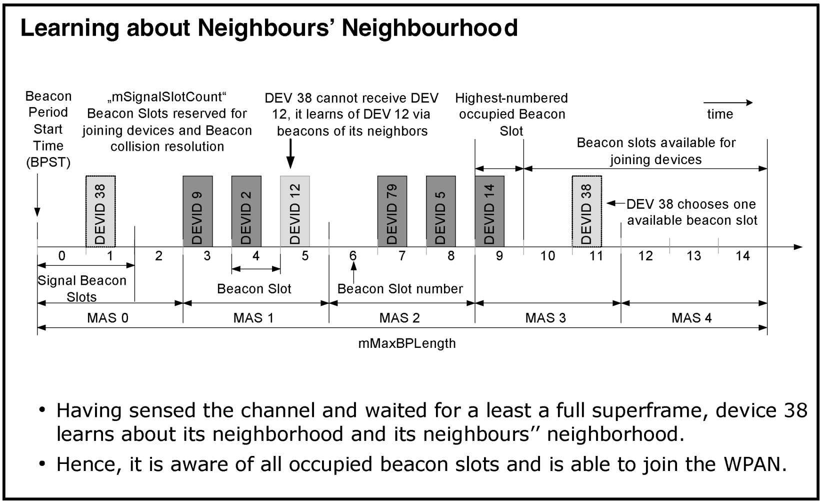


Figure 6.18 Beacon period and leaning about neighborhood.

Once a DEV is powered-up it scans for an empty beacon slot during at least one superframe. Then, it announces its presence in one randomly chosen Signal Beacon Slot. The beacon is sent a second time in a randomly chosen beacon slot in between the highest-numbered beacon slot and the end of the BP, as depicted in Figure 6.18. If all beacon slots are occupied, a device proceeds to send during the Signal Beacon Period and extends the BP by adding its beacon to the succeeding MAS of the BP of the next superframe.

Every *n*th superframe a DEV does not transmit a beacon to be able to detect beacon collisions. The value of *n* depends on the implementation. Additionally, a DEV may detect a beacon collision if neighboring DEVs report an empty beacon slot or a different DEVID than its own in the corresponding beacon slot.

Establishing a single, joint BP with overlapping WPANs is important for energy conservation, since the BP is the only time period a DEV must stay awake and be able to receive. Thus, battery-powered DEVs may stay in sleep mode mainly and only need to power up during the BP and DRP periods with which they are involved.

A BP may be contracted to reduce the BP length. Therefore, the highest-numbered DEV that does not experience any collisions may shift its beacon slot to the earliest empty beacon slot in the next superframe.

With overlapping networks, a coexistence support procedure is needed, since the WPANs involved are likely to have different BP placements. Hence, DEVs that detect alien BPs must refrain from interfering alien BPs and related DRP reservations. Since more than one BP transmitted at the same location means extra overhead and reduction of energy conservation for sleeping devices, WiMedia MAC specifies an algorithm to merge coexisting BPs. After announcing a protection DRP period for the alien WPAN, DEVs start shifting their beacons with the goal to merge their BP into the other to result in a single BP.

6.7.3 Simulative Performance Analysis

We use event-driven stochastic simulation to analyze the efficiency of the WiMedia MAC layer on top of the MBOA OFDM PHY. The delay results are presented as empirical Complementary

CDF, using the discrete Limited Relative Error (LRE) algorithm to gain sufficient statistical accuracy of correlated samples (Schreiber, 1988). All results presented in this section are within a maximum limited relative error of 5 %.

The Wireless Access Radio Protocol 2 (WARP2) simulation environment developed at the Chair of Communication Networks, RWTH Aachen University is used to gain the results. It is programmed in Specification and Description Language (SDL) using Telelogic's TAU SDL Suite. The error model used in WARP2 to accurately simulate the wireless medium is presented in Mangold *et al.* (2001a). To evaluate the efficiency of the WiMedia MAC layer, all simulations are performed with ideal channel condition, i.e., packet errors may happen only because of frame collision. In the following all DEVs are assumed within mutual reception range, i.e., no hidden stations are present in the simulation. The PHY parameters shown in Figure 6.11 apply.

We evaluate the maximum achievable throughput of WiMedia MAC using a simple scenario 1 with one DEV transmitting and one DEV receiving. In this simulation both DRP and PCA with a user priority of 7 are considered with the three kinds of acknowledgment policies and frame aggregation. The simulated superframe structure comprises a BP carrying 24 beacon slots and a data period with 248 MASs that is used as the TXOP length in DRP mode. The TXOP length of PCA mode is 1024 μs and the QoS parameter set of user priority UP7 used in this simulation is given in Figure 6.19. All results with ACK policy set to Block-ACK are derived with a burst buffer size of 16 frames. The aggregation function is evaluated with a timeout value of 100 μs. The MIFS and Stream Mode preamble are used in the stream mode transmission.

It can be seen from Figure 6.20 that throughput of DRP is always several Mb/s higher than with PCA, when comparing at the same ACK policy. The reason is that with PCA, even when applying the highest UP, the send DEV must wait for duration AIFS and perform a backoff before transmitting, decreasing throughput. As expected, in both DRP and PCA modes the No-ACK policy achieves the highest throughput, followed by the Block-ACK which reaches a slightly lower throughput than No-ACK due to the overhead of transmitting acknowledgment frames. As visible, frame aggregation substantially contributes to boost throughput with short packets that then may reach the same throughput as long packets.

QoS Parameter Sets

User Priority	AC	CWmin	CWmax	AIFSN
1	AC_BK	63	1023	7
2	AC_BK	63	1023	7
0	AC_BE	63	1023	3
3	AC_BE	63	1023	2
4	AC_VI	31	1023	2
5	AC_VI	15	31	1
6	AC_VO	7	15	1
7	AC_VO	3	7	1

- 8 priorities are used.
- Lowest priority: 1
- Highest priority: 7
- In general, the larger CWmin, CWmax, AIFSN, the lower the priority

Figure 6.19 PCA QoS settings in used in simulation.

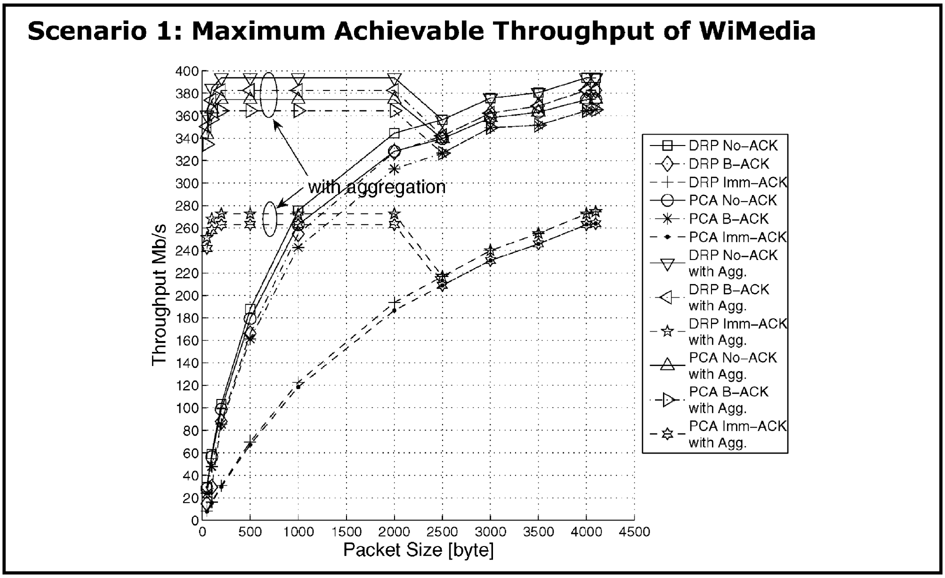


Figure 6.20 Throughput efficiency of WiMedia MAC under two access methods DRP and PCA, each combined with three ACK policies. Frame aggregation is used as a function of packet size. Data rate of PHY is 480 Mb/s. Scenario 1.

Simulation scenario 2 assumes a duplex route connection where two stations share channel capacity. Throughput and delay are evaluated to compare DRP mode (applied at both stations) to PCA mode assuming user priorities UP7 and UP0, respectively. The packet length is fixed to 1800 bytes, the ACK policy is set to Imm-ACK and frame aggregation is disabled.

Figure 6.21 presents throughput versus application traffic load. Under DRP, the two unidirectional links of the duplex connection have their own channel resources reserved, so that the same throughput is achieved by both. Under PCA access control, DEVs must contend, independent of the transmit direction for the channel, achieving about the same throughput. PCA with UP7 results in a higher throughput than with UP0, owing to shorter AIFS time duration and smaller backoff window size.

Delay performance results in Figure 6.22 indicate that PCA results are competitive with DRP under light traffic load (50 Mb/s), but not under heavy load. Under DRP access control the C-CDF of delay is bounded and with small variance of delay.

A WPAN scenario 3 has also been evaluated with a mix of services: One pair of DEVs with bidirectional VoIP (150 kb/s each direction, 120-byte packets, R1-2), a wireless streaming server providing HDTV to two clients (24 Mb/s each, 1500-byte packets, R3-4), one DEV handling file transfer (FTP) at 30 Mb/s (R5–6) and one DEV handling FTP at 100 Mb/s (R7), both with 1500-byte packets. The DEVs mentioned all use DRP. In addition, two DEVs handling best effort traffic (1500 bytes, R8–9) operating under PCA complete scenario 3.

From Figure 6.23 it can be derived that DRP enables guaranteed throughput support to high priority service classes, while PCA access fits the needs of low priority background services. From Figure 6.24 it can be seen that DRP results in well-bounded delay for the different service

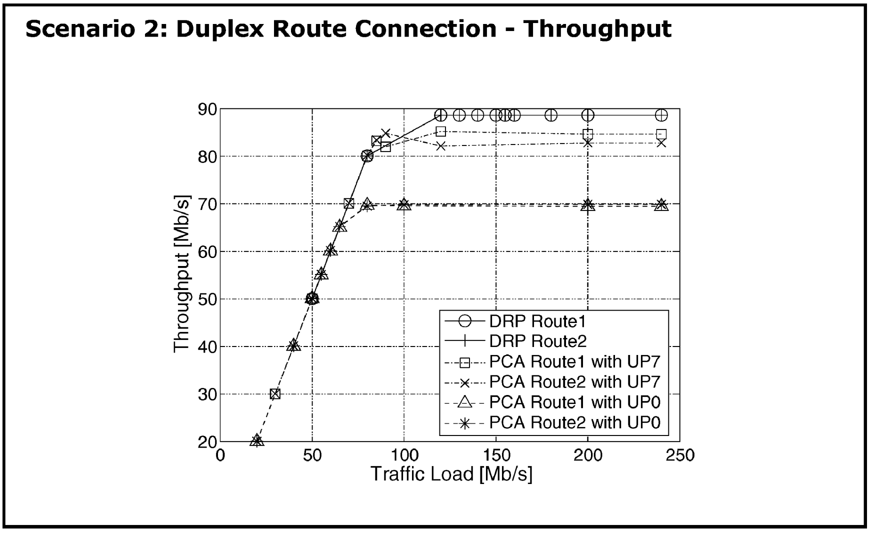


Figure 6.21 Throughput versus traffic load on a duplex route between two stations. DRP and PCA (with UP7 and UP0) compared. ACK policy is Imm-ACK. Packet length = 1800 B. PHY-mode: 480 Mb/s. Scenario 2.

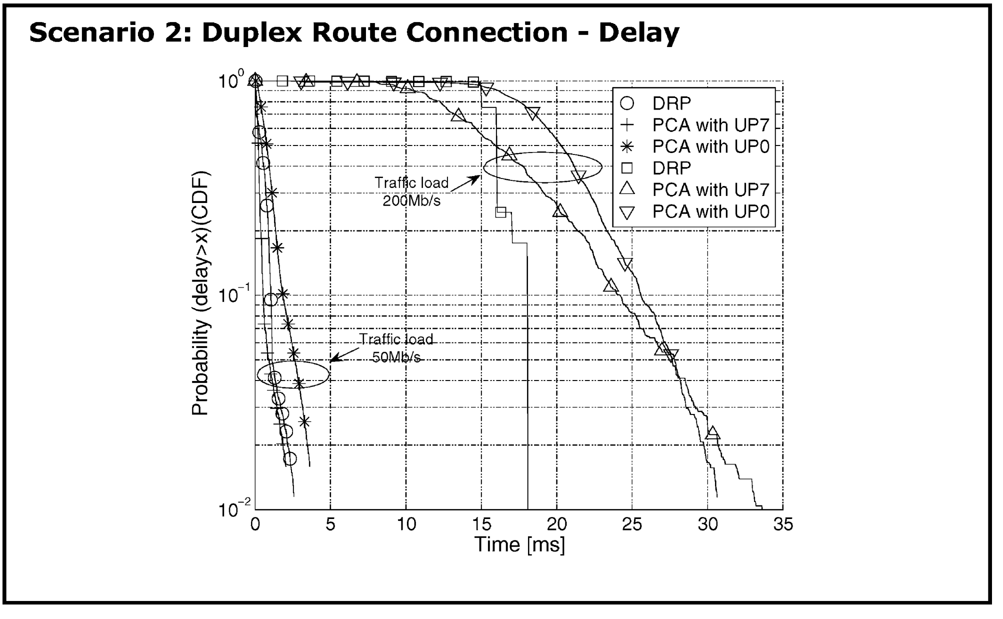


Figure 6.22 C-CDF of delay under light traffic load (50 Mb/s) and heavy traffic load (200 Mb/s) of the duplex route between two stations (Scenario 2) under DRP and PCA with UP7 and UP0. PHY-mode equivalent to 480 Mb/s.

Scenario 3: Mixed Scenario – Throughput

Route	Traffic Load (Mb/s)	Packet Size (B)	Access Method	Ack Policy	Throughput (Mb/s)
R1	0.15	120	DRP	No-ACK	0.15
R2	0.15	120	DRP	No-ACK	0.15
R3	24	1500	DRP	No-ACK	23.98
R4	24	1500	DRP	No-ACK	23.98
R5	30	1500	DRP	Block-ACK	30.00
R6	30	1500	DRP	Block-ACK	30.02
R7	100	1500	DRP	Block-ACK	99.89
R8	200	1500	PCA	Imm-ACK	13.86
R9	200	1500	PCA	Imm-ACK	15.04

Figure 6.23 Throughput result for the services of scenario 3.

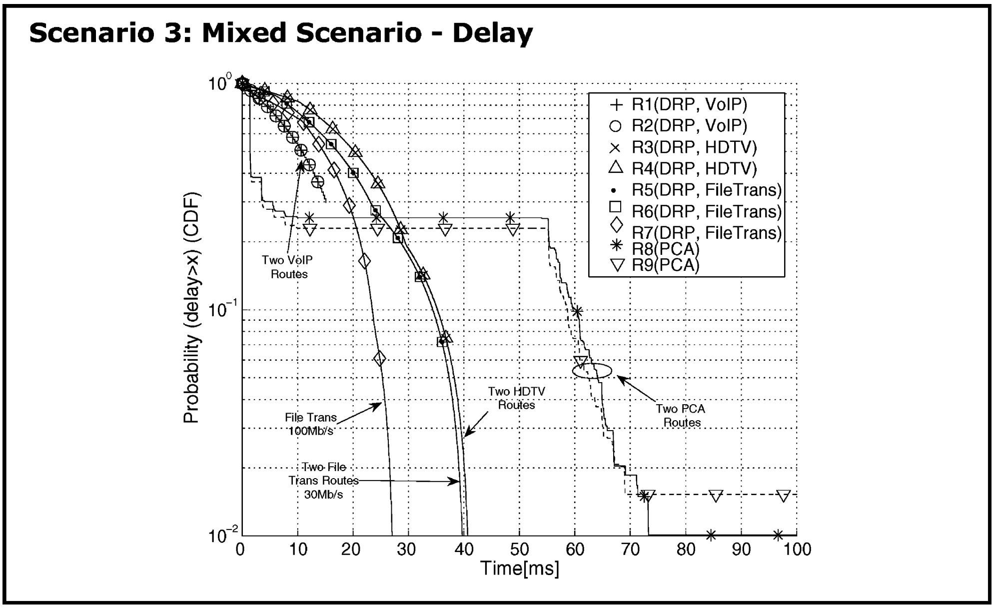


Figure 6.24 Delay evaluation of the mixed scenario with seven DRP routes and two PCA routes sharing a single 480 Mb/s channel.

classes, but low priority best effort (BE) service under PCA experiences large delay for some packets.

The shape of the PCA delay distribution reflects the slot reservation by DRP. In some superframes, BE packets may immediately access the channel under PCA, when MASs do not use up the channel capacity, and the delay is then very small. In other superframes BE packets experience large delays, owing to occupying the channel by DRP.

6.7.4 Conclusion

Simulation results indicate the WiMedia MAC protocol to be very efficient supporting real-time services very well using the DRP protocol. PCA (that is equivalent to EDCA) as a contention-based access protocol is not suited for delay-sensitive service classes.

Based on the complete PHY and MAC specifications of WiMedia a very powerful WPAN with data rates of several hundred Mb/s can be set up. Several companies have announced chipsets for 2006: WiMedia seems well positioned for the race of competing WPAN standards.

6.8 Next-generation WPAN Technologies

6.8.1 Market Perspective

Today's main applications for WPAN technology are characterized by low data rate, short range communication to synchronize Personal Digital Assistants (PDAs), cell phones and other portable devices. Some recent mobile devices include high bandwidth demanding devices such as portable MP3 players with internal hard disks, smart phones and digital cameras. These devices are designed for the price-sensitive market of consumer electronics. Thus, they use as a general purpose interface, the USB version 2.0 (USB, 2000), which supports up to 480 Mb/s data rate. Therefore, a successful wireless (WPAN) standard must be competitive to USB 2.0, since it is a key element for a seamless connectivity to existing devices.

6.8.2 PHY Technology

Current proposals for 802.15.3a (IEEE, 2005i) propose different technologies for high-speed WPANs. The proposed solutions based on DS-UWB and MB-OFDM have been discussed in this chapter.

The OFDM technology is widely approved and stable. OFDM devices have proven to be reliable and robust. Silicon design, simulation methodology, capacity analysis and design experience for OFDM based systems are available. Thus, a successful market introduction of the WiMedia PHY layer is very likely.

The DS-UWB technology claims to have benefits such as low complexity and low energy consumption. However, no circuit designs are available yet. To become a reliable and mature PHY technology, DS-UWB will need several rounds of development. In addition, existent Intellectual Property Rights make license cost of DS-UWB uncertain.

In summary, from a perspective of an early roll out to the market, OFDM-based technology appears of advantage compared to DS-UWB, so it is very likely that high-speed WPAN products will be available during the next few years. The potential of OFDM to be used in multi-user networks is expected to be further enhanced by OFDMA operation, see Section 2.3.5.

6.8.3 MAC Design

IEEE 802.15.3 describes a MAC protocol with central control. Different from that, the WiMedia MAC operates under decentral control. WPANs suffering from frequent topology changes (due to mobility and energy-conserving devices, which switch on and off) would benefit from using a MAC protocol with distributed control, since no central instance is needed. In a highly dynamic network, election and re-election of cluster heads, cluster establishment and intercluster forwarding add additional overhead to a WPAN network that may be prevented through distributed control.

Future WPAN technologies will be limited more and more by signal attenuation and reduced coverage range, since higher carrier frequencies will come into focus, and spatial frequency re-use will become a more important issue for license exempt bands.

Mesh technology (IEEE, 2005h) appears then to be a key element for future WPANs. The related MAC protocols will be based on distributed control, providing efficient spatial frequency reuse. Standard IEEE 802.15.3 MAC provides good services in application scenarios similar to today's Bluetooth technology. However, this standard cannot support establishment of fast-converging mesh networks, efficient relaying and self-organization, limiting its use in the future. New MAC protocols like MDCF (see Section 12.2) will be required able to extend the radio range by multi-hop relaying support. Besides this, support of coexistence of same and different standard wirelesss networks in the same band will be a key element for the latest wireless technology, which intelligently adapts to changing radio environments as discussed in the context of cognitive radio networks in Section 13.2.

Note

1. Reproduced by permission of © 2005 IEEE. Source: G. R. Hiertz, Y. Zang, J. Habetha, and H. Sirin, "Multiband OFDM Alliance – The Next Generation of Wireless Personal Area Networks," in Proc. of *2005 IEEE Sarnoff Symposium*, Princeton NJ, USA, April 2005. (Section 6.7 (in parts)).

7

IEEE 802.16 Wireless Metropolitan Area Networks

Christian Hoymann and Bernhard H. Walke

Wireless last-mile technology is becoming a challenging competitor to conventional broadband wired last-mile access systems such as Digital Subscriber Line (DSL), cable modems and even fiber optic cables. The IEEE has developed a standard for Broadband Wireless Access (BWA) systems namely IEEE 802.16. The amendments to the base standard enhance the system to support advanced antenna systems, mobile Subscriber Stations (SSs) as well as mesh and multi-hop deployments. This broadens the system's applicability far beyond the pure fixed wireless last-mile access.

This chapter is outlined as follows: After an introduction to the scope of 802.16 in Section 7.1, the basic deployment concepts, the reference model and targeted frequency bands of 802.16 are discussed in Section 7.2. Section 7.3 gives an overview on the structure of the 802.16 working group and its organization for standardization – all currently active task groups and their objectives are introduced. The multi-carrier Physical layer (PHY) of 802.16 that is based on Orthogonal Frequency Division Multiplex (OFDM) is described in detail in Section 7.4. Thereafter, the Medium Access Control (MAC) layer is introduced in Section 7.5 with a focus on the Time Division Duplex (TDD) mode of 802.16. Different system profiles are outlined in Section 7.6. The performance of 802.16 is evaluated in Section 7.7 and finally this chapter is summarized in Section 7.9. The subsequent Chapter 8 gives a detailed description of the mesh option of 802.16 and its benefits.

7.1 Scope of 802.16

The Wireless Metropolitan Area Network (WMAN) IEEE 802.16 specifies four different PHYs including two single-carrier (SC and SCa) modes, an OFDM and a mode based on Orthogonal Frequency Division Multiple Access (OFDMA) (IEEE, 2004a). One SC mode, which has been

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

specified for frequency bands between 10 and 66 GHz targets Line-of-Sight (LOS) communication whereas all other modes are designed for frequency bands below 11 GHz that in addition must cover Non-LOS (NLOS) links.

The OFDM-based transmission mode of the IEEE 802.16 standard has been standardized in close cooperation with the European Telecommunications Standards Institute (ETSI) whose standard is named High PERformance Metropolitan Area Network (HiperMAN) (ETSI, 2003b, 2003c). Thus, the HiperMAN standard and the OFDM-based transmission mode of IEEE 802.16 are nearly identical. Both OFDM-based physical layers shall comply with each other and a global OFDM system should emerge (Koffman and Roman, 2002). Both standards form the basis for the Worldwide Interoperability for Microwave Access (WiMAX) certified technology. The WiMAX Forum is an industry-led, non-profit corporation formed to promote and certify compatibility and interoperability of broadband wireless products such as IEEE 802.16 and HiperMAN (WiMax-Forum, 2005).

The main advantage of BWA technologies over wired systems such as DSL and cable modems results mainly from the high costs of the labor-intensive deployment of cables. “A 200-square-kilometer service area costs a DSL provider over $11 million. The same area can be served wirelessly for about $450,000” (Cherry, 2003). Apart from being wireless the above-mentioned BWA systems IEEE 802.16 and HiperMAN have been designed to meet today’s most promising challenges: NLOS operation capability cuts the deployment costs. Large cell radii allow for rapidly deployable infrastructure networks. This will decrease time to market for new broadband services, which will be crucial for the success of new operators. Networks become even more scalable by utilizing the optional mesh deployment. The system performance enables operators to offer services requiring high-peak bit rates. A substantial Quality-of-Service (QoS) support for packet-based services is provided by the system.

7.2 Deployment Concept, Reference Model and Target Frequency Bands

7.2.1 Deployment Concept

WiMAX’s network deployment consists of a Point-to-Multipoint (PMP) architecture where Base Stations (BSs) are the central, controlling units. On the one hand, they are connected to the operator’s core network and on the other hand they provide the wireless interface towards SS (see Figure 7.1). The BS’s connection to the core network is not part of the WiMAX network itself. It can be achieved by using cable or radio links. For instance DS3/E3 services provided by the plesiochrone respectively synchronous digital hierarchy, Frame Relay or DSL are available (Orthman, 2005). Alternatively, the WiMAX network itself can establish the backhaul to connect BSs, wirelessly, to the fixed network.

In the PMP deployment, each SS has a direct link to its BS. The wireless link might have LOS or NLOS characteristics. A repeater might even reinforce the signal. WiMAX business models foresee a variety of potential markets (Orthman, 2005). Thus, WiMAX can provide high-speed Internet access for residential customers, Small Offices and Home Offices (SOHO). It can alternatively supply Small to Medium Sized Enterprises (SME) with DSL or leased line services. As a local loop bypass, it will carry packet data as well as circuit-switched voice. Cellular operators will have the opportunity to replace its wired backhaul partly by a WiMAX network. Possibly even more relevant is the potential as backhaul for Wireless Local Area Network (WLAN) hot spot. Since 802.11 Access Points (APs) do need a high-capacity, cost-efficient backhaul solution for their rapid growth. A concept for realizing a meshed backhaul network based on 802.16 that

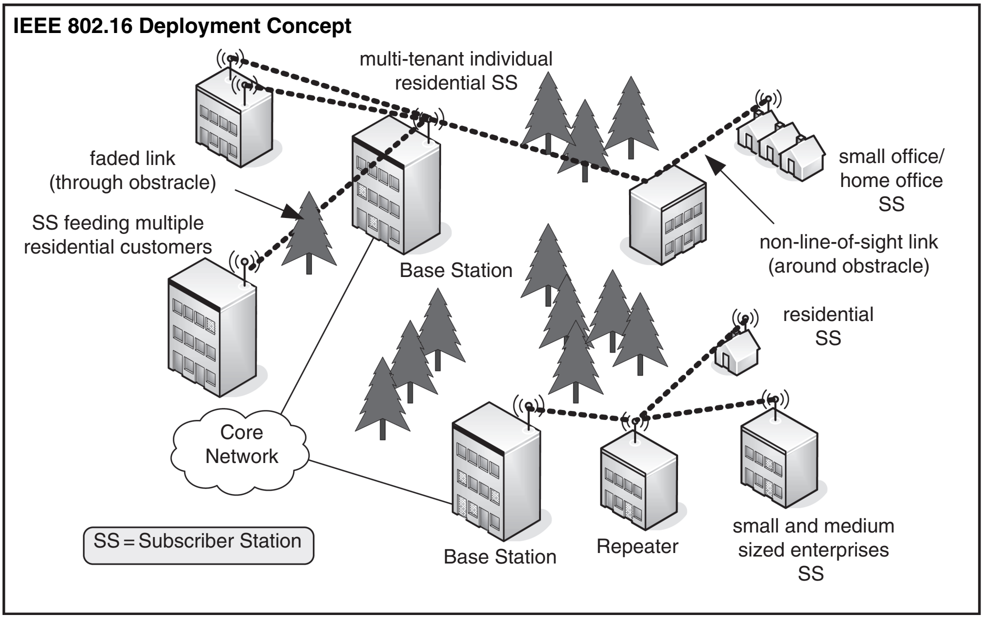


Figure 7.1 Deployment concept of IEEE 802.16 according to ETSI (2003a). (SS = Subscriber Station.)

connects APs serving 802.11 stations on the same frequency channel is introduced in Section 8.3. Other applications for WiMAX can be public safety services and private networks.

The IEEE standard foresees an optional mesh deployment. Within the mesh deployment, SSs no longer need a direct link to the BS, but they might be associated to a relay station that forwards data to/from the BS, see Section 7.3.8.

7.2.2 Reference Model

The IEEE reference model follows general IEEE 802 guidelines similar to the other WGs of 802 and specifies the Medium Access Control (MAC) and the PHY layer. Higher layer protocols as well as the management plane are outside the scope of the standard. Figure 7.2 shows that the MAC comprises three sublayers.

The service-specific Convergence Sublayer (CS) interfaces higher layers. It classifies external Service Data Units (SDUs) and associates them to the proper MAC connection. The CS may also process SDUs, e.g., to reduce overhead by performing Payload Header Suppression (PHS). Two CSs specifications are provided for interfacing Asynchronous Transfer Mode (ATM) as well as packet-based protocols such as IP, Point-to-Point Protocol (PPP), or IEEE 802.3 (Ethernet). Both CSs are outlined in Section 7.5.1.

The MAC Common Part Sublayer (CPS) carries key functions such as system and channel access, connection management and the application of QoS. Section 7.5.2 details the CPS.

Below the MAC CPS resides the security sublayer, which provides authentication procedures, a secure key exchange and encryption functions.

The IEEE 802.16 PHY specification defines multiple PHYs, each appropriate to a particular frequency range and application. Thus, it is left unspecified how to develop, implement and

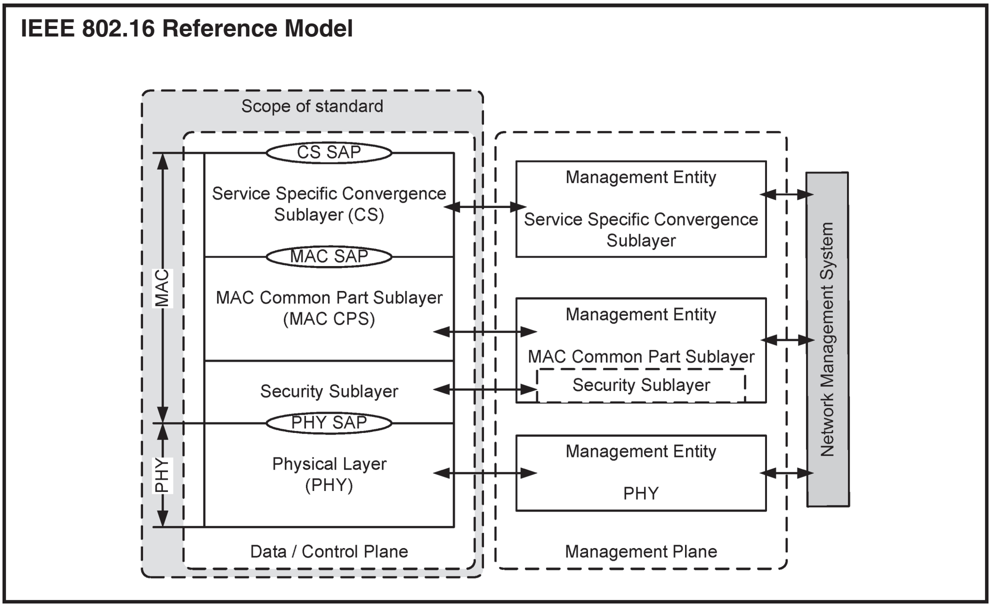


Figure 7.2 IEEE 802.16 reference model illustrating the scope of standard (IEEE, 2004a) and the protocol separation into layers.

deploy optimized systems with respect to available frequency bands, cell planning, equipment cost and targeted services. Another reason for having several PHY specifications was the lack of support for a single common PHY during standardization. The PHY specifications supported by 802.16 are discussed in Section 7.4.

7.2.3 Target Frequency Bands

Since the electromagnetic propagation conditions in the huge foreseen frequency range are not uniform, the IEEE 802.16 standard targets two different frequency regions between 10 and 66 GHz and below 11 GHz, down to 2 GHz.

At high frequencies, LOS radio propagation is necessary in order to enable proper radio communication, see Section 2.1. Since the wavelength is short the attenuation owing to shadowing is severe. Under LOS conditions the effect of multipath propagation can be neglected. Only the main path carries significant energy. Another advantage at high frequencies is that channel bandwidth is, typically, large in the order of 25–8 MHz, compared to radio services operated at lower carrier frequencies. Frequency bands preferred by the Electronic Communications Committee (ECC)/European Conference of Post and Telecommunications Administration (CEPT) for BWA ranges from 24.5 to 26.5 GHz and from 27.5 to 29.5 GHz (ERC, 2000, 2001c). Other potential bands are identified at 31.8 to 33.4 GHz and at 40.5 to 43.5 GHz (ERC, 1999, 2001a).

Below 11 GHz, the radio propagation conditions allow for NLOS system operation. Thus, the system has to deal with shadowing and multipath propagation resulting in variations of the received signal power resulting from both long-term and short-term fading. In license-exempt bands, users competing in the band may cause interference, hence regulation limits the allowed transmission power more restrictively there, compared to licensed bands. The interference has to be handled by the system, e.g., by means of an advanced power management and/or Dynamic Frequency Selection (DFS). Precautions against mutual interference of systems are introduced in Chapter 3 in the context of radio spectrum regulation.

Since a single harmonized frequency band below 11 GHz is not available, IEEE (2004a) and ERC (1998) recommend different bands, depending on the world region of interest. Licensed bands are available at 2.5 and 3.5 GHz in the majority of all countries. Some regions, e.g. Europe, additionally allocate bandwidth at 10 GHz as preferred bands for BWA (ERC, 1996). Due to the suitable amount of low-cost spectrum, license-exempt frequency bands at 2.4 and 5 GHz have been targeted nearly worldwide, see Figure 3.1(a).

Radio frequency technology for these frequencies is inexpensive since it is well known from 802.11a/b/g devices. License-exempt spectrum is strategically important for enabling grass root deployments in under-served, low-populated rural and remote markets. It might also be attractive to introduce better support of QoS by operating 802.16 in these bands. This can be done stepwise to replace 802.11 systems. In Section 9.2 it is described that 802.16 is excellently suited to be operated in coexistence with 802.11 systems, able to control access to and occupancy of the medium, completely.

New bands of interest are, especially, located at lower frequencies. The WiMAX Forum advances the allocation of licensed and license-exempt spectrum in lower frequency bands. Sub 1 GHz frequency bands, especially vacant or analog TV spectrum, are expected to become available after television stations' transition from analog to digital broadcasting has finished. For example in the US, the Federal Communications Commission (FCC) has already allocated spectrum for BWA in the 700 MHz range (WiMax-Forum, 2005). The WG 802.22, as introduced in Section 13.6.2, is targeting at the realization of reusing TV band without interfering the incumbent radio systems.

7.3 History and Different Subgroups

7.3.1 History

Shortly after the ETSI Broadband Radio Access Networks (BRAN) had started its effort on standardization of BWA, the IEEE Standards Association (IEEE-SA) established the IEEE 802.16 working group in 1999. In the same year, both organizations agreed on a cooperation to harmonize their standards. IEEE 802.16 specified the initial standard for WMANs operating in frequencies above 11 GHz. It was published in 2002 as IEEE 802.16-2001 (IEEE, 2002a). ETSI BRAN published its standard High Performance Radio Access Network (HiperACCESS), which is based on the same single-carrier physical layer, in 2002 as well.

In response to market needs and significant industry interest, the standardization efforts proceeded by extending the standard to licensed and license-exempt bands ranging from 2 to 11 GHz. Three additional PHYs including one single-carrier mode, an OFDM and a mode based on OFDMA have been specified. The resulting amendment to the base standard was published in 2003 as IEEE 802.16a-2003 (IEEE, 2003c). In the same year, ETSI published its corresponding standard named HiperMAN that applies OFDM only (ETSI, 2003b, 2003c). It had been designed using the basic MAC of the IEEE 802.16 standard. HiperMAN had been developed in close cooperation with IEEE 802.16, such that the HiperMAN standard and the OFDM-based subset of the IEEE 802.16a-2003 standard will seamlessly interoperate.

Further on, task group 802.16c developed an amendment for 10–66 GHz system profiles to aid interoperability specifications. TG 802.16d had been set up to specify 2–11 GHz system profiles. Both groups were additionally reviewing the corresponding standards to correct errors and inconsistencies. Before 802.16d could finish, the work transitioned to a revision effort of IEEE standard 802.16-2001 as modified by task group “a” and “c”. The standard developed during this revision was published in 2004 as IEEE standard 802.16-2004 (IEEE, 2004a), replacing 802.16-2001, 802.16c-2002 and 802.16a-2003. From the date of publication, the new standard document supersedes the previous volumes.

7.3.2 IEEE 802.16-2004 – Base Document

The IEEE standard 802.16-2004, as it emerged out of the revision process, is considered as the root document. The basic MAC and four incompatible PHYs are specified. An ATM and a packet CS as well as system profiles are part of the standard. It covers all frequency bands below 66 GHz. The process of standardization has already been finished and only a corrigendum (IEEE 802.16-2004/Cor1) is going on that corrects errors, inconsistencies and ambiguities in the standard. It does not contain new material.

7.3.3 IEEE 802.16/Conformance

Conformance tests ensure that equipment is performing according to a general, independent and not to a personal interpretation of the standard. Together with successful interoperability, verified conformance leads to certification. The final goal of certification is the global interoperability of fully functional equipment and systems. The conformance task group works out the general, independent conformance specifications for BSs and SSs based upon the 802.16 air interfaces. The documents describe the Protocol Implementation Conformance Statement (PICS), the Test Suite Structure and Test Purposes (TSS&TP) and the Radio Conformance Tests (RCT) needed to develop a standardized Abstract Test Suite (ATS). The documents are published as IEEE standard 802.16/Conformance01 to 802.16/Conformance04.

7.3.4 IEEE 802.16.2 Coexistence

Coexistence in BWA Systems of IEEE 802.16 operating in licensed bands is standardized in the task Group IEEE 802.16.2. This task group provides a recommended practice for the design and coordinated deployment of BWA systems in order to control interference and facilitate coexistence among these systems and with other applicable systems that may be present. It analyzes appropriate coexistence scenarios and provides guidance for system design, deployment, coordination and frequency usage. It generally addresses licensed spectrum between 2 GHz and 66 GHz.

The IEEE standard 802.16.2-2004 was published in 2004 (IEEE, 2004c). 802.16.2 suggests threshold parameters, such as the distance between two interfering base stations, to assess the necessity for inter-operator coordination. These parameters are used to define guidelines for geographical spacing and frequency reuse. A concept of using Power Spectral Flux Density (PSFD) values is introduced in order to trigger different levels of initiatives taken by an operator to give notifications to other operators. Maximum values are defined for the PSFD that can be tolerated as a result from co-channel interference originating from adjacent operators. Frequency guard bands, recognition of cross-polarization differences, antenna angular discrimination, spatial location differences, the use of adaptive antennas and frequency assignment substitution are suggested to reduce the probability of interference.

7.3.5 IEEE 802.16e Mobility

In order to increase the market for BWA solutions task group 802.16e takes advantage of the ability of wireless media to support mobility. By providing enhancements to support SSs moving at vehicular speeds, the specified system fills the gap between high data rate wireless LANs and high mobility cellular systems. The amendment IEEE 802.16e-2005 for a combined fixed and mobile operation in licensed bands below 6 GHz was published in 2006 (IEEE, 2006b).

7.3.6 IEEE 802.16f/g/i Network Management

The purpose of the Network Management Task Group is to provide 802.16 equipment with procedures and services to enable interoperable and efficient management of network resources, mobility and spectrum. Within this group, the management plane behavior in 802.16 fixed and mobile devices is standardized. Amendment IEEE 802.16f-2005 defines a Management Information Base (MIB) for the MAC and PHY. It was published in 2005. IEEE 802.16g creates standardized procedures and interfaces for the management of 802.16 devices. Amendment IEEE 802.16i provides mobility enhancements to the MIB for the MAC and the PHY as well as associated management procedures.

7.3.7 IEEE 802.16h License Exempt

The IEEE 802.16h License-Exempt Task Group is developing improved mechanisms for enabling coexistence among license-exempt systems based on IEEE 802.16. The standardization efforts target for instance MAC enhancements and policies. Additionally, 802.16h also focuses on the coexistence of such systems with primary radio systems when vertically sharing spectrum as introduced in Section 13.3.3. The amendment is not limited to license-exempt bands, but operation may take place in all bands where 802.16-2004 is applicable.

A distributed architecture for radio resource management is suggested in IEEE (2005e), which enables communication and exchange of parameters between different networks formed by one 802.16 BS and its associated SSs. Each BS has a Distributed Radio Resource Management (DRRM) entity to execute the spectrum-sharing policies of 802.16h and to build up a database for sharing information related to actual and intended future usage of radio spectrum.

802.16h proposes a coexistence protocol to realize all functions required for coexistence as for example detecting the neighborhood topology, to register to the database or the negotiation for sharing radio spectrum. The DRRM uses the coexistence protocol to communicate with BSs and regional systems databases operated in license-exempt interacting with MAC or PHY.

7.3.8 IEEE 802.16j Mobile Multi-hop Relay Study Group

The Mobile Multi-hop Relay study group develops a relay mode for fixed and mobile terminals. 802.16 relay nodes shall either extend the coverage or enhance the throughput of a BWA network by leveraging the higher throughput over multi-hop links. Details to 802.16j can be found in Section 8.1.4.2.

Different from the existing mesh mode, the Mobile Multi-hop Relay group focuses on tree structures and thus excluding inter-SS/Mobile Station (MS) communication. The relay mode shall be compatible to the PMP mode and shall support OFDMA. Fixed, nomadic and mobile relays are taken into account, which may be owned and operated by the network operator, or which may be client terminals.

7.3.9 ETSI BRAN HiperACCESS and HiperMAN

The BRAN technical committee of the ETSI and IEEE working group 802.16 are cooperating closely with each other to harmonize the interoperability standards for BWA networks. ETSI BRAN standardized the HiperACCESS as well as the HiperMAN system.

HIPERACCESS is a broadband multimedia fixed wireless access system to allow for a flexible and competitive alternative to wired access networks. HIPERACCESS is based on a single-carrier transmission and is targeting high frequency bands, especially the 40.5–43.5 GHz band. The published specifications include PHY and Data Link Control (DLC) layer specifications, convergence layers for the cell and packet-based core networks, and conformance test specifications.

HIPERMAN is an OFDM-based BWA system operating at radio frequencies between 2 GHz and 11 GHz. Due to the close cooperation HIPERMAN and the OFDM-based subset of IEEE 802.16a-2003 have the same MAC and PHY. Thus, they will interoperate seamlessly. The published specifications include the PHY and DLC layer specifications, electromagnetic compatibility and radio spectrum matters, and conformance test specifications.

7.3.10 WiMAX Forum

The WiMAX Forum is an industry-led, non-profit corporation formed to promote and certify compatibility and interoperability of broadband wireless products. Formed in 2001, member companies support the industry-wide acceptance of the IEEE 802.16 and ETSI HiperMAN standards in order to accelerate the introduction of these systems into the marketplace. By August 2004, more than 320 companies had joined the WiMAX Forum, including equipment manufacturers, operators, system integrators, silicon and component makers, and application providers.

7.3.11 Wireless Broadband (WiBro)

After the allocation of 100 MHz of spectrum in the 2.3 GHz band in 2002, the Telecommunications Technology Association (TTA) of Korea standardized the national Wireless Broadband (WiBro) system in 2004. Three licensed 27 MHz bands have been assigned to three different operators, each band containing three 9 MHz channels. Thus, a frequency reuse factor of one allows for a cell structure with one to three sectors.

The WiBro system shall support mobile personal SSs, such as cell phones, Personal Digital Assistants (PDAs), handheld PCs and laptop PCs moving with speeds up to 60 km/h. The targeted throughput varies between a minimum of 128 kbit/s/512 kbit/s and a maximum of 1 Mbit/s/3 Mbit/s for uplink/downlink, respectively. The WiBro specification is a subset of the standards IEEE 802.16-2004, 802.16e and 802.16-2004/Cor1. It is based on the OFDMA PHY using 1048 subcarriers in TDD mode. Harmonization activities are currently working towards compatibility between the standards WiBro and WiMAX.

7.4 Physical Layer

The initial version of the standard (IEEE 802.16-2001) specified only one single-carrier PHY. It targets frequency bands between 10 and 66 GHz, in which LOS communication is mandatory due to the propagation characteristics. The current standard (IEEE 802.16-2004), which supersedes all previous versions, specifies three additional PHY techniques. They were designed for frequency bands below 11 GHz that allow for NLOS links. The new modes include a single-carrier, an OFDM and OFDMA based mode.

The OFDM-based transmission mode has been standardized in close cooperation with the ETSI standard HiperMAN. Both OFDM-based protocols shall comply with each other in order to form the basis for the WiMAX certified technology. In the following section, the OFDM based PHY layer of 802.16 is presented in detail.

7.4.1 Orthogonal Frequency Division Multiplexing in 802.16

IEEE 802.16 specifies a multi-carrier PHY mode based on OFDM. As outlined in Section 2.3.5, an OFDM system splits the available system bandwidth into orthogonal subcarriers. By reducing the bandwidth per subcarrier by a factor of N, the symbol duration becomes N times longer. Thus, the symbol becomes relatively robust to delay spread caused by multipath propagation. The introduction of a guard time, called cyclic prefix, nearly eliminates the influence of delay spread, if well designed in its length. Nevertheless, OFDM systems are highly sensitive to frequency offset errors, which destroy the orthogonality of subcarriers. Furthermore, the addition of a multitude of harmonic sinusoidal waves results in a high peak to average power ratio. Thus, the transceiver's power amplifier has to be linear over a wide range of power.

In OFDM transmission, each single subcarrier band can be seen as a narrowband flat fading channel. This allows for a relatively simple equalizer at the receiver. For instance, equalization might be performed by means of only one complex division per subcarrier.

The IEEE 802.16 OFDM PHY was designed for both LOS and NLOS operation in frequency bands below 11 GHz. It can operate in licensed and license-exempt spectrum. Under license-exempt operation the system makes use of transmit spectral masks and channelization schemes specified for the unlicensed spectrum. These special features are captioned under the name Wireless High-speed Unlicensed Metropolitan Area Network (WirelessHUMAN). Link distances, i.e., cell sizes, vary strongly based on the used frequency bands, the allowed transmit power, the environment, propagation conditions and antenna gain. The system targets distances between 2 km and 4 km for NLOS and up to 15 km for LOS.

For the IEEE 802.16 OFDM PHY mode, TDD and Frequency Division Duplex (FDD) variants are defined. Channel bandwidths vary from 1.25 to 28 MHz. The actual sampling frequency depends on the chosen channel bandwidth and is slightly higher than the nominal bandwidth (see Figure 7.3). The OFDM PHY is based on a 256-point Fast Fourier Transform (FFT), resulting in 256 subcarriers. Thus, the subcarrier spacing is calculated to one 256th of the sampling frequency. Eight subcarriers are used for pilot tones, 55 subcarriers are used as guard carriers. On the guard carriers and on the Direct Current (DC) subcarrier nothing is transmitted at all. Due to the even number of subcarriers, the upmost subcarrier spacing remains unused and can be seen as the 28th upper guard carrier. The remaining 192 data carriers are used for data transmission.

Figure 7.4 shows the basic modules of an IEEE 802.16 transmitter–receiver chain. In the following the different modules and their corresponding functionality are outlined. The very first block "MAC Source" represents the MAC layer. It generates the current MAC frame. The frame is converted into a bit stream, which is given to the first block of the PHY, i.e., the Randomizer. The functions of the modulation, pilot insertion, preamble insertion and Inverse FFT (IFFT) module are introduced in Section 2.3.5.

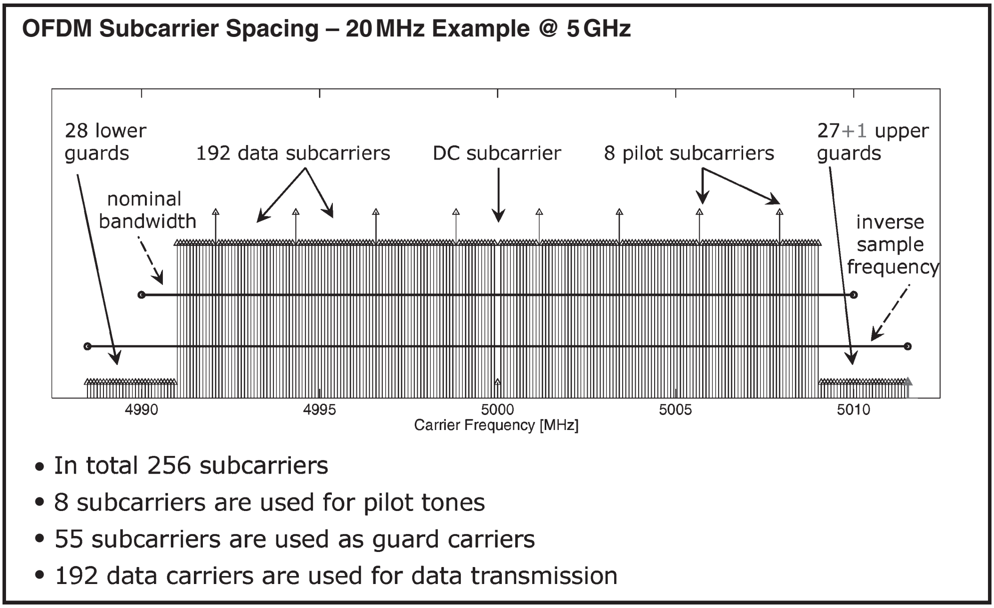


Figure 7.3 OFDM subcarrier spacing 20 MHz example with 5 GHz center frequency.

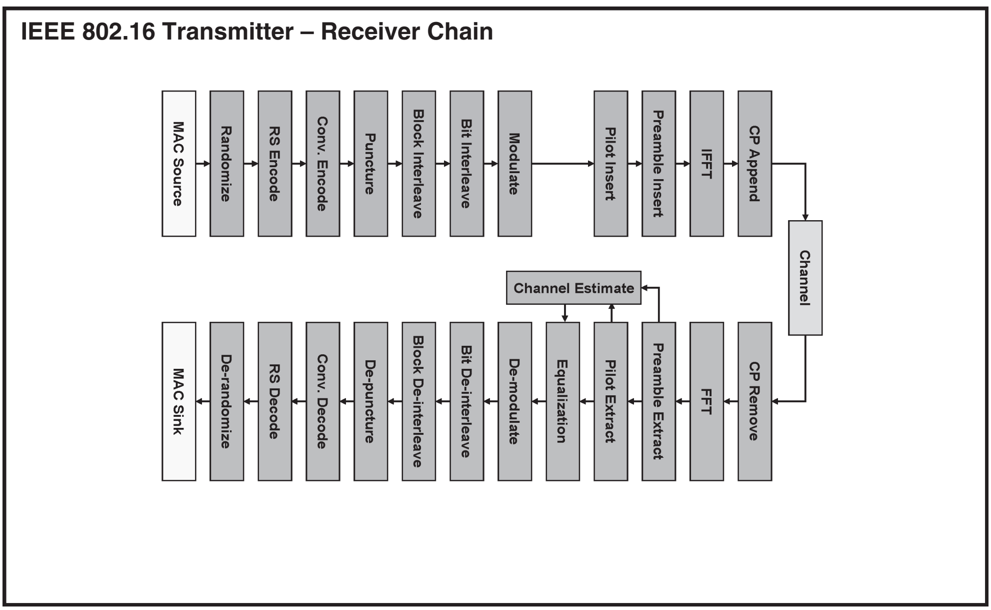


Figure 7.4 Basic modules of the IEEE 802.16 transmitter–receiver chain.

7.4.1.1 Randomizer

The Randomizer (or Scrambler) adds a pseudo-random binary sequence to the original Downlink (DL) and Uplink (UL) bit stream. This random bit sequence avoids long rows of zeros or ones for a better coding performance of the following Forward Error Correction (FEC) modules. The addition is individually done on each burst. The pseudo-random binary sequence generator is initialized with the Base Station ID (BSID), the modulation and coding scheme used for the burst and the frame number. At the end of each burst, the Randomizer adds padding bytes (0xFF) to increase the size of a burst up to an OFDM symbol boundary. The module appends one tail byte of zeros (0x00) for the following convolutional coder in order to start and end in the well-known and predefined all-zeros state.

7.4.1.2 Forward Error Correction

The principle of FEC is described in general in Section 2.5.1. In 802.16, the FEC scheme consists of the concatenation of a Reed–Solomon (RS) outer code and a Convolutional inner Code (CC). A concatenated code has been specified in order to achieve a superior error correction performance for a given complexity compared to a single but longer code. The RS coder corrects burst errors at byte level. The actually applied code depends on the PHY mode and is derived, i.e., shortened, from a systematic RS ($N = 255, K = 239, T = 8$) code using the Galois field (28). N is the number of overall bytes after encoding, K is the number of data bytes before encoding, and T stands for the number of data bytes that can be corrected. The RS code is particularly useful for OFDM links in the presence of multipath propagation, since bursty errors might frequently occur. The CC corrects independent bit errors. A CC code can easily and efficiently be decoded, e.g., using the Viterbi algorithm. The CC decoder can benefit from softbit input generated by the de-modulation and de-puncturing blocks. The concatenation of both codes is made rate-compatible by the following puncturing module. Based on four puncturing patterns bits are removed to realize different code rates. The punctured bits are replaced at the receiver side either as hard (either 1 or 0) or soft bits (no information, i.e., a value of 0.5). The support of Block Turbo Coding (BTC) and Convolutional Turbo Coding (CTC) is optional. The table (a) depicted in Figure 2.9 shows the resulting data block sizes for each modulation and coding scheme (PHY mode).

7.4.1.3 Interleaving

The Interleaver is composed of a block and a bit interleaver. The block interleaver maps adjacent coded bits onto non-adjacent subcarriers to overcome burst errors. This can be done by writing the bit stream into 12 rows of an array row-wise and by reading the bits out column-wise. Alternatively, a deterministic formula can be applied to modify the indices of the bits. The bit interleaver maps adjacent coded bits alternately onto less and more significant bits of the constellation to avoid long runs of unreliable bits. The Quadrature Amplitude Modulation (QAM) is effected by this as a single bit, e.g., using 16-QAM, can either switch once (more significant bit) or twice (less significant bit) while moving through the constellation diagram (see bits b_0 and b_1 in Figure 2.8).

7.5 Medium Access Control Layer

The scope of the IEEE 802.16 standard comprises the data and control plane of the MAC and the PHY as illustrated in Figure 7.2. The MAC includes a service-specific convergence sublayer that interfaces higher layers. The MAC CPS realizes key functions and the security sublayer is located below the MAC CPS. The management plane is specified in three IEEE network management standard amendments (IEEE 802.16f, g, i) for the Fixed as well as the Mobile Management Information Base (MIB) and for Procedures and Services as described in Section 7.3.6.

7.5.1 Service-Specific Convergence Sublayer

The service-specific CS provides any transformation or mapping of external network data, received through the CS Service Access Point (SAP). This includes the classification of external network SDUs and (if required) the processing of SDUs. Classifying incoming SDUs means to associate them with the proper connection identified by the Connection Identifier (CID) which is shown in Figure 7.5. Since a CID is associated with a certain level of QoS, the association of an SDU to a CID facilitates the delivery with the corresponding QoS constraints. The CS processes higher layer SDUs to suppress unused higher layer protocol information. After classification and PHS, the SDU is delivered to the corresponding MAC CPS SAP. At the receiving CS entity, the suppressed header is reconstructed before it is handed over to the higher layer protocol via the SAP. Since PHS is an optional feature, incoming SDUs can also be delivered without any modifications. The standard provides two CS specifications, an ATM and a Packet CS.

7.5.1.1 Packet Convergence Sublayer

The Packet CS is used for packet-based higher layer protocols such as Internet Protocol (IP), Point-to-Point Protocol (PPP) and IEEE 802.3 (Ethernet). Classification of SDUs is based on classifiers that consist of a reference to a CID, a classifier priority and a set of protocol-specific

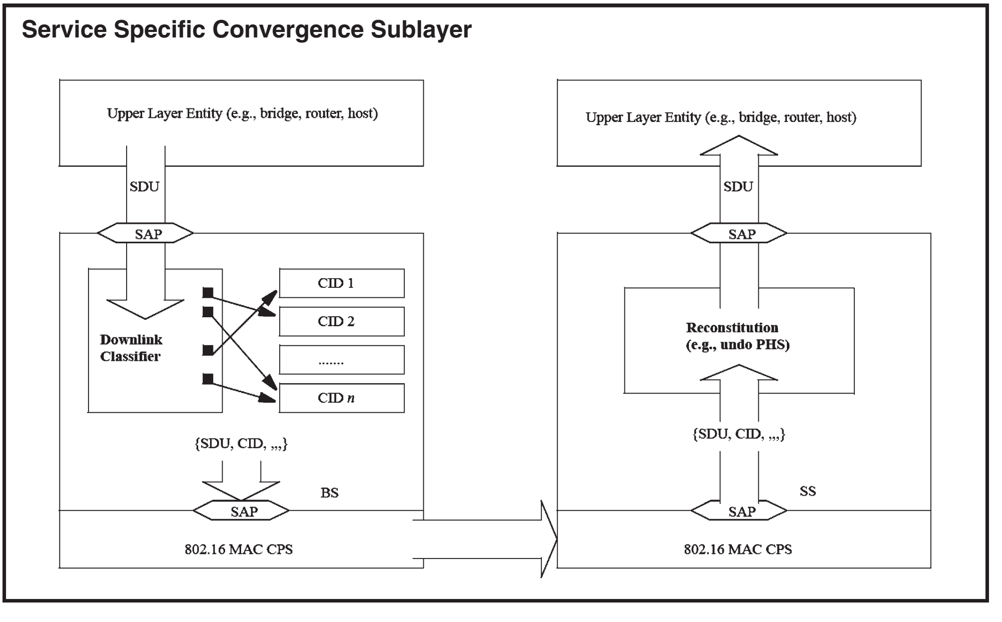


Figure 7.5 Service specific convergence sublayer [IEEE, 2004a].

matching criteria. Characteristic protocol entries are used as matching criteria as for instance the IP or Ethernet source/destination address, protocol source/destination port range, IP type of service/differentiated services codepoint, or the IEEE 802.1D-1998 user priority. If several classifier rules match with an incoming SDU, the classifier priority specifies which rule is to be applied.

Since the Packet CS handles various higher layer protocols, various header entries might have to be suppressed and reconstructed. Therefore, the optional PHS functionality defines a mechanism to adaptively suppress specific bytes of an unspecified SDU. All bytes of a specific region (PHS Field) will be suppressed by the sending entity unless they are marked by the PHS Mask. Compressed Packet CS SDUs are prefixed with an 8-bit PHS Index. The receiving entity reassembles the original SDU by adding the bytes that are stored in the PHS Field associated to the CID.

7.5.1.2 ATM Convergence Sublayer

Classification within the ATM CS depends on the switching mode of the corresponding ATM connection. An ATM cell is uniquely identified by a pair of values, namely the Virtual Path Identifier (VPI) and Virtual Channel Identifier (VCI). If virtual path switched mode is applied, ATM cells having a certain VPI are mapped to the associated MAC connection. Thus, ATM cells with different VCI are transmitted on one single MAC connection. If the ATM connection operates in VCI switched mode, the combination of VCI and VPI is mapped to the CID of the connection on which it is transported. Thus, in the ATM CS the classifier's matching criteria is either the combination of VCI and VPI or the VPI only.

The optional PHS mode also depends on the switching mode of the ATM connection. In virtual path switched mode, the 5-byte ATM header is compressed to a 3-byte header containing among others the VCI. Thus, the VPI is suppressed. In virtual channel switched mode, both the VPI and the VCI are suppressed resulting in a compressed ATM header of only 1 byte. The receiving ATM CS entity reconstructs the original ATM header by adding the corresponding VPI and/or VCI and by recalculating the header checksum.

7.5.2 MAC Common Part Sublayer

The MAC CPS provides system access, bandwidth allocation, connection establishment and connection maintenance. The MAC CPS receives from the convergence sublayer data classified to particular CIDs. QoS is applied to the transmission and scheduling of data over the PHY.

IEEE 802.16 is optimized for PMP configurations where several SSs are associated with a central BS. As an optional feature, the standard allows for a flexible mesh deployment where direct communication between stations is possible. Since the mesh frame structure is not compatible with the PMP frame, the mesh deployment is especially foreseen for wireless backhaul networks based on IEEE 802.16 as described in Section 8.1.4. Additional to PMP, an amendment for multi-hop communication in tree-based deployments is currently under specification in IEEE 802.16j as also introduced in Section 8.1.4. The introduction of relay stations, which decode and forward data, extends the coverage area of a BS or increases the achievable capacity within a given area. The multi-hop amendment requires being PMP-compliant so that legacy SS will be able to participate in relay enhanced networks.

Section 7.4 outlines the specification of four different IEEE 802.16 PHYs, while one single MAC is controlling the access to the medium. Hence, the MAC protocol is PHY independent in general, but some mechanisms are PHY specific. PHY-specific parts mainly focus on the MAC frame structure and the corresponding signaling messages. Wherever necessary, the OFDM PHY-specific part of the MAC protocol is described in the following when introducing the 802.16 MAC.

7.5.2.1 Duplex Modes

Two duplexing techniques are specified for 802.16, namely TDD and FDD, which are both introduced in Section 2.2. In short, FDD operation implies paired frequency bands, which are typically allocated in licensed spectrum. Thus, DL and UL operate on separate frequency channels. The asymmetry between DL and UL is predefined by the spectrum allocation and is therefore static. Thus, capacity cannot be shifted during operation between DL and UL. In full-duplex FDD, stations simultaneously receive and transmit on both channels. Consequently, two Radio Frequency (RF) filters, two oscillators and two synthesizers are required. On the one hand, the increased need for components makes FDD devices more power consuming and more expensive. On the other hand, the MAC software can be less complex since DL and UL are not strictly synchronized (Bisla *et al.*, 2004). In order to avoid expensive hardware a Half-Duplex FDD (HFDD) mode is supported. In HFDD DL and UL are still operating on separate frequency channels, but stations do not simultaneously transmit and receive. The resulting radio complexity is comparable to the complexity of TDD devices. A desirable network deployment, in which the BSs operate in FDD and the SSs in HFDD, combines the possibility to utilize both channels simultaneously with competitive user devices.

TDD overcomes the static asymmetry of FDD by sharing a common frequency channel for DL and UL transmission in the time domain. Hence, capacity can be dynamically shifted in adapting the switching point between DL and UL in time. Since stations do not receive and transmit at the same time, only a single RF filter, one oscillator and one synthesizer are required. This results in cost- and power-efficient devices, but the MAC scheduler of the BS tends to be more complicated since it has to synchronize many stations' time slots in both DL and UL direction (Bisla *et al.*, 2004). In order to switch between receive and transmit phases turnaround gaps, i.e., guard intervals, have to be introduced between both phases. Assuming low mobile SSs, the reciprocity of the radio channel can be exploited in TDD systems, because the transmitter can take advantage of the channel knowledge available at the receiver. In license-exempt spectrum, the IEEE 802.16 TDD mode is mandatory.

7.5.2.2 Frame Structure

IEEE 802.16 supports a frame-based transmission. The MAC frame duration may vary between 2.5 to 20 ms although a highly dynamic variation is not foreseen during normal operation as a change of frame duration forces all SSs to resynchronize. Figure 7.6 illustrates the frame structure of an OFDM-based MAC layer operating in TDD mode. Each frame consists of a DL subframe and a UL subframe. The DL subframe always precedes the UL subframe. Subframes are separated by Receive/Transmit Transition Gaps (RTG) and Transmit/Receive Transition Gaps (TTG), respectively. The gaps allow stations to switch their PHY processors between transmit and receive states. The partitioning between DL and UL may vary dynamically to efficiently handle an asymmetric traffic load. In the FDD mode, similar DL and UL subframes are present, but they are located on separated frequency channels.

The **DL subframe** consists of only one PHY transmission starting with a preamble. The long preamble at the beginning of each frame is composed of two OFDM symbols. It is used by the SSs

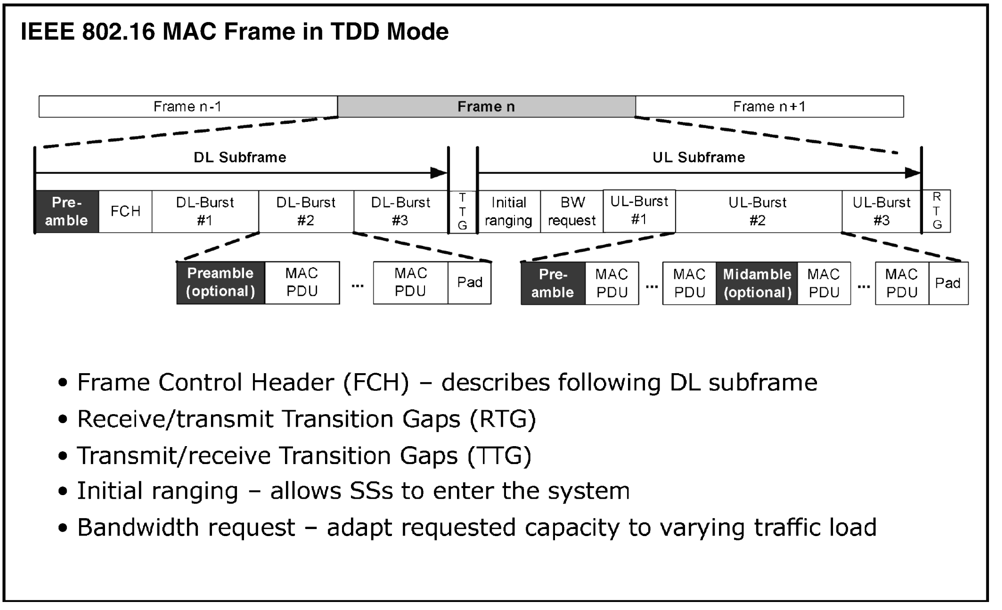


Figure 7.6 IEEE 802.16 medium access control frame in TDD mode.

for time and frequency synchronization. The following Frame Control Header (FCH) occupies one OFDM symbol and describes the following DL subframe. The FCH is followed by one or multiple DL bursts, which should be ordered by decreasing modulation robustness. While the burst with the most robust modulation, e.g., BPSK, is transmitted first, the last burst is modulated using the highest PHY mode that is supported, e.g., 64 QAM. This ordering should simplify the PHY and improves the synchronization of SSs. Each DL burst consists of MAC Protocol Data Units (PDUs) scheduled for DL transmission. Optionally, a DL burst might start with a short preamble. A short preamble has the duration of one OFDM symbol and allows for an enhanced synchronization and channel estimation at the receiving SSs. MAC PDUs transmitted within the same DL burst are all encoded and modulated using the same PHY mode. Nevertheless, PDUs of one DL burst might be associated to different connections and/or SSs. In DL as well as in UL direction the burst length is an integer number of the OFDM symbol length so that burst and OFDM symbol boundaries match each other. In order to end up on OFDM symbol boundaries, padding bytes (0xFF) are added at the end of each burst.

The **UL subframe** consists of contention intervals and one or multiple PHY transmissions, each transmitted from a different SS. A contention interval may be scheduled for initial ranging and another one for bandwidth request purposes. In order to reduce the probability of collisions, the contention intervals are slotted in the time domain. Initial ranging slots allow SSs to enter the system by requesting the basic management CIDs, by adjusting its power level and frequency offset and by receiving its timing offset. Since the SSs that are accessing the initial ranging contention slots do not have their timing offset yet, the appropriate slot duration depends on the intended cell radius. The slot duration should be at least long enough to transmit a long preamble followed by an initial ranging management message, which is coded and modulated with the most robust PHY mode (BPSK 1/2), plus the round trip delay. SSs access the second type of contention slots in order to transmit bandwidth requests. Hence, SSs can adapt the requested UL capacity to their varying traffic loads. The slot duration of bandwidth request slots may be shorter than the duration of initial ranging slots. Taking the right timing offset into account, only a short preamble plus a bandwidth request header is transmitted. Numerous mechanisms to allocate UL resources more efficiently are foreseen in 802.16. These mechanisms include, for instance, dedicated bandwidth request slots, piggybacked requests or unsolicited resource grants. Besides contention phases, the UL subframe contains one or multiple UL bursts. Each burst starts with a short preamble. For better synchronization and channel estimation at the receiving BS, optional midambles might be periodically included in the UL burst. Like in the DL, all PDUs transmitted within one single burst are coded and modulated with the same PHY mode. A UL burst is transmitted by one single SS. Thus, PDUs are associated to connections of only one SS. Analog to the DL, padding bytes are added at the end of each UL burst to end up on OFDM symbol boundaries.

7.5.2.3 Frame Control

Since the MAC frame is dynamically composed out of various bursts, the BS broadcasts MAC management messages to describe the current realization of the MAC frame. SSs receive these messages and gather the information needed to receive and transmit PDUs. Figure 7.7 shows the basic MAC management messages used to specify the current MAC frame realization. The management messages contain time references to the corresponding elements of the MAC frame. The arrows in Figure 7.7 indicate these references, such as burst start times and burst lengths.

The FCH and the DL MAP define the access to the DL subframe. The FCH is composed of the Downlink Frame Prefix (DLFP). The DLFP specifies up to four DL bursts, which are directly following the FCH. The FCH, whose length is one OFDM symbol, is transmitted with the mandatory modulation BPSK and with a code rate of 1/2. Taking the zero tail byte for

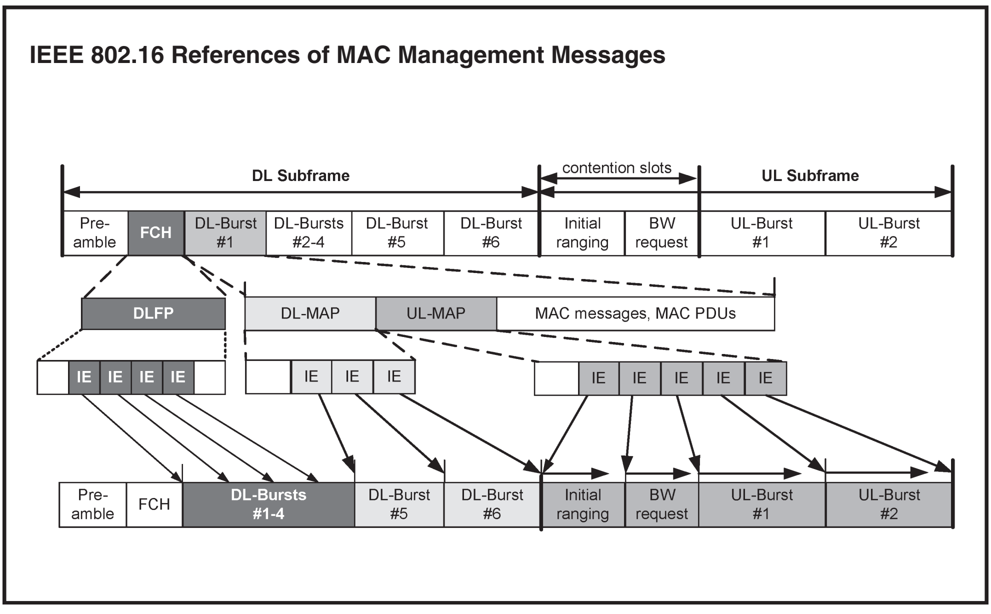


Figure 7.7 IEEE 802.16 references of MAC management messages. Reproduced by permission of © 2006 IEEE[1].

the CC into account, the DLFP has a size of 11 bytes. The DLFP contains four Information Elements (IEs), each specifying one DL burst. The DLFP IE contains the length and the PHY mode of the corresponding DL burst. The PHY mode is not given directly, but as a reference to a certain burst profile, which is named Downlink Interval Usage Code (DIUC). A burst profile is a detailed description of the corresponding PHY mode. It is contained in the Downlink Channel Descriptor (DCD). The DCD defines the characteristic of the physical DL channel. It is occasionally broadcasted by the BS. The burst profile details the specific coding algorithm (RS+CC, BTC or CTC), the code rate and the modulation scheme. Since the start time of a burst is not explicitly included in the IE, it has to be calculated as the sum of all burst lengths of the preceding bursts. The IE additionally informs whether the optional preamble is transmitted at the beginning of the DL burst or not. If the DL subframe consists of four or fewer bursts, the DLFP is sufficient to specify the entire subframe. However, if the DL subframe is made up of more than four bursts, an additional DL MAP has to be transmitted that specifies the remaining bursts.

The very first DL burst contains the broadcast MAC control messages, i.e., DL and UL MAP as well as the DL and the UL channel descriptor (DCD, Uplink Channel Descriptor (UCD)). Analog to the DCD, the UCD defines the characteristic of the physical UL channel. If it is applicable to all SSs within the intended coverage area of the BS, the first DL burst may be coded with a more efficient PHY mode than BPSK 1/2. This reduces the overhead due to signaling, but is also reduces the potential coverage area of the BS.

If present, the **DL MAP** is included in the very first PDU of the first DL burst, thus, it immediately follows the FCH. Among others, the DL MAP contains one IE for each burst of the DL subframe that has not been described by the FCH yet. The information included in the DL MAP is only relevant for the DL subframe of the current MAC frame. The left PDU in Figure 7.8 shows that the DL MAP IE is made up of four entries, the CID of the addressee of the

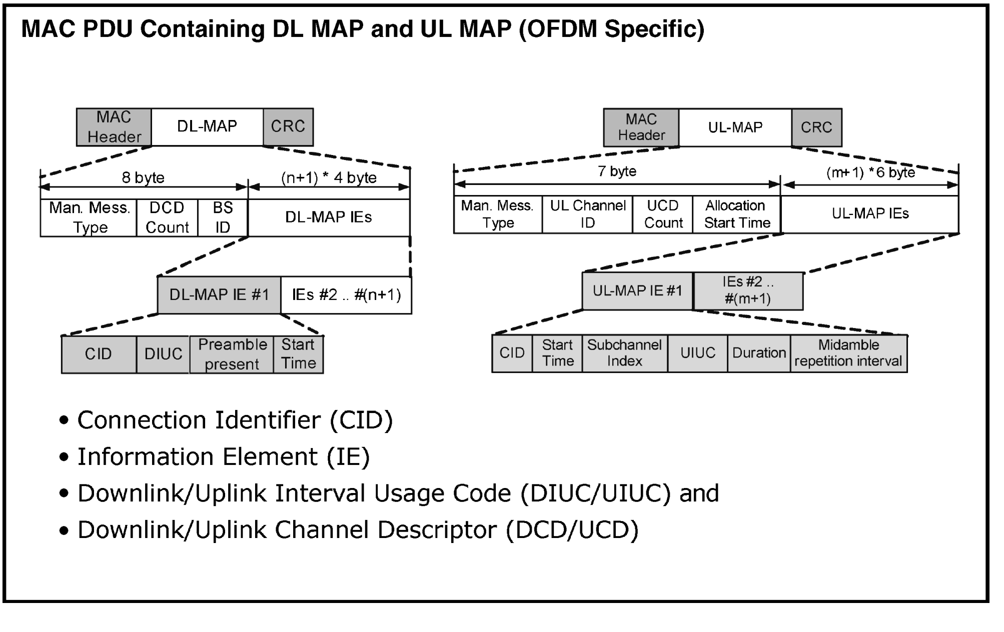


Figure 7.8 MAC PDU containing DL MAP and UL MAP respectively (OFDM specific).

burst, the DIUC, the Start Time of the burst and a bit to indicate whether the optional preamble is prepended to the DL burst. If all PDUs of one burst are to be transmitted to the same SS, the burst is directly addressed to this SS. No other SS has to decode this particular burst. If the specified DL burst contains MAC PDUs for several SSs, the CID of the corresponding DL MAP IE is set to a multicast or the broadcast CID. All addressed SSs have to start decoding the burst at the specified start time with the given PHY mode. The information to which connection the received MAC PDUs are associated can be taken from the MAC header of the particular PDU. Like in the DLFP, the DIUC is a reference to the burst profile used for the corresponding burst. The start time of the following DL burst is automatically taken as the end of the current one. That means the burst duration is implicitly given by subtracting the burst start time from the start time of the following burst. This calculation strictly relies on a sequential nature of bursts. The last IE indicates the end of the MAP and refers to the end of the subframe. Besides the start time, which is set to the end of the DL subframe, this last IE is empty.

Optionally, the DL MAP IE might be extended so that it contains additional information for specific purposes as for instance the Channel Measurement IE, which is used by a BS to get a channel measurement report from a SS. If bursts of the DL subframe are transmitted with Space-Time Coding (STC) or Advanced Antenna System (AAS) techniques, they can be identified by means of the AAS IE and the STC IE, respectively. A Physical Modifier IE indicates a cyclic shift of the following optional DL preambles. In this way the reception of concurrent DL bursts sent by the BS in SDMA mode might be enhanced. In this case, a Concurrent Transmission IE is used to specify the burst duration. The explicit indication of the duration overcomes the restriction of the sequential nature of DL bursts. Knowing the start time and the duration, the BS is able to flexibly arrange concurrent DL bursts for SDMA (Hoymann, 2006).

The **UL MAP**, which is shown in the right-hand MAC PDU of Figure 7.8, assigns capacity to the UL subframe. Like the DL MAP, the UL MAP contains IEs to specify the bursts. Depending on the effective start time of the corresponding UL subframe (included in the UL MAP as Allocation Start Time), the UL MAP is relevant for the UL subframe of the current or of the following MAC frame. Each UL MAP IE specifies one UL burst including the contention intervals for initial ranging as well as for bandwidth request.

The UL MAP IE consists of six elements as illustrated in Figure 7.8. The CID is the unique address of the SS that is scheduled for the particular UL burst. The Start Time and the Duration of the corresponding UL burst are additionally defined. Hence, unlike the regular DL subframe the UL subframe does not rely on a sequential structure of bursts. UL bursts do not implicitly end at the beginning of the following one. Thus, idle times can be included in between two bursts by delaying the start of the second burst. For SDMA operation, UL bursts might be scheduled concurrently by specifying the same start time for more than one burst. The Subchannel Index is used to indicate which OFDM subcarriers shall be used for the transmission. Subchannelization can only be used if the SS is able to transmit on a subchannel basis. This capability has to be negotiated during the network entry. Analog to the DIUC, the Uplink Interval Usage Code (UIUC) is a reference to a burst profile. The UCD message, which contains the burst profiles, is occasionally broadcasted. The profile details the coding algorithm, the code rate and the modulation scheme that should be used by the SS to transmit the corresponding UL burst. In addition to the mandatory short preamble at the beginning of each UL burst, midambles might be included in the UL burst on a periodic basis. The request to include midambles is indicated by the Midamble Repetition Interval. Together with the preamble, the midambles allow for an enhanced synchronization and channel estimation at the BS.

Like the DL MAP IE, the UL MAP IE can be extended for specific purposes, such as the request to change the SS's transmit power, the switch to AAS enabled traffic, or the indication of the cyclic shift of preambles and midambles. The cyclic shift allows for an enhanced joint detection of SSs during the UL subframe.

The above-described FCH and the MAPs are PHY layer independent, while the detailed inner structure of the DLFP and the IEs are PHY specific.

7.5.2.4 Packet Data Unit Format

MAC PDUs carry user data, i.e., SDUs from the convergence sublayer, as well as MAC management messages. PDUs consist of a fixed-length MAC header, a variable-length payload and a 32-bit Cyclic Redundancy Check (CRC). The CRC is optional for PDUs that contain user data but it is mandatory for PDUs that carry MAC management messages. The size of the MAC header is six bytes. Thus, the minimum size of a PDU consisting only of the header is six bytes. The Length field, which is included in the header, specifies the entire length of the PDU including header, payload and CRC. Since the Length field is encoded with 11 bits, the maximum possible PDU length is 2047 bytes. Thus, the variable PDU payload may range up to 2041 bytes. This enables the MAC protocol to tunnel various higher layer traffic types without knowledge of the format of those messages. The structure of a MAC PDU is shown in Figure 7.9.

Two different header formats are defined, the Generic MAC and the Bandwidth Request Header format. The generic header is used for any PDU that contains user data or management messages. It is composed of eight fields as depicted in Figure 7.9: The Header Type identifies the generic header format. The Encryption Control field indicates if the payload is encrypted. If it is, the entry Encryption Key Sequence (EKS) field gives a reference to the encryption key that has been used. The Type field is a set of bits that indicates subheaders and special PDU payloads, such as

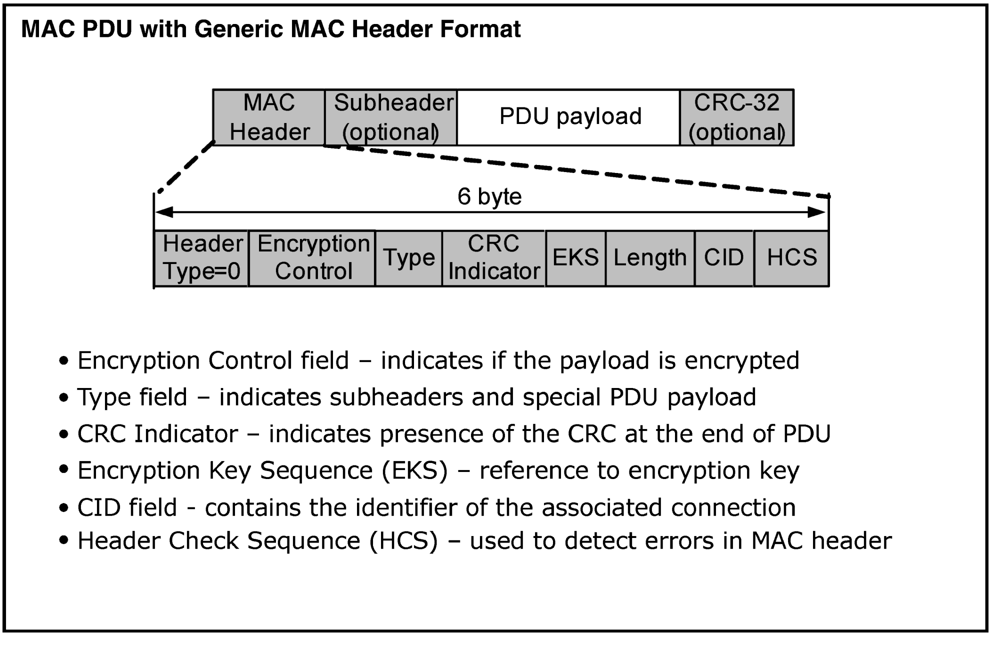


Figure 7.9 MAC PDU with generic MAC header format.

an ARQ feedback payload or Mesh, Packing and Fragmentation subheaders. Some subheaders directly follow the generic header, while others might occur within the payload. The presence of the CRC at the end of the PDU is denoted by the CRC Indicator. The CID field contains the identifier of the associated connection. The Header Check Sequence (HCS) is used to detect errors in the MAC header.

The bandwidth request header is only used by SSs to request bandwidth in the UL subframe. The corresponding header format contains entries to identify the connection, for which bandwidth is requested and to specify the amount of data that is waiting for transmission. The type of the request can be either incremental or aggregate.

7.5.2.5 Fragmentation and Packing

User data can be directly encapsulated in a MAC PDU payload, i.e., a single MAC SDU becomes the payload. For flexibility and/or efficiency reasons, SDUs may be fragmented and/or packed. These features can be enabled on a connection basis during the connection setup.

Fragmentation is the process of dividing SDUs into several fragments with the aim of allowing flexible and efficient use of available bandwidth. Each fragment is prefixed with a Fragmentation Subheader (FSH) and encapsulated into a MAC PDU. Since an SDU that exceeds the maximum PDU payload length of 2041 bytes has to be fragmented, this feature is mandatory. The reassembly of SDUs at the receiver side is performed by means of the FSH. The FSH contains a Fragment Sequence Number (FSN) and the position of the fragment within the SDU. The position is encoded in the Fragment Control (FC) field, whose entry marks the fragment as first, middle (in between), or last fragment.

Packing is the process of encapsulating multiple SDUs into one single PDU in order to reduce the overhead due to small SDUs. If packing is enabled for a connection, the transmitting side has full discretion whether or not to pack. Packing is an optional feature, but unpacking is a mandatory one. In order to pack variable-length SDUs, each SDU is prefixed with a Packing Subheader (PSH), which contains the SDU length. At the receiver, the SDU can be recovered by means of the PSH. If a connection carries only fixed-length SDUs, the fixed SDU length can be signaled during the connection setup. In this way, fixed-length SDUs can be packed without further subheaders and thus no additional overhead occurs. Simultaneous fragmentation and packing is also possible. For this purpose, the PSH contains the FC and the FSN. Figure 7.10 shows a MAC PDU that contains the last two fragments of an SDU and an unfragmented SDU. Depending on the station's capability, the length of the FSN field in the FSH and PSH is either 3 or 11 bits, so that the FSN is counted modulo 8 or modulo 2048. The length of the FSHs and PSHs varies accordingly. For Automatic Repeat Request (ARQ) enabled connections, the FSN is 11 bits.

7.5.2.6 Automatic Repeat Request

The IEEE 802.16 ARQ mechanism is an optional part of the MAC protocol. By means of the connection setup procedure, ARQ functionality can be enabled on a connection basis. Four different ARQ mechanisms with different types of Acknowledgments (ACKs) are defined in the standard: a selective ARQ, a cumulative ARQ, a cumulative with selective ARQ, and a cumulative with block sequence ACK ARQ. These basic ARQ protocols are introduced in Section 2.5.2.

All ARQ mechanisms are working on numbered ARQ data packets in 802.16 referred to as ARQ blocks. For this purpose, the SDU, which is transmitted on an ARQ-enabled connection, is logically partitioned into fixed-size ARQ blocks. The SDU is prefixed with a fragmentation

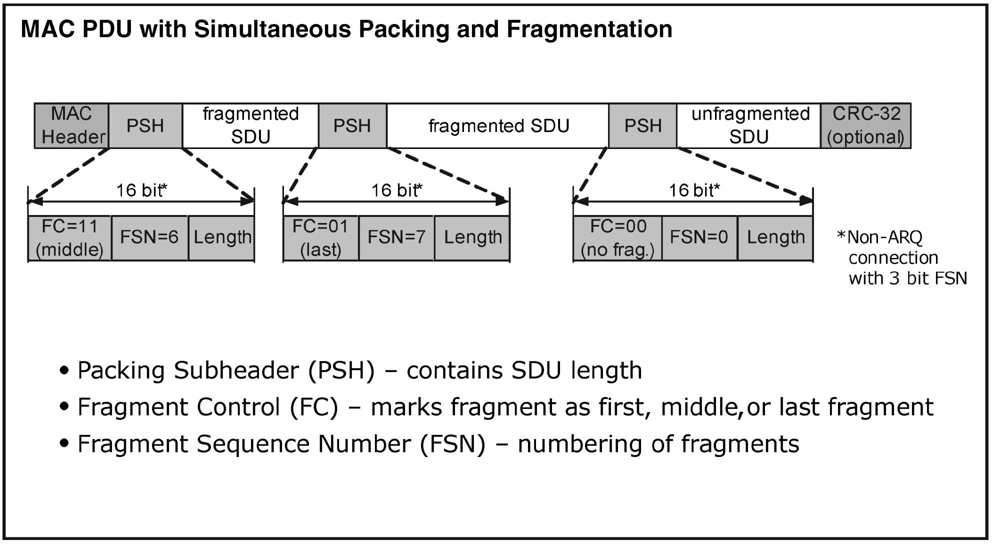


Figure 7.10 MAC PDU with Simultaneous Packing and Fragmentation.

or a packing subheader. On ARQ-enabled connections, the FSN, which is contained in the subheader, becomes the Block Sequence Number (BSN). The BSN is set to the number of the first logical ARQ block of the SDU. Then, the compound of the subheader and the SDU is ready for transmission or retransmission. It is also possible to fragment SDUs, which have been partitioned into ARQ blocks, but fragmentation shall only occur on ARQ block boundaries.

ARQ feedback, i.e., positive or negative ACK, is transmitted either as a standalone MAC management message or piggybacked. The format of the ARQ feedback depends on the type of the ARQ in use. The feedback message of the cumulative ARQ, which has the smallest size, contains the BSN of the last successfully received ARQ block. The feedback of the other ARQ types contains one or more bitmaps to acknowledge ARQ blocks. The message of the selective ARQ contains up to four 16-bit bitmaps that selectively acknowledge ARQ blocks. The message format of the selective with cumulative ARQ equals the one of the selective ARQ. Its feedback additionally acknowledges received blocks cumulatively, which are not covered by the bitmaps. The ACK format of the block sequence ARQ type is optimized to acknowledge entire sequences of ARQ blocks at once.

7.5.2.7 Connection Identifier

The 802.16 standard defines independent connections for the transport of user data, i.e., SDUs from the convergence sublayer, and for the transport of signaling data, i.e., MAC management messages. In general, a MAC connection is a logical mapping between MAC peer entities. A CID uniquely identifies a connection.

A transport connection is a unidirectional flow of packets that is provided with a particular QoS. Particular CIDs are uniquely associated with both UL and DL connections. Thus, stations can setup numerous connections whose partition between DL and UL direction must not be symmetric.

Contrary, management connections are established as bidirectional flows identified by one single CID. A management connection can carry messages for a number of transport connections.

The standard defines three types of management connections, which shall be used for different purposes. The Basic Management Connection is used to transmit short and time-urgent messages, such as ARQ feedback and channel measurements. The CID of the Basic Management Connection is a kind of unique identifier of a station, thus, it can be used to address a station, e.g., in an UL MAP IE. The Primary Management Connection transports longer, more delay-tolerant messages, such as encryption or connection management messages. Every SS establishes both the Basic and the Primary Management Connection during the network entry. So-called managed SSs need the last type of management connection, i.e., the Secondary Management Connection. Managed SSs do not configure themselves: They receive their configuration over the network. In order to receive their configuration file, managed SSs use higher layer protocols, such as Dynamic Host Configuration Protocol (DHCP), Trivial File Transfer Protocol (TFTP) and Simple Network Management Protocol (SNMP). The Secondary Management Connection carries signaling messages of these higher layer protocols.

CIDs for transport and management connections range from 1 to 65278. Other CIDs are reserved for a specific use, such as initial ranging or broadcast. Besides unicast CIDs for transport and management connections, multicast CIDs can be assigned as well. For this purpose, groups of SSs are associated to one single multicast CID. This multicast CID can be used to poll this group of SSs as introduced in Section 7.5.2.10.

7.5.2.8 Network Entry

A new SS synchronizes to the appropriate BS and makes use of the procedure for network entry and initialization. After the successful network entry, the SS (or the BS) can establish transport connections.

Figure 7.11(a) shows the message sequence chart of the corresponding MAC management message exchange. In order to find an appropriate BS, the SS scans the possible channels of the DL frequency band of operation. An SS should store its last operational parameters in nonvolatile storage in order to check this DL channel first. If this DL channel is not available, the SS continuously scans the other channels until it finds a valid DL signal. The signal, the SS is looking for, is the well-known DL preamble. Since the preamble is periodically transmitted, the SS gathers the frame duration and it synchronizes in time. While receiving the preamble, the SS can estimate the wireless channel. The resulting channel estimates initialize the channel equalization. Thus, the FCH / DLFP, which are following the DL preamble, can be decoded. The first DLFP IE points to the first DL burst, which periodically contains the DL MAP and the DCD. These messages contain all necessary DL parameters required for enabling a DL transmission. When receiving the DL MAP and the DCD successfully, the SS is synchronized.

After the successful PHY and MAC synchronization, the SS performs ranging. In the process of ranging, the SS acquires the right timing offsets, it corrects its frequency offsets and it adjusts its transmission power. For this purpose the SS decodes the UCD in order to obtain the required uplink parameters. From the UCD, the SS extracts a set of transmission parameters for a possible UL channel. The UCD contains, e.g., the uplink center frequency and the size of the initial ranging contention slots. After that, the SS extracts the current position of the UL burst scheduled for initial ranging from the UL MAP.

The SS randomly chooses an initial ranging contention slot to send its Ranging Request (RNG-REQ) message as shown in Figure 7.11(b). The message is addressed to the reserved initial ranging CID. If the SS chooses a proper transmission power and if no collision occurs, the BS will successfully receive the ranging message. The BS responds with the Ranging Response (RNG-RSP) message. This response contains the SS's basic and the primary management CID.

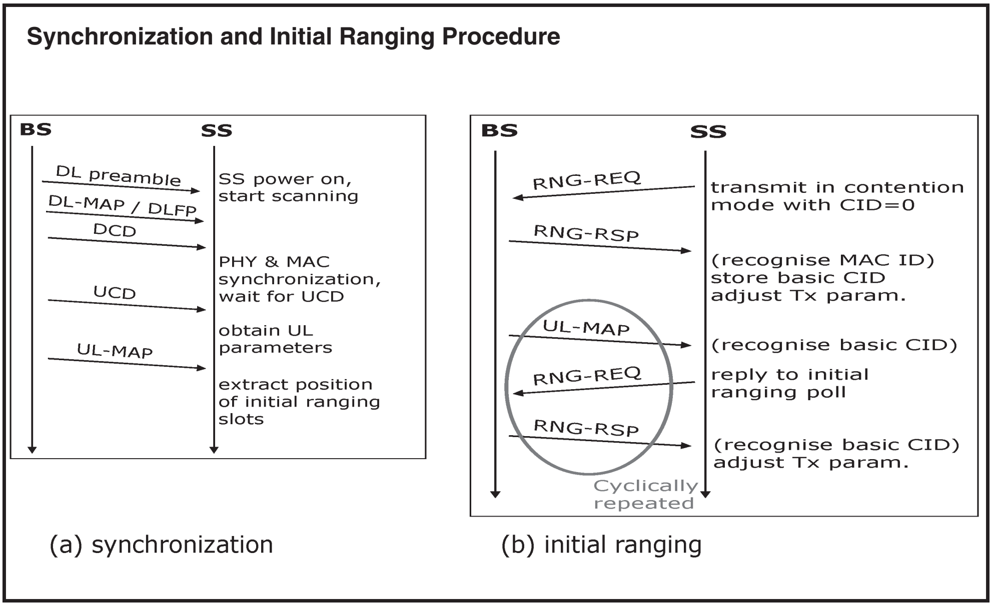


Figure 7.11 Message sequence chart of synchronization and initial ranging procedure.

Furthermore, transmit power level and offset frequency adjustments are included. The timing offset for the SS is given, as well. The exchange of ranging request/response messages can be repeated until the BS notifies a successful ranging. Since the SS has already got its basic management CID, this ongoing message exchange must not be performed on contention slots. The BS can address ranging intervals to the SS's basic management CID.

After the process of ranging, the SS informs the BS about its basic capabilities. The BS responds with the intersection of the SS's and the BS's capabilities. The basic capabilities regard, for instance, the length of the transition gaps TTG and RTG, the maximum transmit power per PHY mode, and the modulator/de-modulator capabilities. Thereafter follow the authentication of the SS and an exchange of encryption keys. After its authorization, the SS performs registration. It is the last step before the SS is allowed to enter the network. In order to register with a BS, the SS sends a registration request that informs, e.g., about its ARQ and CRC support and the convergence sublayer capabilities. Now, an SS successfully entered the 802.16 network. It is ready to establish transport connections. Managed SSs have not yet finished their network entry. In order to become manageable, they receive their secondary management CID during the process of registration. Afterwards, managed SSs establish IP connectivity to download their configuration file from a given server. When the download is completed, the network entry is completed.

7.5.2.9 Connection Management

In order to describe the connection management in 802.16, two terms are defined with a slightly different meaning: a service flow and a connection. The 802.16 standard defines a service flow as a MAC transport service that provides unidirectional transport of packets. It is identified by a 32-bit Service Flow Identifier (SFID). If an SS or a BS wants to use the transport service and the chosen service flow is admitted, then a 16-bit CID is assigned to the service flow. Thus, a service flow can be called a connection once the transport service is realized.

The activation of a service flow, i.e., the establishment of a connection, follows a two-phase activation model that is often utilized in telephony applications. First, the initiating station conserves network resources, such as bandwidth or memory, for an admitted connection request. Once the complete end-to-end connectivity has been established, the resources are activated. The process of activation performs policy checks and admission control on resources as quickly as possible. If possible, this is done before the far end of a connection request is informed. This procedure should prevent potential theft-of-service scenarios.

A service flow has several attributes. It has at least an SFID and a direction. A service flow is characterized by a set of QoS parameters. Due to the activation model, three different QoS parameter sets can be distinguished. First, the Provisioned QoS Parameter Set is provided, e.g., by the network management system. The BS (and optionally the SS) may choose to activate the service flow by requesting the activation. Following the two-phase activation process, the BS checks the SS's authorization, admits the flow and conserves resources. The Admitted QoS Parameter Set defines the set of parameters, for which resources have been reserved. The level of admitted QoS is always lower than the level of provisioned QoS. A CID has been associated to a service flow having a non-zero admitted QoS parameter set. The type of an admitted service flow may remain admitted for a while or it migrates to active. The type changes to active once both peer entities have confirmed the end-to-end connectivity. The service flow now has a non-zero Active QoS Parameter Set. Again, the activated level of QoS is lower than the admitted level of QoS of the service flow.

The QoS parameter sets may contain various parameters. The standard foresees traffic performance parameters, such as the maximum traffic burst length, the minimum reserved traffic rate, the tolerated jitter, the maximum latency, or a maximum sustained peak traffic rate. Other parameters affect the configuration of the user plane. Such parameters may configure the ARQ mechanism or specify the convergence sublayer including payload header suppression. Additionally, the 802.16 standard contains a hook to encode vendor-specific parameters.

The corresponding message exchange follows a three-way handshake of Dynamic Service Addition (DSA) MAC management messages. The DSA sequence, which is composed of a request, a response and an acknowledgment, takes place on the SS's primary management connection. A similar message exchange is specified for the modification of existing connections. By means of the Dynamic Service Change (DSC) mechanism, QoS parameters can be modified if the request is accepted. However, only QoS parameters that are related to the traffic performance can be modified. The ones that configure the user plane are exempt from modification. Finally, connections can be released by exchanging Dynamic Service Deletion (DSD) management messages.

Burst profile management

UL and DL transmissions are associated with a burst profile. The burst profile details the specific coding algorithm, the code rate and the modulation scheme of the corresponding DL and UL burst, respectively. The BS has full control over the association of burst profiles to DL and UL transmissions. The BS decides about the burst profile based on the signal quality, i.e., SINR perceived during the transmission.

The BS can measure the UL SINR during the reception of UL bursts. Thus, the BS chooses the proper burst profile based on the measured SINRs of one or more previous UL transmissions. The UL burst profile, which an SS should use during a given UL burst, is specified in the UL MAP. In the DL, the BS cannot measure the SINR perceived at the receiving SS. Therefore, the SS measures the DL SINR and signals this information back to the BS. In order to reduce the signaling overhead, not every measured value is transmitted. Each SS monitors the observed DL SINR and compares its mean value with the allowed range of operation. If the mean SINR leaves the allowed range, the SS requests a change of the burst profile. The range of operation is defined by DL SINR thresholds. Each burst profile has mandatory entry and exit thresholds, which are encoded in the DCD message. Figure 7.12 shows the thresholds of an example burst profile A. If an SS receives DL bursts with the profile B, but the mean SINR is higher than the entry threshold of profile A, the SS requests the change to the more efficient profile A. If, after a while, the mean SINR drops again and if it passes the exit threshold of profile A, the SS requests the change to the more robust profile B. The resulting DL burst profile, which a BS applies during the transmission of a DL burst, is specified in the DL MAP.

The standard defines two methods to request a change of the burst profile. If the SS has a granted UL allocation, i.e., if a UL burst is scheduled for its basic management CID, the SS sends a Downlink Burst Profile Change Request (DBPC-REQ) in that burst. The BS responds with the Downlink Burst Profile Change Response (DBPC-RSP). If no UL allocation is available, the SS uses the initial ranging contention slots. It sends an RNG-REQ message, which is addressed to the basic management CID, and the BS responds with the corresponding RNG-RSP.

7.5.2.10 Bandwidth Requests and Uplink Scheduling Services

To allow for an adaptive UL transmission of data streams, which may have variable throughput requirements, the standard defines different types of bandwidth request mechanisms. In general, the request mechanisms follow a semi-distributed QoS approach, i.e., bandwidth is requested on

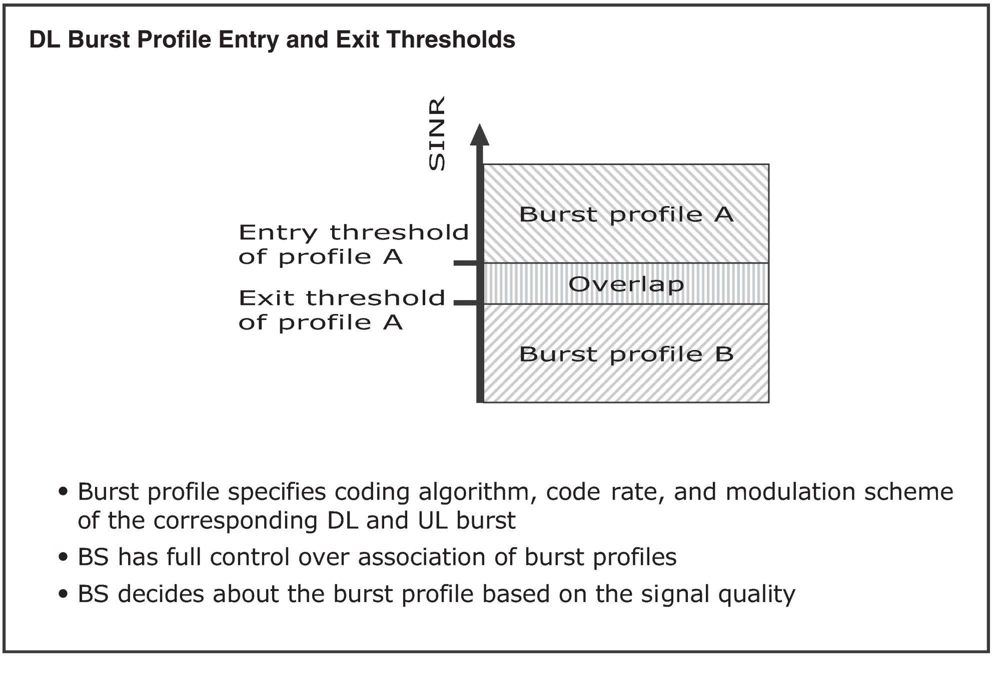


Figure 7.12 DL burst profile entry and exit thresholds.

connection basis (individual transport CID), but it is granted per SS (basic management CID). By partitioning the granted bandwidth to different connections, even the SS has to maintain QoS and fairness of connections.

One way of requesting bandwidth is to transmit a preamble plus a bandwidth request header during a contention interval. The interval can be either addressed to the broadcast or a multicast CID. The bandwidth request header requests the transmission of data for a given transport CID. The amount of data, which is encoded in the message, specifies the aggregate number of bytes waiting for transmission. This request can also be transmitted during any regular UL burst dedicated to the SS's basic management CID. In such bursts, the same bandwidth request header is transmitted as a standalone MAC management message, which specifies either the aggregated or incremental amount of data.

Another way to request bandwidth is "piggybacking". An SS sends a grant management subheader piggybacked to a regular PDU. This subheader requests bandwidth for the same UL connection (transport CID) that carries the PDU. Thus, this mechanism can only be used for connections that currently have UL allocations. Piggyback requests can only specify incremental requests.

BSs can also trigger bandwidth requests. For this purpose, BSs individually poll SSs by allocating UL bursts specifically for UL bandwidth requests. Regular UL MAP IEs specify these UL bursts, but the bursts have only the size of a preamble plus a bandwidth request header. Thus, SSs have the opportunity to send a bandwidth request header that lists the current amount of UL data. An SS can stimulate the BS to poll it by means of a poll-me bit. The poll-me bit is part of the grant management subheader. By setting this bit, the SS indicates its need to be polled. Thus, the SS can transmit a bandwidth request header to request bandwidth for any CID.

In the above-mentioned mechanisms, SSs request bandwidth so that the BS assigns resources on demand. Nevertheless, the BS can also grant UL allocations to SSs without prior notice. For that purpose, it assigns a UL burst to the SS's basic management CID.

In order to efficiently allocate UL bandwidth, UL scheduling services specify which type of UL bandwidth request shall be used for the corresponding connection. Four types of scheduling services allow for different levels of flexibility and efficiency. The type of the scheduling service and the corresponding parameters are negotiated during the connection setup. They are part of the QoS parameter set that individually characterizes each service flow.

Unsolicited Grant Service

The Unsolicited Grant Service (UGS) supports real-time data streams that generate fixed-size data packets on a periodic basis, such as T1/E1 or Voice over IP (VoIP) without silence suppression. The scheduling service assigns fixed-size grants, i.e., UL bursts, at periodic intervals. These UL bursts shall be used for data transmission. The size of the grants and the duration of the time period is specified by the QoS parameter set associated to the service flow. Since service flows using UGS are prohibited from using contention-based request mechanisms, signaling overhead and latency due to bandwidth requests is eliminated.

Real-time Polling Service

The real-time Polling Service (rtPS) supports real-time streams that generate variable size SDUs on a periodic basis, such as MPEG video. Using this scheduling service the BS periodically grants UL allocations. The size of the UL bursts meets the flow's real-time needs and it allows the SS to specify the size of the desired grant. Thus, during the granted UL allocation, the SS sends PDUs containing user data and it requests bandwidth for the next grant either by means of a standalone bandwidth request header or by means of a piggybacked subheader. The minimum and maximum data rate as well as the allowed latency is specified by the QoS parameter set that is associated to the rtPS service flow. Service flows using rtPS are prohibited from using contention-based bandwidth request opportunities. This service requires more signaling overhead than UGS, but supports variable grant sizes for efficient data transport.

Non-real-time Polling Service
The non-real-time Polling Service (nrtPS) supports streams of delay-tolerant SDUs for which a minimum throughput is required, such as FTP. The scheduling service offers unicast polls on a regular basis. Minimum and maximum data rates are specified by the corresponding QoS parameter set. Besides polling, contention-based bandwidth request opportunities may be used by service flows using nrtPS.

Best Effort service
The Best Effort (BE) service supports data streams that require no specific QoS. Service flows using BE scheduling services can be handled on a space-available basis. Thus, BE connections can use contention-based request opportunities. Only a maximum traffic rate is given in the QoS parameter set of the BE service flow.

7.5.3 Security Sublayer

The security sublayer provides subscribers with privacy across the 802.16 network by encrypting connections between SS and BS. The encryption of the payload of a PDU is indicated in the MAC header. The type of encryption and its usage is negotiated during connection setup.

7.6 System Profiles

The standard contains various optional features such as ARQ, PHS and CRC capability. Their implementation is left to the equipment manufacturer. Besides that, many configuration parameters are not fixed so that SS devices can operate with different duplexing modes, have different channel bandwidths or operate in different spectrum bands. To reduce the implementation complexity, the WiMAX Forum started to define system profiles. These profiles list sets of features and parameters that are assumed as being typical implementation cases. Additionally, the process to certify standard conformance and interoperability, i.e., the WiMAX Forum certification, is based on these profiles. The WiMAX system profiles were transferred to the IEEE standardization group, which finalized the profiles and included them in the standard document.

A system profile consists of a set of profiles, each one listing features for a specific purpose. Thus, a system profile is composed of a MAC, a PHY and an RF profile as well as the duplexing selection and a power class. A profile specifies optional features of the standard as "required" or "conditionally required". It does not change the "mandatory" status if it is specified in the standard. Options that do not appear in the profile remain optional. The standard contains separate profiles for the different PHY layer specifications SC, SCa, OFDM and OFDMA. This section outlines the OFDM system profiles.

7.6.1 MAC Profiles

Two MAC profiles are defined, one for PMP (called *profM3_PMP*) and another one for mesh deployments (called *profM3_Mesh*). Both profiles require a Packet CS that supports IPv4 and Ethernet traffic. CRC capability is necessary as well. The profile *profM3_PMP* demands the uplink scheduling services BE and nrtPS, whereas *profM3_Mesh* only requests the BE service. The mesh profile, additionally, requires ARQ functionality. Other features, such as PHS or AAS remain optional.

7.6.2 Physical Layer Profiles

The PHY layer profiles follow the naming convention *profP3_BW* where BW is the provisioned channel bandwidth, which is also the intended channelization. Profiles for 1.75, 3.5, 5.5, and 7 MHz bandwidth define the operation in licensed bands and *profP3_10* defines the license-exempt operation with 10 MHz channel bandwidth. All profiles require specific minimum performance level of the transmitter and the receiver. BSs operating in PMP mode shall support frame durations of 5, 10 and 20 ms. SSs shall allow for transition gaps (TTG and RTG) than 100 μs. Licensed operation additionally requires requires 64-QAM modulation capability, whereas unlicensed operation requires Dynamic Frequency Selection (DFS). DFS is introduced in the context of 802.11 in Chapter 5.

7.6.3 RF Profiles, Duplexing Modes and Power Classes

RF profiles are defined for license-exempt operation only. Three RF profiles define 10 MHz channels in the middle U-NII (5.275–5.335 GHz), the upper U-NII (5.740–5.830 GHz), and in the CEPT band C (5.735–5.835 GHz).

The duplexing mode is either TDD or FDD, furthermore, SSs in FDD are either half- or full-duplex. The power classes specify the range of the devices' transmit power. The intended transmit power classes vary from below 14 dBm (for class *profC3_0*) up to above 23 dBm (for class *profC3_23*).

7.7 Space Division Multiple Access

As one of the first standards IEEE 802.16 includes means to integrate adaptive antenna techniques. Comparable approaches are currently being standardized by the 3GPP for UMTS or by the IEEE for 802.11n. These advanced antenna techniques have a significant impact on the capacity and service quality provided by wireless links and the efficient use of the available spectrum (Ghosh *et al.*, 2005). An initial approach to support Space Division Multiple Access techniques in Wireless ATM systems has been presented in Vornefeld *et al.* (1999).

The simultaneous transmission can be accomplished by pre-distortion or beamforming techniques. The concurrent reception is known as joint detection. In general, the concurrent transmission/reception of data to/from different spatially separated channels is called Space Division Multiple Access (SDMA), see Section 2.3.4. It provides another degree of freedom to conventional TDMA, FDMA or CDMA based medium access.

Systems that operate in SDMA must fulfill several requirements. First, the PHY layer must be able to dynamically adapt its receive and transmit characteristics. Therefore, the antenna array control and a proper algorithm to calculate the antenna weights are implemented in the PHY entities. Second, the enhanced PHY layer has to offer its SDMA services to the MAC layer so that the MAC can leverage the new features. This is done by extending the services offered at the PHY Service Access Point (SAP). Third, the MAC protocol has to cope with the spatial domain. Different peer MAC entities communicate by means of standardized MAC PDUs. Beside the time and frequency domain, these PDUs have to be able to coordinate and utilize the spatial domain. Last, the MAC behavior, which is not standardized, is affected by the spatial domain as well. Vendor-specific algorithms, such as scheduling, have to efficiently handle the new spatial domain.

These four aspects, which enable a system to operate in SDMA, are detailed in the following subsections. Following that, performance results of cellular SDMA enabled WiMAX systems are presented.

7.7.1 PHY Layer Comprising an Antenna Array

Instead of a single antenna, an SDMA enhanced PHY layer uses multiple antenna elements forming an antenna array to transmit and receive signals. The individual antenna elements are spatially separated and arranged in a specific geometric layout (e.g. on a line or circle). The distances between the elements are typically on a scale of $\lambda/2$, where λ is the wavelength of the carrier signal. By controlling the antenna elements individually, beamforming algorithms allow to focus the transmit power into certain directions to increase the received signal strength. It is also possible to steer nulls into certain directions to decrease co-channel interference. An antenna pattern, which defines the transmit and receive characteristic of the array, is steered by applying a weight, i.e., a complex number to each antenna element. Thus, a pattern is represented by a weight vector $\mathbf{w}_i$ which contains one weight per antenna element, see Figure 7.13. If multiple patterns are applied, one weight vector per pattern has to be calculated ($\mathbf{w}_0 \ldots \mathbf{w}_{K-1}$). Beamforming maximizes the SINR by focusing the transmitted energy into the desired direction. At the same time it minimizes the emitted energy towards (all) other directions. The linear nature of the antenna elements enables an antenna array to apply two patterns simultaneously. It can transmit one signal into one direction while it transmits another signal at the same time into another direction. Thereby, both receivers experience a sufficient SINR. Since an antenna is a reciprocal element, the same principle is valid during the reception of signals (Godara, 1997). Here, joint detection techniques allow an antenna array to receive different signals simultaneously. Both signals can be separated and the bit streams can be recovered individually.

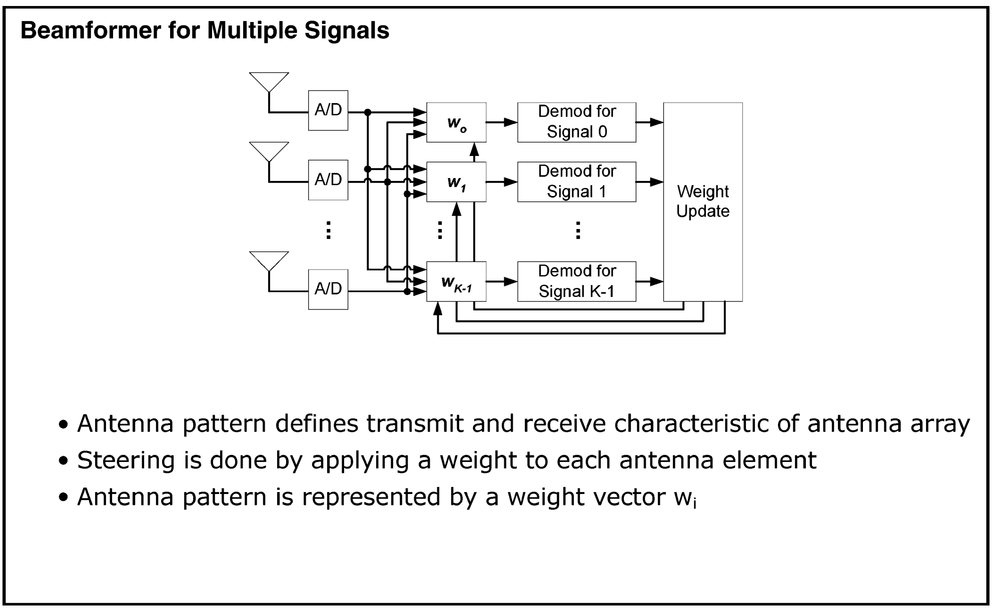


Figure 7.13 Beamformer for multiple signals.

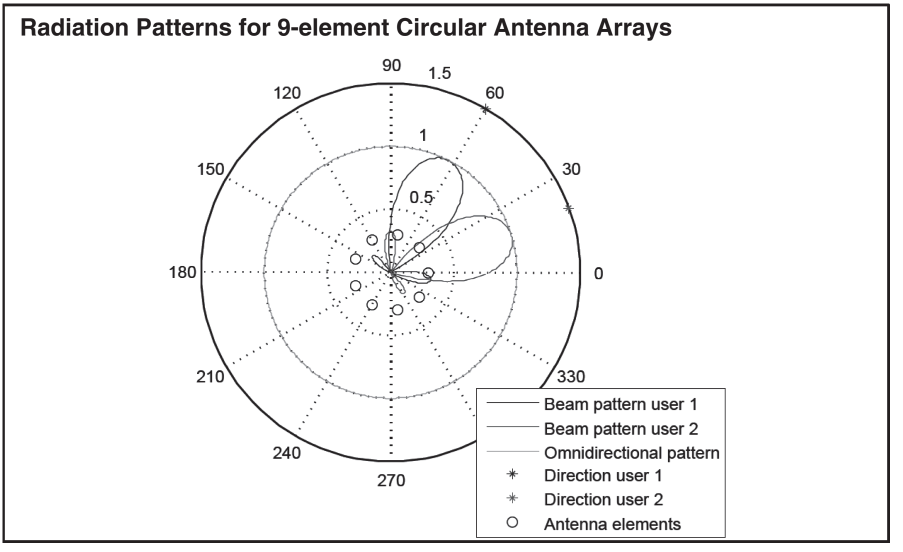


Figure 7.14 Radiation patterns for nine-element circular antenna arrays.

The beamforming algorithm used in the following is the Optimal Beamformer as described in Godara (1997). It uses an array correlation matrix that describes the correlation of the signal strengths received at the array's antenna elements. When building this correlation matrix, the array geometry is taken into account. Using all users' direction of arrivals and the received powers, the correlation of the received signals between all combinations of antenna elements can be computed. Then, the following optimization problem is solved to derive the optimal weight vector: Minimize the total power absorbed from (or emitted to) all directions while assuring unity response in the direction of the desired user.

The result is a beam pattern that realizes a gain of one in the direction of the user. As the algorithm minimizes the total radiated energy, it usually does this by setting nulls in the direction of undesired users. Figure 7.14 shows an exemplary beam patterns for a nine-element Uniform Circular Array (UCA). Each pattern achieves a gain of one in the direction of the desired user while nulling the undesired user. The size of the sidelobes can be influenced by assuming a different level of interference when building the array correlation matrix. Note that if sidelobes exist that have a gain higher than one, the entire pattern is normalized so that the maximum gain is one. The reason is that in most frequency bands the antenna's transmit power is limited by the Equivalent Isotropic Radiated Power (EIRP).

7.7.2 Enhanced PHY Service Access Point

A PHY layer that is composed of smart antennas has to offer its new services to the MAC layer. This is done at the PHY Service Access Point (SAP). New services control the beam patterns that are applied during the reception or transmission of data and they provide feedback about the spatial separability of stations. Neither the MAC nor the PHY SAP is standardized in IEEE 802.16. Their implementation is left to the manufacturer. In the following an exemplary realization of the service "SINR estimation" is given.

In order to estimate the correct SINR, the BS's transmit and the receive cases have to be differentiated. Figure 7.15(a) illustrates all relevant signals during the concurrent reception of data from spatially separated stations. It can be seen that an optimized antenna pattern is applied at the base station. It maximizes the receive power from the wanted signal S from SS A and minimizes the intra-cell interference I_{intra} from SS B that is concurrently transmitting. If more than one station is scheduled to transmit in SDMA, all intra-cell interferences have to be counted as well. The third fraction of the received signal strength is generated by neighboring cells that are operating on the same frequency. By minimizing the sidelobes of the pattern the beamforming algorithm tries to minimize the inter-cell interference I_{inter}^{BS} received at the BS.

The estimation of the SINR during SDMA reception has to consider all above-mentioned signals and filters them through the optimized antenna pattern, thus the signals in the following equation get the index^{opt}. Finally, receiver noise N_{Rx}^{BS} has to be considered. The SINR at the BS during the reception of one SS, which is spatially separated from other SSs can be calculated as:

$$SINR\_Rx_{estimated} = \frac{S^{opt}}{N_{Rx}^{BS} + I_{\text{inter}}^{BS\,opt} + \sum\limits_{\substack{\text{SDMA} \\ \text{stations}}} I_{\text{intra}}^{opt}}$$

Figure 7.15(b) illustrates the relevant signals during SDMA transmission. An optimized antenna pattern is applied at the BS that maximizes the wanted signal S to SS A. All other patterns, which are optimized for concurrent transmissions to other SSs, minimize the intra-cell interference I_{intra} to SS A. Again, neighboring cells generate inter-cell interference I_{inter}^{SS} at the receiver. In the transmit case the sidelobes of the transmit pattern do not affect the SINR at the SS, but they are responsible for inter-cell interference at neighboring cells. During SDMA transmission the wanted signal S and the intra-cell interference I_{intra} are filtered through their corresponding optimized antenna patterns. Since it is assumed that SSs are receiving omni-directionally, the

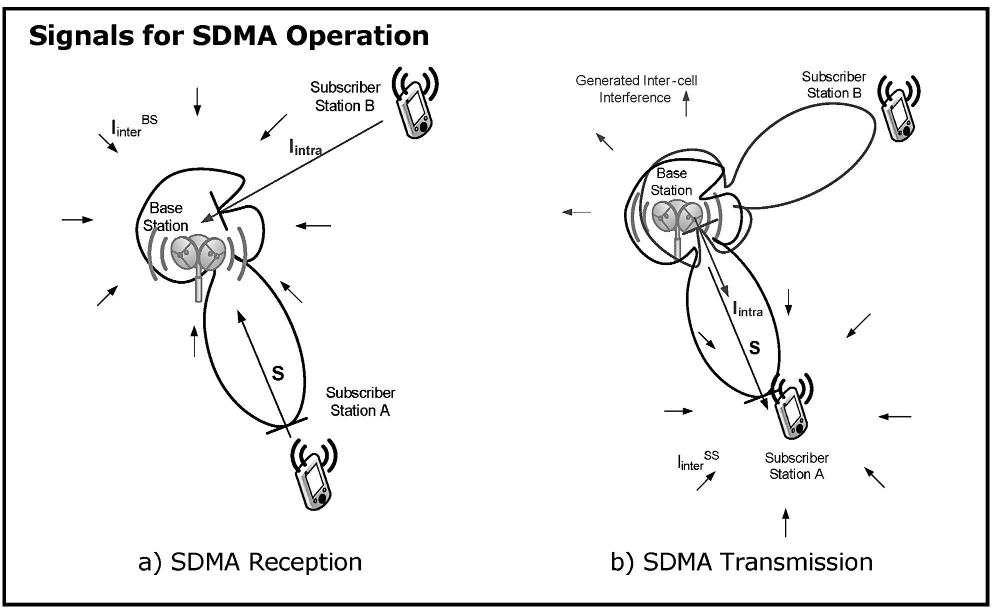


Figure 7.15 Signals for SDMA operation.

inter-cell interference is not filtered. Again, receiver noise N_{Rx}^{SS} has to be considered. This results in the following equation:

$$SINR\_Tx_{estimated} = \frac{S^{opt}}{N_{Rx}^{SS} + I_{\text{inter}}^{SS} + \sum\limits_{\substack{antenna\\patterns}} I_{\text{intra}}^{opt}}$$

7.7.3 SDMA Enhanced Medium Access Control Layer

The standard IEEE 802.16-2004 contains DL- and UL MAP information elements for various purposes. For the DL MAP, a *DL MAP concurrent transmission IE format* can be used that allows specifying one of a set of parallel downlink bursts for SDMA transmission. For the UL MAP, the regular IE already contains all fields that are required for a concurrent reception in SDMA mode.

So the MAC protocol allows for arranging concurrent bursts, which are transmitted or received in SDMA mode. Additionally, optional DL preambles and UL midambles might be included and a possible cyclic shift of preambles and/or midambles further supports channel estimation as well as time and frequency synchronization and therewith the robustness of SDMA. All these enhancements enable the system to fully support SDMA (Hoymann, 2006). Figure 7.16 shows the SDMA enhanced IEEE 802.16 MAC frame. The antenna characteristics are drawn above the frame. Inside the DL burst 1, the DL and the UL MAP are highlighted. The arrows show the timing information included in the IEs, i.e., start time and burst duration.

Starting with the DL preamble, the first part of the frame is sent omni-directionally. Preamble, FCH and DL burst 1 are broadcast as the antenna pattern above the frame indicates. This is necessary because all SSs need to decode the MAPs. At the beginning of the next burst the antenna characteristic is adapted. The BS applies the antenna weight factors to the antenna elements. Thus, the BS can send one data stream containing DL burst 2 in the direction of a SS while at the same time the BS can send a different data stream containing DL burst 3 in the direction of another SS. The number of data streams is limited by the capability of the antenna array to form

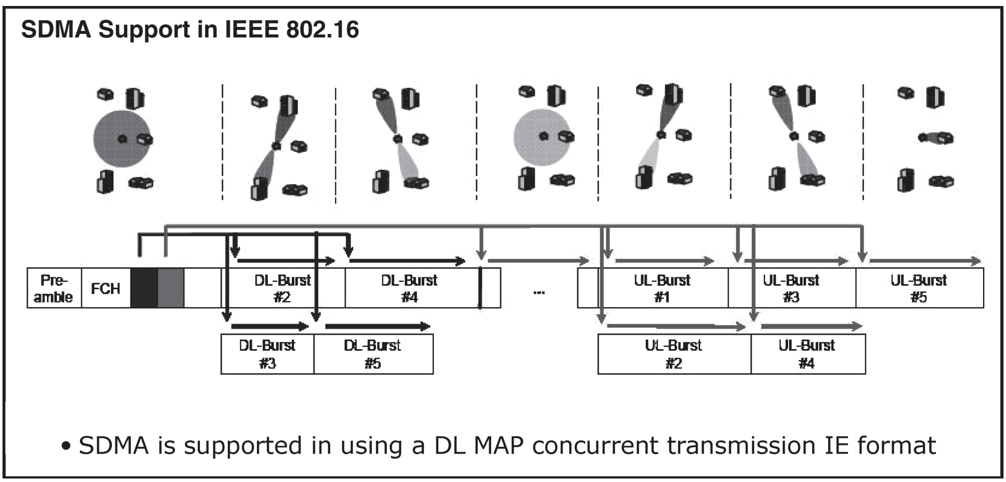


Figure 7.16 SDMA support of IEEE 802.16 using concurrent transmission DL MAP IEs.

patterns that sufficiently separate the different signals. To enable a parallel reception of different data streams at the BS, the UL MAP specifies the duration and the start time of each UL burst. With the UL MAP IE it is possible to let different SSs start their UL transmission bursts at the same time only separated in space.

7.7.4 SDMA Scheduling

Although scheduling algorithms are not specified in the standard, SDMA scheduling is one of the key requirements for a system to operate in SDMA mode (Hoymann, 2006). In the following a hierarchical TDMA/ SDMA scheduling approach is introduced.

SDMA scheduling algorithms look for groups of users that are spatially separable. Users of the same spatial group will be scheduled in one SDMA slot while users of different groups will be scheduled in different time slots (regular TDMA scheduling) (Fuchs *et al.*, 2005). The performance of the spatial grouping directly affects the system performance.

On the one hand, grouping algorithms can search through the whole combinatorial set of possible solutions to find the optimal one. On the other hand, computationally less complex heuristics do not try to find the optimal solution, but a sub-optimal one. In order to set the boundaries of the potential performance gains of different algorithms, three grouping strategies have been evaluated in randomly generated scenarios (Monte Carlo simulation).

The lower bound is marked by the random grouper. The scheme randomly selects a partition. Due to the random nature, the SDMA performance of the resulting schedule is sometimes worse than the TDMA one. This is reflected by the minimum gain of the random grouper, which is 0.69. Gains smaller than one occur so rarely during the simulation that their probability is not visible in Figure 7.17(b). Its mean gain compared to the non-SDMA schedule is 1.79. The results of the

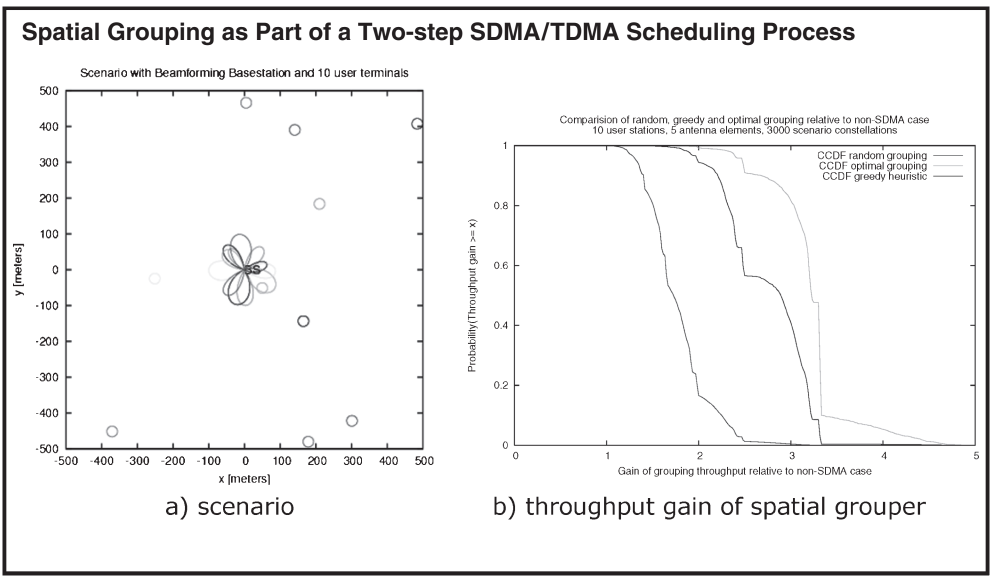


Figure 7.17 Spatial Grouping as Part of a Two-step SDMA/TDMA Scheduling Process. The applied scenario (a) and the throughput gain of spatial grouping (b) are shown.

max throughput grouper resides in between the boundaries. The algorithm first selects the spatial group that results in the highest throughput. This group is scheduled in the first time slot. Taking the set of unserved SSs, the next group is chosen that again results in the highest throughput. The next slot is scheduled for that group. As long as there are unserved SSs left, slots are scheduled. The result is again a partition of the set of SSs into spatial groups. The computational complexity of the max throughput grouper is far below the optimal grouper since it does not search through the entire combinatorial set of possible partitions. Its mean gain is 2.76, which is a promising result.

7.8 Performance Evaluation of 802.16

7.8.1 Multi-user Multi Phy Mode Scenario

The IEEE 802.16 frame structure and therewith the system capacity depends on the number of SSs and on the number of different PHY modes in use. For each SS a UL burst including preamble, padding and possibly midambles is allocated and a UL-MAP IE is inserted in the UL-MAP. For each PHY mode a DL burst is scheduled that might contain a preamble and padding bytes. Additionally, a DL-MAP IE is included in the DL-MAP. Thus, the number of active SSs and the number of PHY modes in use should be realistically covered by the scenario. The scenario introduced in the following supports multiple SSs with different modulation and coding schemes.

7.8.1.1 PHY Layer Configuration and PHY Mode Distribution

The number of SSs using certain MCSs is derived from the surface area covered by the corresponding scheme. To calculate the surface area covered by each PHY mode, the maximum distance between BS and SSs depending on the modulation scheme needs be estimated. This estimation is based on the minimum SINR an SS should receive to avoid data loss. Switching points between modulation schemes depending on receiver SINR are proposed in IEEE (2004a). With the minimum SINR, the maximum distance an SS should have from its BS can be calculated. The noise depends on the system bandwidth and on the temperature of the receiver as introduced in Section 2.1.9. The calculation of the path loss (L_F) between transmitter and receiver in a free-space scenario without any interfering obstacles is described in detail in Section 2.1.1. For the following calculation no other sources of interference are taken into account and perfect transceivers are assumed. In this case the receiver SNR is defined as

$$P_R[dBm] = P_T[dBm] - L_F[dB]$$
$$SNR[dB] = P_R[dBm] - N[dBm].$$

P_R denotes the receive power and P_T the transmit power. Power values are given in dBm whereas the pathloss is measured in dB. With the equations from Section 2.1, the distance between BS and SS in dependency on its signal power and its noise is given by

$$d = \frac{\lambda}{4\pi} \cdot 10^{d\,\frac{P_T[dBm]-SNR[dB]-N[dBm]}{20}}.$$

It is assumed that the scenario system is being operated in the upper 5 GHz band. An example for this is the unlicensed band starting at a frequency of 5.47 GHz that is restricted in Europe to outdoor use. The typical system bandwidth in this frequency band is 20 MHz. The noise in a 20 MHz band can be calculated to 8×10^{-14} watt, which equals -100.97 dBm. As discussed in Section 3.2.3, spectrum regulation limits for this frequency band the maximum allowed Equivalent Isotropic Radiated Power (EIRP) to 1 W. Thus, the transmitters are assumed to have a transmission power P_t of 1 W equal to 30 dBm. The calculated SNR and the switching points can be seen in Figure 7.18.

Each switching point between two different PHY modes results in a certain cell radius. The radius of the last switching point, i.e., BPSK 1/2 marks the cell boundary. In this free-space example the cell area has a maximum radius of approximately 7.4 km. The parts of the surface area of the cell that are covered by specific PHY modes are regions lying between two concentric circles. The area of the annulus ($F_{Annulus}$) formed by two circles of radii R_1 and R_2 is

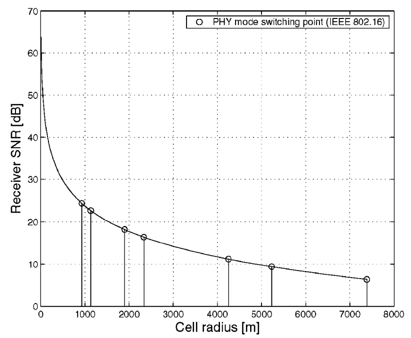
$$F_{Annulus} = \pi \cdot \left(R_1^2 - R_2^2\right)$$

The cell boundary in an ideal cellular deployment is a hexagonal cell so that the area belonging to the last mode, i.e., BPSK 1/2 is not a whole annulus but certain fractions of it are cut. The area covered by BPSK 1/2 can be calculated as

$$F_{BPSK1/2} = \frac{3}{2}\sqrt{3}R_{BPSK1/2}^2 - \pi R_{QPSK1/2}^2$$

The area per PHY mode is a certain fraction of the whole cell area. The proportion of each surface area per PHY mode is listed in the table depicted in Figure 7.18. Note that the distribution of PHY modes in a hexagonal cell neither depends on the frequency band in use nor on the transmission power. One can see that the annulus where the most robust PHY mode BPSK 1/2

Receiver Signal-to-Noise Ratio vs. Distance



Modulation	Coding Rate	Receiver SNR (dB)	Surface [%]	Number of SS
BPSK	½	6,4	39.40	6
QPSK	½	9,4	20.56	3
QPSK	¾	11,2	27.95	4
16 QAM	½	16,4	4.10	1
16 QAM	¾	18,2	5.15	1
64 QAM	2/3	22,7	0.92	0
64 QAM	¾	24,4	1.92	0

Figure 7.18 Receiver SNR vs. distance and usage of PHY modes.

must be used is represented over proportionally. The sensitive and powerful modes in the more inner circles of the cell cannot be utilized very often because their range is limited to a small area.

Assuming a constant density of SSs within the whole cell, the proportion of the annulus of a PHY mode can be converted to a proportion of SSs using this MCS. In the example scenario, a total number of 15 SS per BS is active. The number of 15 subscriber stations results from the analysis conducted in Section 7.8.2.1 where a maximum of approx. 15 SSs could be scheduled in one MAC frame assuming a fixed payload size of 381 bytes. Of course more SSs can be handled by the system, but their payload has to be fragmented into smaller blocks, then or the corresponding connections have to be scheduled interleaved in subsequent MAC frames. Thus, the overall MAC capacity is distributed to the different SSs.

Now the number of SSs using a certain PHY mode can be calculated as the percentage rate of the covered area times the total number of SSs. The resulting number of SSs (see the 5th column) corresponding to the percentages of PHY mode utilization (see the 4th column) is listed in the table depicted in Figure 7.18. Each SS of the multi-user scenario has the same data rate, i.e., 1/15 of the overall data rate. Thus, having for example a system throughput of 30 Mbit/s, each SS is offering 2 Mbit/s (1 Mbit/s DL, 1 Mbit/s UL).

7.8.1.2 MAC Layer Configuration and Performance Metric

The example system with 20 MHz bandwidth operates in the 5 GHz bands in TDD mode. The frame length is set to 10 ms. Figure 7.19 illustrates the MAC frame that is used in the analysis. The scenario deals with several DL and UL connections. The DL subframe consists of the DL preamble, the FCH and several DL bursts. The optional DL preambles are disabled but the CRC is applied to the PDU. According to the standard, the as well as the RTG should be below 100 μs, here they are set to 26 μs, i.e., two OFDM symbols.

The UL subframe starts with three ranging slots. To allow for the transmission of the long preamble and the ranging request management message (RNG_REQ, 13 byte) and to cope with a round trip delay of stations being up to 8000 meters away, eight OFDM symbols are considered for each ranging slot. Each of the 10 bandwidth request slots containing a short preamble and the BW-REQ management message is considered to be two OFDM symbols long. Several UL bursts are scheduled in the UL subframe. The optional midambles are disabled. Packing and fragmentation are initially shut off so that every burst is padded to fill up an integer number of OFDM symbols. The following performance evaluation of the system throughput does not take ARQ into account.

To calculate the resulting static system throughput MAC and PHY overhead is subtracted. Thus, all frame elements which do not contain payload have been taken off (white parts of Figure 7.19). The payload of the MAC PDUs remains. Now the bit rate on MAC level can be calculated by dividing the payload of the MAC frame by the frame duration.

7.8.2 Performance Analysis

The IEEE 802.16 MAC frame structure and the PHY layer characteristics have been modeled by an analytical system model implemented in MATLAB. The model takes all the above introduced features into account and calculates various different MAC and PHY layer measures. The scheduling is based on a fair Round Robin algorithm where all SSs are treated equally. First the DL subframe is filled up and afterwards the UL bursts are allocated. Starting from the highest

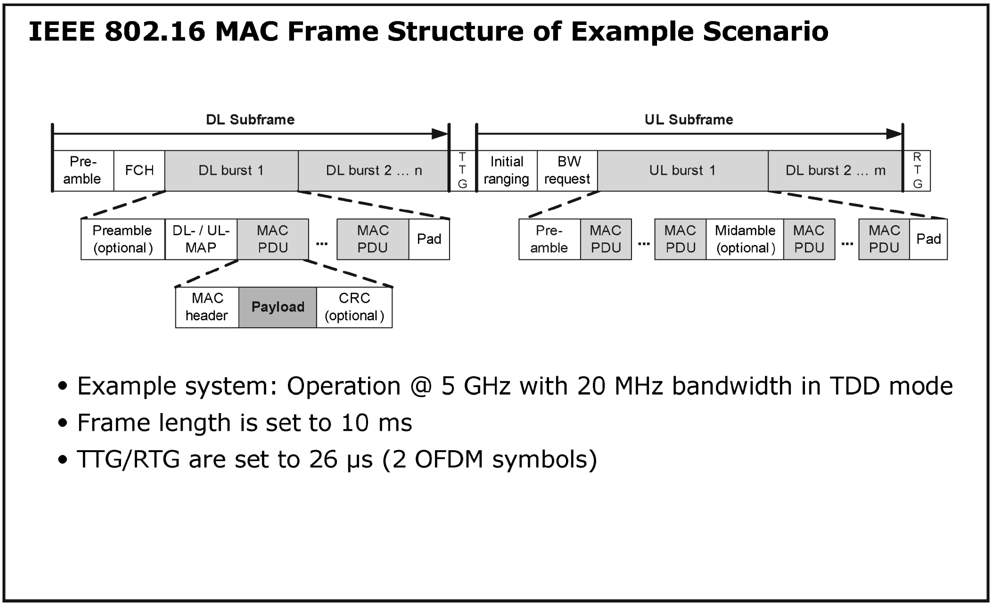


Figure 7.19 IEEE 802.16 MAC frame structure of example scenario.

PHY mode one PDU of each SS that can be served by that mode is scheduled for transmission. This is repeated for the next lower PHY mode etc. When one PDU of each SS in the cell has been scheduled, the algorithm restarts as long as there is free capacity in the subframe. If necessary, DL and UL MAPs are included and updated. Optional preambles are prepended and CRCs are appended if required. In the end each burst is filled up with PDU fragments or bursts are padded up to an integer number of OFDM symbols. Thus, the MAC frame is entirely filled. Characteristics of the wireless channel are not regarded in the model. The transmission is assumed to be error-free. In the following the model is used to evaluate the IEEE 802.16 system performance and the influence of various different MAC and PHY layer configurations.

7.8.2.1 System Performance of the Example Scenario

From the basic assumptions of the PHY layer that have been made for the example scenario (20 MHz bandwidth in the upper 5 GHz frequencies using a CP of 1/4) the basic OFDM parameters can be calculated. The resulting values are listed in the last column of table (a) in Figure 7.20. It can be seen that in the targeted frequency band a CP of 1/4 of the useful symbol duration leads to a guard period of 2.78 μs. This CP has been chosen to deal with delay spreads for NLOS operation in suburban areas (compare to table depicted in Figure 7.18). The OFDM symbol duration is calculated to 13.89 μs.

Using different PHY modes, a certain amount of data can be carried by a single OFDM symbol. The throughput on PHY level can be calculated by dividing the amount of data per symbol by the symbol duration. The results vary from 6.91 Mbit/s for BPSK 1/2 up to 62.21 Mbit/s for 64 QAM 3/4. All resulting PHY throughput values are listed in table (b) of Figure 7.20.

(a) Basic OFDM Parameters

OFDM Parameters	Value	Scenario
Bandwidth BW		20 MHz
Sampling Rate $F_s = 1/T$	depends on BW	23.04 MHz
Useful Time T_B	$256 \cdot T$	11.11 μs
T_G / T_B	$^1/_4$, $^1/_8$, $^1/_{16}$, $^1/_{32}$	$^1/_4$
CP Time T_G		2.78 μs
Symbol Time T_{sym}	$T_G + T_B$	13.89 μs
Carriers N_{FFT} /Data Carriers	256/192	

(b) PHY Modes and Corresponding Throughput

Modulation	Coding Rate	PHY throughput (analysis) [Mbit/s]	MAC throughput (analysis) [Mbit/s]	MAC throughput (simulation) [Mbit/s]
BPSK	½	6.91	6.1	5.8
QPSK	½	13.82	12.19	11.8
QPSK	¾	20.74	18.59	17.6
16 QAM	½	27.65	24.69	23.8
16 QAM	¾	41.47	37.19	35.2
64 QAM	2/3	55.30	49.68	46.9
64 QAM	¾	62.21	55.78	52.7

Figure 7.20 (a) Basic OFDM Parameters and (b) PHY Modes and corresponding throughput on PHY and MAC level.

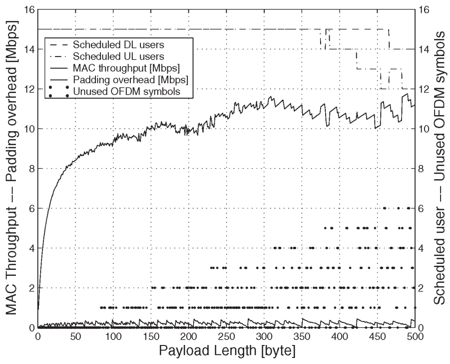
To show the corresponding MAC capacity of the different PHY modes a **single user scenario** is evaluated first. Serving a single user the MAC frame only contains one DL and one UL burst. Since a uniform payload length cannot be derived from the higher layer protocols like TCP/IP, UDP, ATM or Ethernet the single user scenario assumes to have a traffic load of Constant Bit Rate (CBR) traffic with a packet size of 381 bytes (Hoymann, 2005). All other parameters are taken from the example scenario, and the MAC frame length is 10 ms.

The resulting throughput on MAC level ranges from 6.10 Mbit/s for a single BPSK 1/2 user up to 55.78 Mbit/s for one 64 QAM 3/4 station. MAC layer throughput results for all PHY modes are listed in table (b) of Figure 7.20. Approximately 90% of the bit rate on PHY level is available to higher layers, or in other words the MAC protocol reduces the bit rate by 10% due to protocol overhead.

The IEEE 802.16 system capacity depends on the number of SSs and on the number of different PHY modes represented in the respective scenario. For the **multi-user multi-mode scenario** the distribution of users and their corresponding PHY modes is assumed as derived in Section 7.8.1.2. In the following the example scenario will be evaluated.

Figure 7.21 plots the MAC throughput of the example scenario versus the payload length. When transmitting small packets the throughput is low because the overhead due to the MAC header and the CRC dominates the small MAC payload. When transmitting ATM cells of 53 byte length, the MAC throughput lies at 8.27 Mbit/s. With increasing payload length the MAC overhead decreases and the throughput reaches a maximum of around 11 Mbit/s (10.67 Mbit/s at 381 byte). Since all SSs are scheduled fairly a MAC throughput of approx. 711 kbit/s can be provided to each SS, where 355 kbit/s is allocated in UL and 355 kbit/s is scheduled in DL direction.

MAC Throughput – Single User Scenario



- Small packets: Throughput is low because of overhead due to MAC header and CRC dominates MAC payload
- With increasing payload length the MAC overhead decreases and throughput reaches its maximum by 11 Mbit/s

Figure 7.21 MAC throughput of example scenario.

Once the packet length exceeds a certain length, one single frame MAC frame cannot handle all stations. The number of scheduled SSs in one DL and one UL subframe is also drawn in Figure 7.21 with the ordinate on the right-hand side of the graph. It can be seen that for a payload length larger than 350 byte, one MAC frame does not have enough capacity to handle all stations at once. Thus, not the entire 15 SSs can be supported. With increasing payload length, first the UL subframe reaches the limit and shortly afterwards the DL subframe cannot serve all users. From this point on the payload would have to be fragmented or the stations would have to be scheduled in subsequent MAC frames to be able to serve all SSs. Since the partitioning of scheduled users onto PHY modes varies with payload lengths larger than 350 byte, the shape of the curve changes at this point.

Fragmentation of packets is disabled in the example scenario. Thus, bursts are padded to end up on OFDM symbol boundaries. Padding bits cannot be used and are counted as overhead. The amount of padding bits per frame is divided by the frame duration and the resulting padding overhead is shown in Figure 7.21. The padding overhead varies between 0 and 450 kbit/s. Its mean value for packets between 1 and 500 byte is 265.55 kbit/s. Since fragmentation is disabled some OFDM symbols might remain unused. In this particular case one OFDM symbol remains to be allocated in the subframe but the MAC PDU is too long to fit into it. Even if the highest PHY mode is used the symbol cannot be filled. The amount of unused OFDM symbols per frame is also plotted in Figure 7.21. The highest PHY mode used in the scenario is 16 QAM 3/4 in which an OFDM symbol contains 72 byte. Once the PDU length exceeds this limit, i.e., the payload length exceed 62 byte, one OFDM symbol might stay unused. If the PDU length exceeds 144 byte, i.e., a payload length of 134 byte, two OFDM symbols might not be allocated. This reasons the steps in the graph of the padding overhead.

The padding overhead and the unused OFDM symbols can be utilized to transmit MAC PDUs if it is possible to fragment at least the last PDU of each burst. Figure 7.22 plots the difference

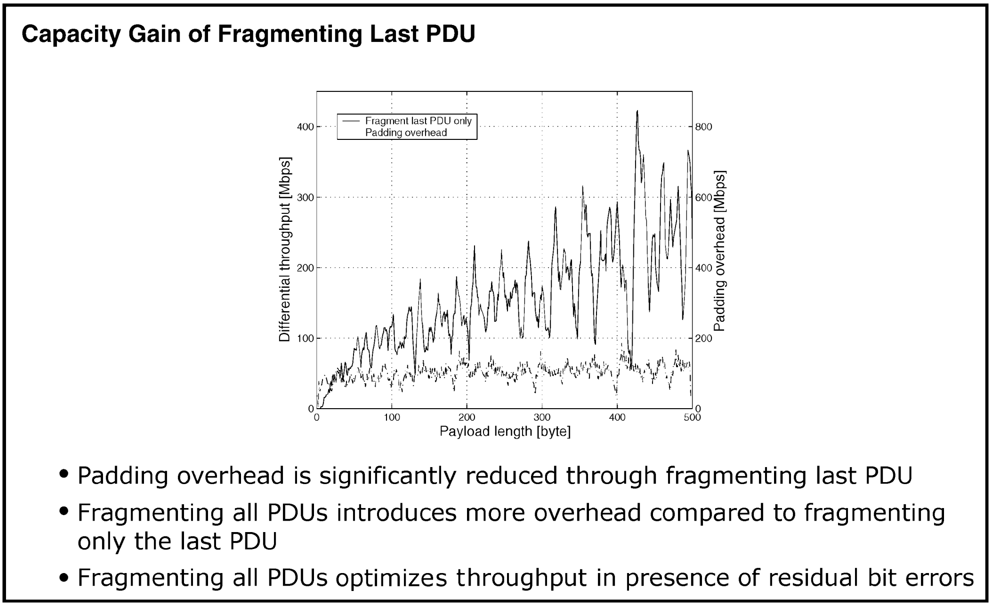


Figure 7.22 Capacity gain of fragmenting last PDU.

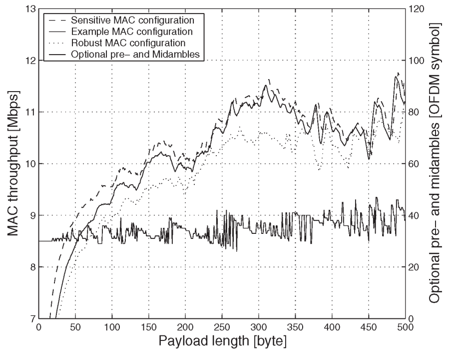
between the MAC throughput without fragmentation and the throughput of a system that allows fragmenting the last payload into PDUs that fill up the bursts. The differential throughput varies between 0 and 420 kbit/s depending on the payload length. The mean value for a payload between 1 and 500 byte is 157.1 kbit/s.

By fragmenting the last payload scheduled in a burst all unused OFDM symbols can be allocated, but the padding overhead can only be reduced, it cannot be eliminated. Due to the size of the MAC header, the fragmentation subheader and the CRC, gaps smaller than 12 byte still have to be padded. The remaining padding overhead is drawn in Figure 7.22. It can be seen that the padding overhead is significantly reduced compared to the non-fragmentation case. But its mean value remains at 67.79 kbit/s compared to 265.55 kbit/s in Figure 7.21.

Fragmenting all PDUs introduces more overhead compared to fragmenting only the last PDU. The subheader included after every MAC header reduces the MAC throughput. Compared to the non-fragmentation case a mean gain in MAC throughput of 43.0 kbit/s can be achieved. Thus, fragmentation of all PDUs does not gain in terms of MAC frame utilization compared to the case of fragmenting the last PDU only. It optimizes the throughput in the presence of residual bit errors. Furthermore fragmentation allows increasing the number of served users per frame as outlined before. Finally, fragmentation is necessary to transmit payload that is larger than the maximum payload length of 2041 byte. If the optional ARQ mechanism is enabled for a specific connection the fragmentation or packing subheader is used to sequentially number the transmitted blocks.

By means of optional preambles prepended to DL bursts IEEE 802.16 provides the means to enhance the stations' capability to synchronize to the BS and to estimate the channel. Optional midambles included in UL bursts facilitate the BS to improve its channel estimation quality. The optional CRC is necessary to detect bit errors in the payload of MAC PDUs. If an error occurs the ARQ mechanism can request to retransmit the erroneous PDU. If ARQ is disabled for a

Capacity of different MAC Configurations



- Sensitive configuration: CRC as well as all optional DL preambles and UL midambles are disabled
- Robust configuration: Preambles are prepended to each DL burst and midambles are included

Figure 7.23 Capacity of different MAC configurations.

particular connection the PDU is simply discarded. These features (optional DL preambles, UL midambles and CRCs) allow for configuring the MAC protocol in a more or less robust way. As a trade off for robustness, the MAC capacity is decreased. In the example scenario the optional preambles and midambles are disabled. CRC is enabled and protects the MAC PDU payload.

Figure 7.23 shows the throughput of a sensitive MAC layer configuration and the throughput of a robust MAC layer configuration. As a reference the MAC capacity of the example scenario is plotted. The sensitive configuration disables the CRC as well as all optional DL preambles and UL midambles. By disabling the CRC each MAC PDU is shortened by 4 byte. Especially in situations where the payload is short, e.g., below 50 byte, the MAC capacity is increased by more than 500 kbit/s. Exceeding the payload length of 150 byte, the influence of the CRC overhead starts to vanish. The MAC capacity of the example scenario and the MAC capacity of the system with a sensitive MAC configuration approach each other. The mean throughput gain for packets between 0 and 500 byte amounts to 255.6 kbit/s. Within the robust MAC layer configuration preambles are prepended to each DL burst and midambles are included after every eighth OFDM symbol of each UL burst. As in the example scenario a 32-bit CRC protects the payload of every PDU. Beside the MAC throughput Figure 7.23 additionally plots the number of OFDM symbols occupied by optional preambles and midambles that have been included in the MAC frame. It can be seen that the overall number of optional pre- and midambles varies around 30. Five preambles are included in the DL bursts of the five different PHY modes of the scenario. The remaining 25 midambles are included in between the OFDM symbols of the UL bursts. Thus, a relatively constant overhead is introduced by this optional feature. A mean capacity loss of 461.9 kbit/s compared to the example scenario is the trade off for the robustness of the MAC configuration.

To protect the useful part of the OFDM signals against inter-symbol interference and inter-carrier interference, a CP is prepended. The CP is configured to be a certain fraction of the useful

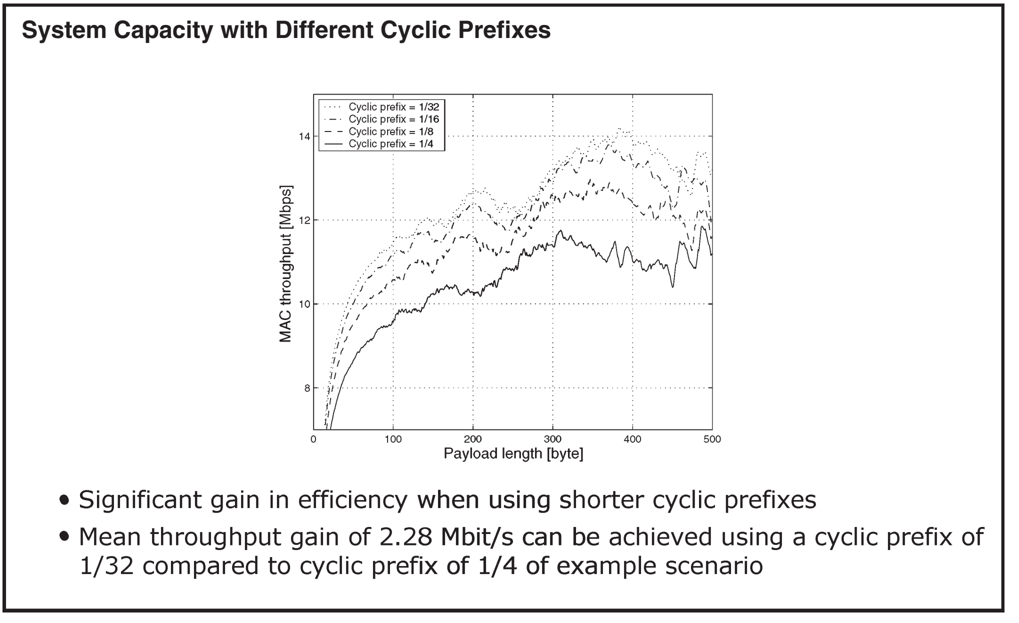


Figure 7.24 System capacity with different cyclic prefixes.

time. Values of 1/32, 1/16, 1/8 and 1/4 are foreseen by the standard. The resulting CP length has to be larger than the maximum delay spread of the environment. Due to the increased symbol duration, the SNR and the number of symbols per second that are transmitted decreases (Engels, 2002). Figure 7.24 plots the MAC throughput of 802.16 systems using different cyclic prefixes. It can be seen that the throughput highly depends on the length of the CP. Thus, the CP should not be made longer than really necessary. Indoors and in open areas, the maximum delay spread is smaller than 0.2 μs as shown in table (b) of Figure 2.9. A CP of 1/32 of the useful time results in a guard time of 0.347 μs and is sufficient for the target environment. Figure 7.24 shows the significant gain in efficiency when using a shorter CP. A mean throughput gain of 2.28 Mbit/s can be achieved using a CP of 1/32 compared to the CP of 1/4 of the example scenario. In LOS operation in suburban areas the maximum delay spread does not exceed 1.0 μs. A CP of 1/8 of the useful time results in a CP length of 1.389 μs and protects the OFDM symbols sufficiently. A mean throughput gain of 1.14 Mbit/s is obtained compared to the reference scenario.

The performance analysis has shown that the IEEE 802.16 standard provides several means to adapt the MAC and PHY layer configuration to the system environment and user demands. With an efficient usage of these features the system performance can be optimized by maintaining the robustness and the operability of the system.

7.8.3 Simulative Performance Evaluation

7.8.3.1 IEEE 802.16 Simulator

A software-based simulator with a prototypical implementation of the IEEE 802.16 protocol has been developed at RWTH Aachen University, Chair of Communication Networks. The protocol stack is

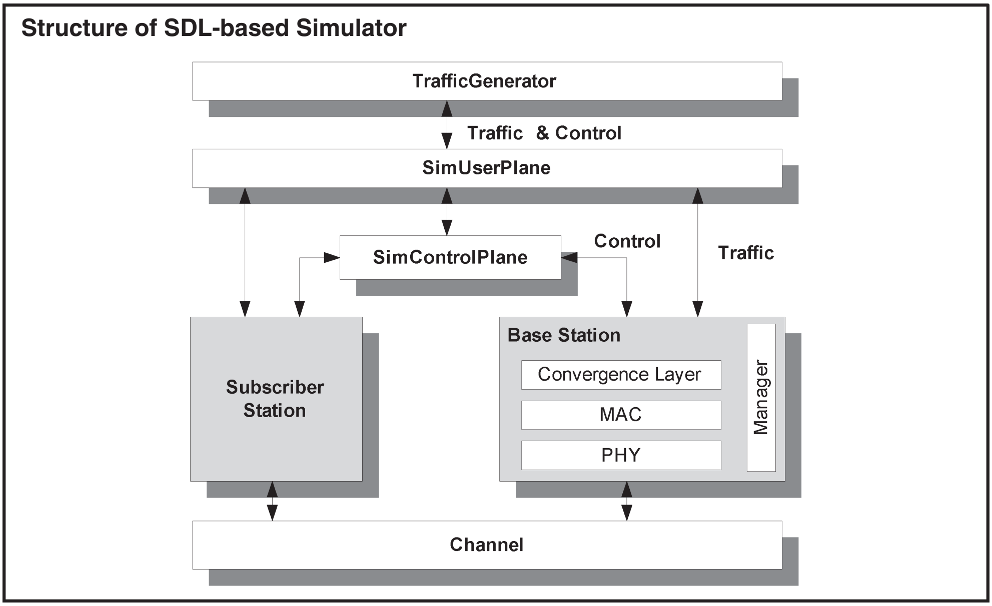


Figure 7.25 Structure of 802.16 simulator based on Specification and Description Language (SDL).

formally specified with the Specification and Description Language (SDL) and is translated to C++ by means of a code generator. The structure of the event-driven simulator is shown in Figure 7.25.

The protocol stacks of SS and BS are implemented. The stack is composed of the convergence layer, the MAC and the PHY layer. Stochastic traffic models generate well-defined traffic loads with characteristic values for several different applications such as MPEG, Ethernet or CBR. Two control blocks manage the simulation, configure the scenarios and stochastically evaluate the transmitted packets.

A physical channel transmits the bursts between the SSs and the BS. Based on the pathloss of the carrier, the interference introduced by other stations and the receiver noise, the channel calculates the SINR for the particular packet. The SINR is mapped to the corresponding error ratio look-up table. These tables introduce the specific behavior of the PHY layer and the wireless channel. The look-up tables are generated by a sophisticated link layer simulation chain, which has been developed during the IST-STRIKE project (Hutter *et al.*, 2004; IST-STRIKE, 2006). The link layer simulation chain models the behavior of transmitters and receivers of 802.16 systems, see Figure 7.4. It implements all relevant transmit and receive blocks such as randomizer, coder, interleaver etc. Beside the transceivers the simulation chain models the real channel characteristics. Different channel models, e.g. the Stanford University Interim (SUI) models or an Additive White Gaussian Noise (AWGN) channel model are available.

7.8.3.2 Simulation Results

The traffic generator generates CBR traffic with a fixed payload size of 381 byte. Payload is encapsulated into MAC PDUs without being packed or fragmented. ARQ is also disabled. To

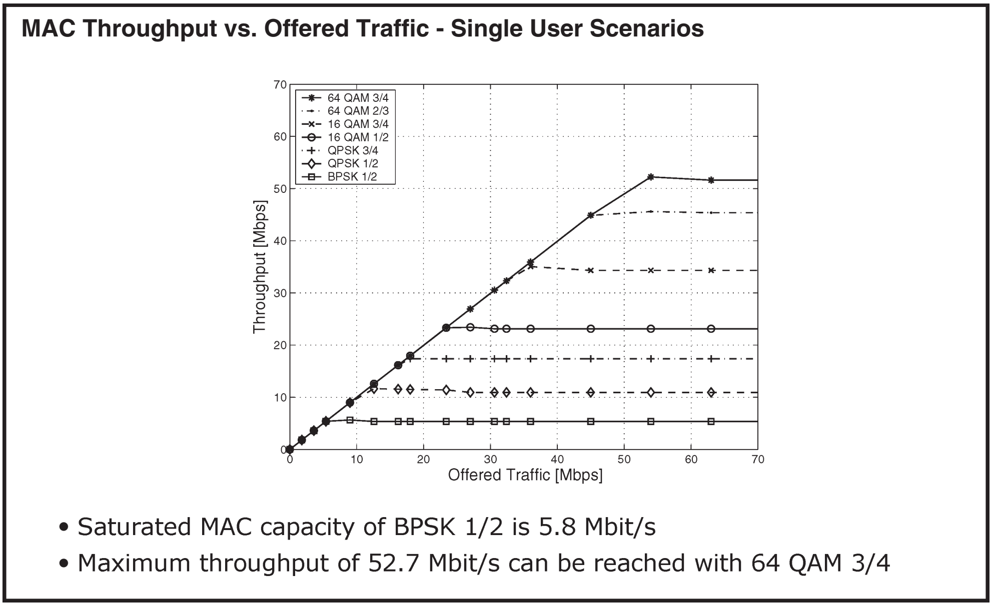


Figure 7.26 MAC throughput vs. offered traffic in single user scenario.

validate the simulator MAC capacity of the different PHY modes with the MAC throughput values of the analysis, a single user scenario is evaluated first. Therefore packet transmission over the channel is assumed to be error-free.

Figure 7.26 illustrates the linear relationship between carried and offered traffic. As long as the offered traffic does not exceed the maximum possible bit rate of the corresponding PHY mode, it is entirely carried. The upmost graph corresponds to the highest PHY mode, i.e., 64 QAM 3/4, and the lowest graph to the lowest modulation and coding scheme, i.e., BPSK 1/2. Having reached the saturation level the MAC throughput stops increasing and it remains nearly constant. The saturated MAC capacity of BPSK 1/2 amounts to 5.8 Mbit/s. The highest PHY mode 64 QAM 3/4 reaches 52.7 Mbit/s. The maximum throughput values obtained with the IEEE 802.16 simulator for all MCSs are presented in table (b) of Figure 7.20 to enable a comparison with the analytic values.

The upper limits of these throughput values have been predicted through the theoretical analysis of the previous sections. They do not match exactly with the IEEE 802.16 simulator since the bandwidth request mechanism in the simulator is based on bandwidth requests. Requests are sent during a contention phase and might collide. Other functionality such as association and the transmission of channel descriptor messages (DCD, UCD) further reduces the capacity slightly.

Figure 7.27 shows the Complementary Cumulative Distribution Function (CDF) of DL and UL packet delay values for 64 QAM 3/4. The packet delay contains queuing and transmission delay. The offered throughput varies between 40 Mbit/s and 50 Mbit/s. The minimum DL delay for an offered traffic of 40 Mbit/s (approximately 76 % of the maximum, see table (b) of Figure 7.20) is around 3 ms, which signifies that these packets were sent right after the DL preamble, FCH

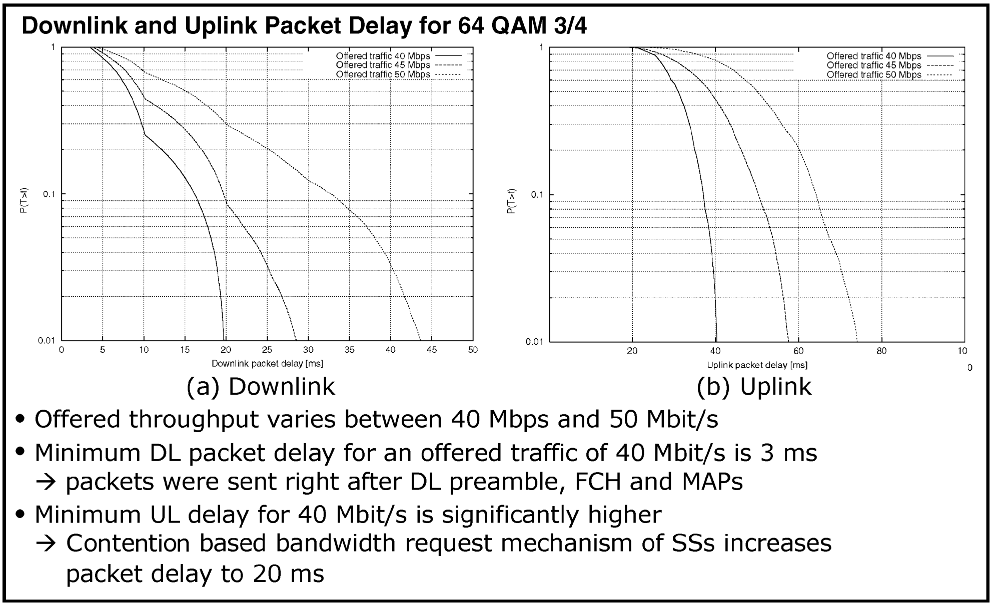


Figure 7.27 Downlink and uplink packet delay for a PHY mode of 64 QAM 3/4.

and the MAPs of the proximate MAC frame. Only 25 % of all arriving DL packets needed more than 10 ms (one MAC frame) to be transmitted.

The minimum UL delay for an offered traffic of 40 Mbit/s is significantly above the minimum DL value. Due to the bandwidth request mechanism of the SSs the minimum delay increased to 20 ms (the MAC frame length assumed is 10 ms). For the uplink scheduling of the services *best effort* and *non-real-time polling*, the SS sends a bandwidth request message to the BS, which allocates UL bandwidth for transmission in the following MAC frames. Other scheduling services such as real-time polling service and unsolicited grant service, which do not rely on accessing contention phases, are foreseen for IEEE 802.16 but have not been considered in this simulation. Additional delay can be counted towards the mean duration of the DL subframe, initial ranging and bandwidth request periods. These periods are scheduled prior to the UL data transmission bursts – 50 % of all arriving UL packets needed more than 30 ms (three MAC frames) to be transmitted. The delay increases for DL and UL if the offered throughput increases towards its maximum.

Finally, the **multi-user multi mode scenario** is evaluated under consideration of transmission errors caused by an AWGN channel. The same scenario as introduced in Section 7.8.1 is studied again. The distribution of users and PHY modes for the scenario is listed in table (a) of Figure 7.20.

The simulated MAC throughput is plotted in Figure 7.28. As long as the offered traffic does not exceed the maximum possible MAC throughput, it can be entirely carried. Reaching the saturation capacity the MAC throughput remains nearly constant. The saturation point of the simulation scenario amounts to approximately 10 Mbit/s. When exceeding the system capacity (beyond saturation), the cell throughput is smaller than the offered traffic, since many packets must be dropped. The MAC throughput capacity found by simulation is confirmed by the calculated throughput shown in Figure 7.21.

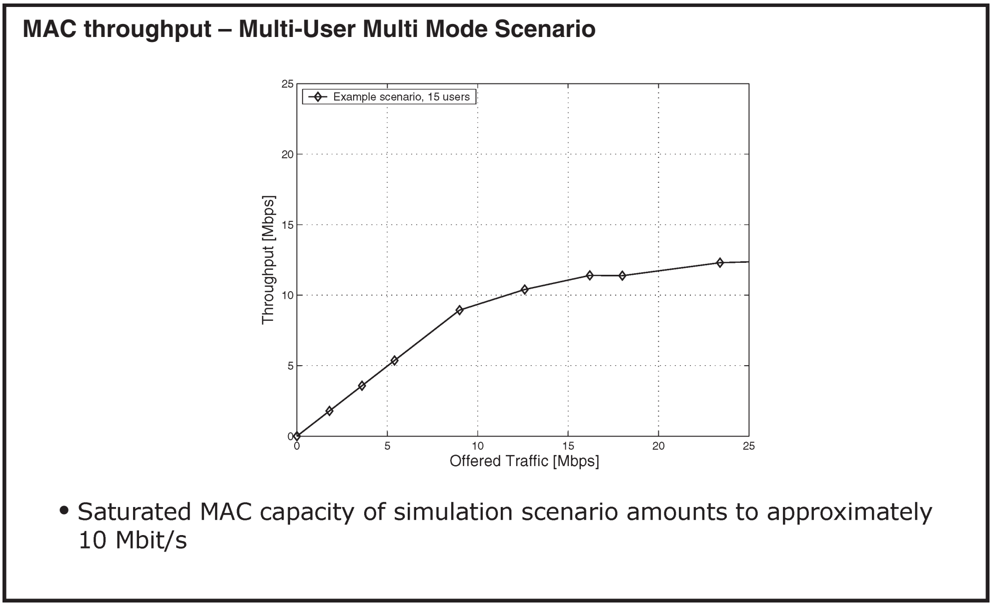


Figure 7.28 MAC throughput vs. offered traffic in the multi-user multi mode scenario.

The analysis of the simulation scenario using the mathematical model results in a MAC capacity of 10.67 Mbit/s. The simulated capacity at system saturation does not match this theoretical value exactly, because the model of the wireless channel introduces transmission errors. Transmission errors lead to packet losses and reduce the MAC throughput. The aforementioned effects of contention-based bandwidth request mechanisms further reduce the MAC throughput. In overload operation, the scheduling gets unfair as SSs needing a robust MCS are not served any more.

7.9 Performance of SDMA Enabled 802.16 Networks

7.9.1 Scenario and Simulation Environment

The evaluated scenario consists of seven cells, each with a central BS and 25 SSs. Measurements are only performed in the central cell. The stations in the surrounding six cells only produce interference for the central cell and are not evaluated. Nevertheless, the same event-driven stochastic simulation, with identical average traffic loads, and with the same degree of detail, is conducted at all 182 stations. The cells have a radius of $R = 1750$ m and an $N = 7$ cell cluster order. Each base station is equipped with a nine-element uniform circular antenna array used to serve the whole cell without sectorization. The SSs are equipped with standard omnidirectional antennas. For both station types, the transmit power is 1 W and no further power control/adaption is performed. A bandwidth of 20 MHz with a mid-frequency of 5.470 GHz is used. All stations are assumed to be fixed. As a rooftop deployment for the subscriber station's antenna is envisioned, the pathloss model presumes LOS conditions. The "C1 LOS" pathloss model for a suburban environment as derived by the WINNER project (WINNER, 2005) is used in the following.

The simulator implements the detailed 802.16 MAC protocol for both base and subscriber stations. A stochastic traffic model generates a well-defined traffic load with a fixed packet size of 1024 bit and a negative-exponential inter-arrival time. The inter-arrival time is adapted to generate various amounts of offered traffic. A Radio Interference Simulation Engine (RISE) models the physical layer as well as the wireless channel. It is based on look-up tables that are generated by a sophisticated link layer simulation chain, which has been developed during the IST-STRIKE project (Hutter *et al.*, 2004; IST-STRIKE, 2006).

7.9.2 Downlink Cell Throughput

In Figure 7.29 the DL cell throughput is plotted versus the total offered DL traffic. The diagonal line marks the total offered traffic. The graphs represent results of different transmission modes. In the (non-beamforming) reference scenario, all transmissions are performed using omnidirectional antennas. In the beamforming scenario with a maximum number of one concurrent data streams, the base station uses smart antennas only to reduce interference. It calculates optimized antenna patterns for the transmission and reception of signals, but only one station is served at a time. This technique is called Spatial Filtering for Interference Reduction (SFIR) and is often proposed to reduce cluster sizes in cellular systems. In the beamforming scenario with a maximum number of four data streams, the smart antenna at the base station is used to serve up to four stations in parallel. The concurrent data streams are spatially separated by optimized antenna patterns (SDMA).

Figure 7.29 shows that the non-beamforming scenario can carry up to 5 Mbit/s offered DL traffic. Since a symmetric partition into DL and UL subframes is assumed this results in a total cell throughput (DL and UL traffic) of approximately 10 Mbit/s. If the traffic load further increases,

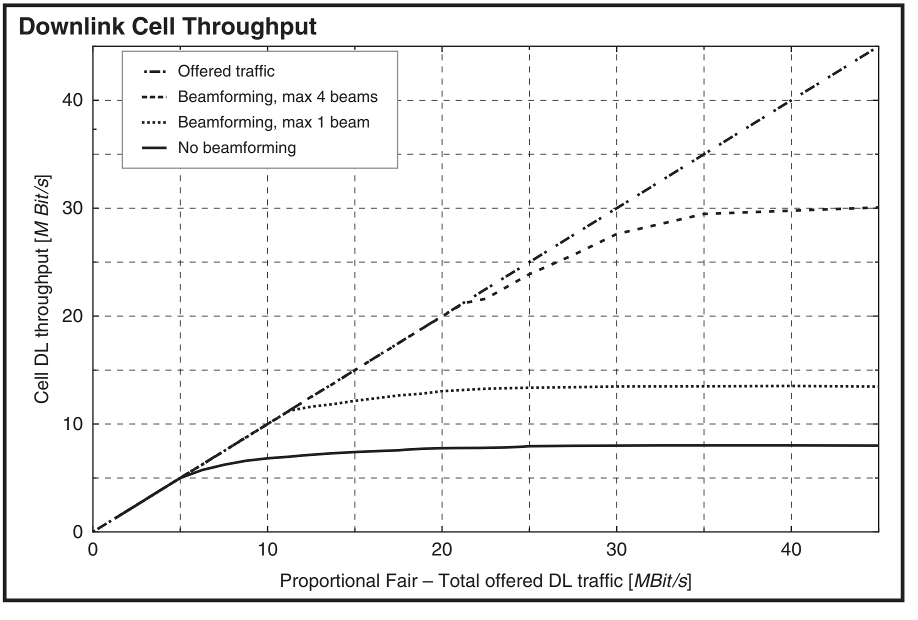


Figure 7.29 DL cell throughput for non-beamforming, spatial filtering, and SDMA.

the system runs into an overload situation, packets have to be buffered and finally discarded. Using SFIR, the intercell interference is reduced and higher order PHY modes are applicable. The increased capacity is reflected by a saturation level of approximately 12 Mbit/s. Using SDMA the base station can serve up to four stations concurrently. Although SDMA introduces intra-cell interference and although it increases the inter-cell interference, the concurrent data streams further increases the cell capacity up to a saturation level of 22 Mbit/s. Taking the saturation levels as a reference, the SFIR transmission achieves a gain of 240 % as compared to the non-beamforming case. The SDMA transmission with up to four beams achieves a more than 80 % gain as compared to the SFIR transmission. When compared with the non-beamforming transmission, it reaches an even higher gain of 440 %.

7.9.3 Signal to Interference Plus Noise Ratio

In Figure 7.30 the CCDF of the signal to interference plus noise ratio level perceived by the subscriber stations while receiving downlink data from its base station is shown. The subscriber stations' random position results in a balanced mix of available SINR values ranging from about 7 dB to more than 30 dB. All PHY-modes listed in Table b of Figure 7.20 are represented.

The beamforming reception with only one beam (SFIR) reaches the highest SINR level: The inter-cell interference from neighbouring base stations is reduced because they are tranmitting through an optimized antenna pattern.

The SINR curve for the SDMA transmission is above the SFIR case. The inter-cell interference in SDMA mode with up to four beams is higher than in SFIR mode because all neighboring cells

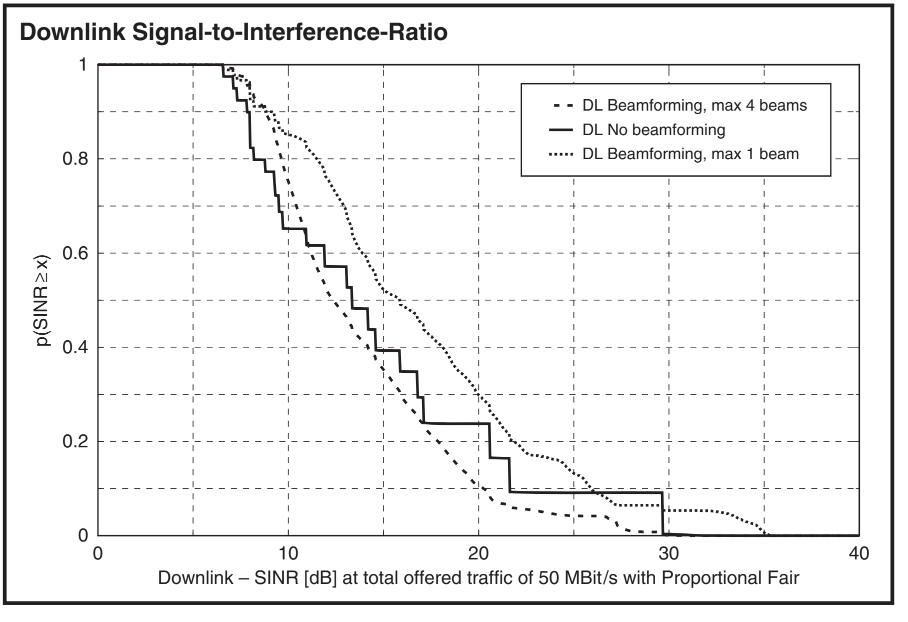


Figure 7.30 DL SINR for different transmission modes at 50 Mbps total offered traffic.

are also using SDMA for their transmissions. This means the number of interference generating transmissions in co-channel cells is up to four times as high as for the SFIR transmission mode.

When comparing the omnidirectional and the SDMA mode, it can be seen that the SINR level of the SDMA mode is partly below the non-beamforming curve. This is due to the following reason. The non-beamforming transmissions cause the highest interference which is not surprising: The omnidirectional antennas at the neighboring base stations transmit a high level of inter-cell interference. The interference level of SDMA stays below the omnidirectional case due to the optimized antenna patterns. However, the optimized (but imperfect) antenna pattern does not only reduce the interference, it might also decrease the signal strength in the desired direction. This effect decreases the SINR under SDMA operation.

7.10 Conclusion

This chapter gives a detailed introduction to the IEEE 802.16 MAC and PHY layer protocol. The MAC frame structure is introduced and the basic control elements are described. Additionally, the basic PHY layer modules of the IEEE 802.16 transmitter–receiver chain are outlined.

An analytical system performance evaluation of an example scenario is presented. Basic OFDM parameters as well as capacities of different PHY modes are calculated. Based on the PHY layer capacity the MAC layer capacity is calculated. It is figured that the MAC overhead of the IEEE 802.16 system can be assumed to be approximately 10 %. Different features of the protocol are evaluated based on a realistic example scenario. The MAC layer configuration is analyzed with respect to throughput and overhead. Optional fragmentation helps to reduce the overhead and to fill up the MAC frame, optimally. Optional features of 802.16 to resist challenging channel conditions are outlined. Their trade off, i.e., a reduced MAC layer capacity, is shown. It is demonstrated that the system can be optimized while maintaining the necessary robustness against environmental challenges.

Furthermore an SDL-based simulator is introduced that implements a prototypical IEEE 802.16 protocol including a sophisticated channel model. By means of stochastic event-driven computer simulation the MAC throughput is evaluated. The simulation results show maximum throughput and delay values obtained for the investigated scenario. The simulation results are compared and validated with previous results obtained by mathematical analysis.

It is shown that the achievable IEEE 802.16 capacity is sufficient to provide a powerful wireless last-mile technology to potential customers, even in a challenging NLOS environment.

Note

1. Reproduced by permission of © 2006 IEEE. Source: L. Berlemann, C. Hoymann, G. R. Hiertz, and S. Mangold, "Coexistence and Interworking of IEEE 802.16 and IEEE 802.11(e)," in Proc. of *IEEE 63rd Vehicular Technology Conference, VTC2006-Spring*, Melbourne, Australia, 7–10 May 2006. (Fig. 7.7).

8

IEEE 802.11, 802.15 and 802.16 for Mesh Networks

Guido Hiertz, Lars Berlemann, Harianto Wijaya, Christian Hoymann, Stefan Mangold and Bernhard H. Walke

Consumers are used to easy operable and reliable products such as Hi-Fi systems and TV sets. One step towards wireless-based enhanced products would be well accepted if there were no QoS degradation from using such technology. Therefore, future wireless home networks need high data rate with low bit error rate. With today's technology, wireless high-speed multi-megabit data rate communication is limited to the close vicinity of a station. Thus, easy deployment of network elements and ease of use are restricted. To provide sufficient radio coverage to multiple users in the home environment, forwarding and routing technology is needed. Similar arguments apply to laboratory, campus and hot-spot urban environments when aiming at wirelessly connecting mobile or movable data terminals.

Mesh networks combine wireless forwarding and routing technology to establish a new class of wireless networks. E.g., in contrast to standard IEEE 802.11 networks, which are mainly Access Point (AP) centered, mesh networks are the wireless equivalent of the Internet in that routers serve to connect links of the network to form an end-to-end route between communicating terminals. In a wireless mesh network, stations may not only be the active sinks or sources of traffic, but also may forward traffic to other stations. The wireless mesh network, by means of forwarding of data frames, serves to increase radio coverage, to enable network connectivity to a station that is outside the receive range of its intended communication partner. Additionally, the mesh network increases network link redundancy, thus supporting a more stable connectivity of stations. Since a wired backbone is merely acceptable or applicable for many customers of Consumer Electronics (CEs), and increase of transmission power to extend the radio range of stations is not desirable or not allowed by radio regulation, mesh technology is the only way to enable networking of CEs in the home environment. Moreover, mesh networking is the evolutionary path for wireless networks, in general. Mesh networks provide the freedom to place wireless devices at arbitrary locations, e.g., in a home or outdoors, without the need for a fixed infrastructure connectivity.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

Approaches to wireless mesh networks in IEEE and industry are introduced in Section 8.1 by means of mesh Wireless Local Area Networks (WLANs), mesh Wireless Personal Area Networks (WPANs) and mesh Wireless Metropolitan Area Networks (WMANs). Section 8.2 concentrates on extensions to IEEE 802.11 and IEEE 802.16 for multi-hop operation. Solutions for centrally controlled homogeneous multi-hop networks are discussed in Section 8.2.1 using the example of 802.16. A realization of a decentrally controlled homogeneous multi-hop network is outlined in Section 8.2.2 using the example of 802.11e. Approaches to heterogeneous multi-hop networks based on 802.16 and 802.11, combined, are introduced in Section 8.3.

8.1 Approaches to Wireless Mesh Networks in IEEE and Industry[1]

Several Working Groups (WGs) in the standardization body of IEEE 802 have formed a Study or Task Group (SG, TG) to develop mesh functionality. Common to all is the goal of reducing deployment costs and enabling autonomous operation. While the most advanced IEEE Mesh group is TG "s" of the 802.11 WG, proprietary solutions already exist on the market.

8.1.1 Differences between Mesh WPAN, WLAN and WMAN

A key difference between mesh WPAN and mesh WLAN is the absence of APs in the WPAN. In a mesh WPAN many more devices need to be wireless routers than in the mesh WLAN, where only APs forward traffic. Because of this, mesh WPANs do not focus on scenarios such as battlefields, disaster area communication, office and campus area networks, as these have already been defined for mesh WLANs. Since the major focus of the mesh WPAN is not on Internet access networks, no gateway to the Internet is needed.

WMAN networks work under different conditions than WPAN and WLAN. Usually, a WMAN is under control of a service provider, which has professional personnel to operate the network. The control of such a network is rather centralized and planned, and mesh concepts in IEEE 802.16 have less ad-hoc character. Although mesh deployment is being discussed, most WMANs will make use of conventional Relay (MMR) technology only. Recent research in IEEE 802.16 SG MMR considers Mobile Multi-hop Relay (MMR) technology where customer devices may enhance the radio range of a central Base Station (BS) as introduced below in Section 8.1.4.

Owing to the different application backgrounds, different technologies are used and different usage cases apply for WLAN, WPAN and WMAN technology. Therefore they need their own protocol extensions for mesh to fit the demands of an upcoming, new class of wireless networks.

Terms and definitions

Each IEEE 802 WG independently develops protocols for relaying or mesh networking support. As the terminology for single-hop operation is different with each WG, dissimilar terms are also used regarding mesh topologies. Furthermore, the protocol design impacts the capabilities of devices. Hence, IEEE 802.11 stations support other functions than IEEE 802.15.3 devices. IEEE 802.16 BSs differ in the functions supported from IEEE 802.11 APs. A precise classification of mesh-related network elements and functions is difficult. However, similar functional aspects lead to the classification shown in Table 8.1.

Table 8.1 Elements of IEEE 802 standards used for mesh.

	IEEE 802.11	IEEE 802.15.3	IEEE 802.16
Fixed User Terminal (FUT)	Station (STA) Every entity in 802.11 is an STA. STAs have the ability to compete on the wireless medium independently of any AP. Medium access is totally distributed. IEEE 802.11 does not consider mobility for classification.	Device (DEV) Devices need a PNC to operate. They are under full control of the central entity. Some DEVs may become a PNC on their own if none is available. IEEE 802.15.3 does not consider mobility for classification.	Subscriber Station (SS) An SS is under full control of the BS. It fully relies on the central entity and its resource assignments. An SS may use fixed wireless access or mobile access.
	QoS STA (QSTA) IEEE 802.11e introduces QoS supporting STAs. Under decentral access control (EDCA), QSTA may provide prioritization of traffic. With an HC, QSTA may rely on central control for QoS guarantee.	DEV DEVs fully rely on the PNC to fulfill QoS demands.	SS SSs fully rely on the BS to fulfill QoS demands.
Mobile User Terminal (MUT)	STA, QSTA Any STA or QSTA may be mobile, but there is no mobility support.	DEV Any DEV may be mobile, but there is no mobility support.	Mobile Station (MS) IEEE 802.16e amends services for MSs.
	Light-weight Mesh Point (LWMP) IEEE 802.11s introduces LWMPs that do not forward traffic but can associate to multiple Mesh APs (MAPs).	DEVs cannot associate with multiple PNCs.	SSs cannot associate with multiple BSs.
Base Station (BS)	Access Point (AP) An AP builds the central node of an infrastructure BSS. The AP forwards any frame from or to the STAs in its BSS. The AP competes on the WM like any other STA.	Piconet Controller (PNC) The PNC has full control over the WM. It assigns the right to use the WM to its DEVs. Channel access is centrally controlled by the PNC.	Base Station (BS) The BS has full control over the WM. It assigns the right to use the WM to its SSs. Channel access is centrally controlled by the BS.
	QoS AP (QAP) A QAP is able to use EDCA for channel access.	PNC The PNC provides QoS support.	BS The BS provides QoS support.
	Portal A portal connects IEEE 802.11 LAN segment with non-802.11 segments.		BS A BS may provide the bridging/routing function to connect the IEEE 802.16 WMAN with other networks.

Table 8.1 (*continued*)

	IEEE 802.11	IEEE 802.15.3	IEEE 802.16
	Point Coordinator (PC) Usually a PC is co-located with the AP. It controls medium access during the Contention Free Period (CFP). During Contention Period it operates like an ordinary STA.	PNC The PNC has full control over the WM at any time.	BS The BS has full control over the WM at any time.
	Hybrid Coordinator (HC) Usually an HC is collocated with the AP. It has highest priority when accessing the channel. It can grab the WM without backoff. For QoS guarantee it does not need to rely on CFP as it can assign Transmission Opportunities (TXOPs) to associated STAs.	PNC The PNC has full control over the WM at any time.	BS The BS has full control over the WM at any time.
	Mesh Point (MP)	Piconet Controller+ (PNC+)	Relay Station (RS)
	An MP participates in the mesh network. It forwards data and is able to calculate a mesh path. However, the MP does not provide the association service.	A PNC+ is part of the mesh WPAN. It interconnects with other PNC+, forwards data and offers access to DEVs.	An RS increases range of the BS. The RS works as extended BS and provides the BS services in an area the BS cannot cover.
Fixed Relay Node (FRN)	MP	PNC+	Fixed Relay Station (FRS)
	An MP may be installed fixed.	A PNC+ can operate immobile.	The FRS is immobile. It is usually deployed by the operator after precise frequency planning.
	MP	PNC+	Nomadic Relay Stations (NRS)
	MPs support mobility of the mesh network. An MP can participate in the Mesh WLAN when moving.	A PNC+ supports mobility.	An NRS's duration of stay is large compared to the session duration. The NRS does not forward data when roaming or moving.
Mobile Relay Node (MRN)	MP	PNC+	Mobile Relay Station (MRS)

MPs support mobility of the mesh network. An MP can participate in the mesh WLAN when moving.	A PNC+ supports mobility.	A MRS forwards data at any time including when being in movement.
Mesh Access Point (MAP) An MAP is an MP that supports the association function. It participates in the mesh WLAN and serves the STAs of its BSS as AP.	PNC+ The PNC+ provides access to the Piconet for DEVs. The PNC+ participates in the mesh WPAN and forwards data.	BS BSs may operate in mesh mode. They mutually serve as BS or associate with each other as SS.
Mesh Portal The Mesh Portal connects the mesh WLAN and other networks.		Mesh BS The Mesh BS has connection outside the mesh WMAN.
	DEV+ DEV+ participate in the mesh WPAN but do not allow DEV to associate with.	Mesh SS, Node Any other entity than the mesh BS in the mesh WMAN is a node.

8.1.2 Mesh WLAN

Driven by the terrorist attacks in 2001, several US administrations demanded new technologies that autonomously operate under decentral control. According to a field survey of McKinsey & Company in 2002, 15 % of the interviewed police officers reported that their radios failed after Tower 1 of the World Trade Center collapsed. Different from systems with central control such as the cellular technologies GSM, IS-95, UMTS, CDMA2000 or TETRA, systems with decentral control that can survive under any failures of elements in the wireless network are requested. Since then, wireless mesh networks became a serious alternative from the perspective of politicians. Alternative routes provide redundancy in mesh networks, if a device fails the mesh network may still be able to operate.

8.1.2.1 802.11s

In 2003 the Standing Committee (SC) Wireless Next Generation (WNG) of IEEE 802.11 received presentations regarding mesh WLAN. On behalf of the proposers, SC WNG requested from IEEE 802.11 WG the formation of an SG. In January 2004 the Mesh SG held its first session. Its main task was the definition of the Project Authorization Request (PAR) and Five Criteria (5C), which are needed to request formation of a new TG. From July on, “ESS Mesh Networking” became TG “s”.

Usage scenarios

The usage scenarios developed by TGs comprise a wide range of applications. Most of them consider static mesh WLAN with a fixed infrastructure:

- Military application
- Public safety
- CE in home environment
- Public access/provider networks
- Office and enterprise networks

Military application of mesh WLAN foresees ad-hoc scenarios, where combatants use the wireless network for distributed, decentralized communication among the troops. Tanks may operate as a backhaul network of the mesh WLAN providing access for soldiers. Tanks move in combat units that have low relative speed to each other. Self-healing and a redundant path for frame exchange are one of the key elements of this usage scenario.

The public safety scenario foresees establishment of ad-hoc wireless communication networks for emergency response in disaster areas. Fire engines may operate as a platform for the mesh WLAN infrastructure. Support for mobile video cameras, VoIP, positioning plans, body monitoring of firefighters and localization are key elements in the public safety scenario. As no infrastructure may be available, the mesh WLAN shall autonomously operate. Mobile battery-operated devices may enlarge the coverage area of fire units and help to interconnect firefighters inside the disaster area.

The CE application scenario foresees cheap devices that can seamlessly be integrated to existing IEEE 802.11 hardware and form mesh WLAN networks using a single transceiver in each device. Cost is a very important issue. The mesh WLAN in the home environment delivers audio/video streams and provides access to the Internet. Instead of expensive wired installation mesh WLAN provides plug-and-play. Auto-configuration and ad-hoc deployment are important elements to support.

Public access networks run by providers can be cheaply deployed using mesh WLAN technology. Unlike traditional hotspots that need connection to a wired backbone, the mesh WLAN infrastructure needs power supply only. Ease of installation outdoors and indoors is an important element for provider-operated mesh WLAN. Especially in outdoor scenarios connection to the fixed backbone may not be available in all desired areas and long-range fiber optics may be too expensive, mesh WLAN can develop new markets and hotspot areas where no service could be provided before.

Figure 8.1 shows an example WLAN-based mesh network, where APs that control a BSS are at the same time Mesh APs (MAP) and wirelessly interconnected to form an Extended Service Set (ESS) with MP1 being a portal at the same time, connecting to the Internet.

In enterprise and office networks, mesh WLANs provide more flexibility. Continuous organizational changes internal to a company require changes to its network as well. With mesh WLAN, topologies can be easily changed, and access to the company network may be provided flexibly anywhere in the office.

Terms and definitions

Each AP and its associated stations form a Basic Service Set (BSS). In its basic form, Task Group s (TGs) defines the mesh WLAN as a network of interconnected APs. The mesh WLAN spans among the BSSs. As defined in IEEE 802.11, several interconnected BSSs may form an Extended Service Set (ESS), see Figure 8.2. In the example it is assumed that BSSs use one frequency channel f_2 and the mesh that realizes the ESS is operated on channel f_1 to avoid conflicts between both types of networks, while all MPs are portal 802.11s in its simplest form permits operation of both networks on the same channel.

An ESS has a single unique SSID. In contrast to the BSSID that equals the MAC address of the AP, the SSID is maintained by the network operator. To interconnect several BSS, IEEE 802.11 uses the Distribution System (DS). IEEE 802.11s provides the means of a DS. As presented in

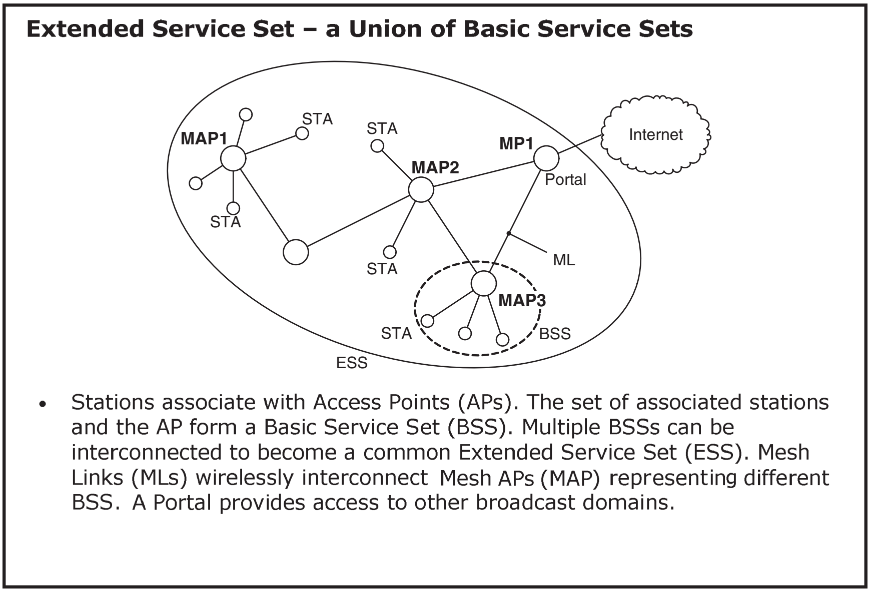


Figure 8.1 Extended Service Sets (ESSs) provide stations with the ability to roam seamlessly. The ESS appears as a single logical broadcast domain to stations. The ESS is transparent to higher layer protocols.

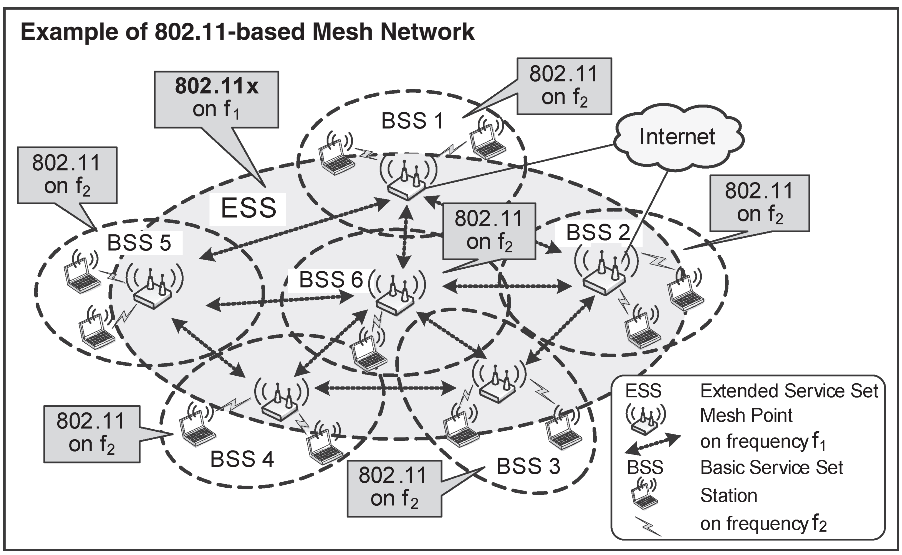


Figure 8.2 Example of an 802.11 based mesh network connecting stations served by APs of BSSs through multi-hop operation across the ESS network to the Internet.

IEEE (2005a), the term Wireless Distribution System is misleading and should not be used to address a mesh WLAN. As APs are not the only devices that may be part of a mesh WLAN, TGs has developed a well-defined set of terms and definitions.

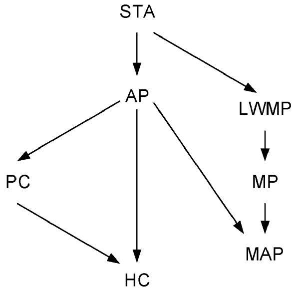
As requested by the PAR, TGs does not mandate changes to IEEE 802.11 stations. The mesh WLAN is formed among APs only. An AP that forwards frames is called Mesh AP (MAP). If the access functionality is missing, it works as forwarder only. Such an entity is called a Mesh Point (MP). Hence, all MAPs are MPs. However, not all MPs are MAPs. Figure 8.3 gives an overview. A mesh portal is an entity corresponding to a standard portal. Mesh uni- and broadcast frames are MSDUs delivered within the mesh WLAN. Between neighboring MPs, a Mesh Link (ML) is used for communication. A concatenated set of MLs from a source MP to a destination MP forms a mesh path. Each intermediate MP on a mesh path operates as immediate receiver or immediate transmitter. It uses the according address fields in the four-address scheme of IEEE 802.11 frames.

Baseline document

During standardization process, 35 proposal intents have been received by TGs. In July 2005, 15 proposals were presented. After rounds of elimination in September, in November the two remaining proposals from Wi-Mesh Alliance (WiMA) and SEE-Mesh merged. The joint proposal became the baseline document during the IEEE plenary meeting in March 2006. The mandatory set of functions includes requirements for security, path selection and Medium Access Control (MAC) mechanisms.

The mandatory MAC is based on IEEE 802.11e and uses Enhanced Distributed Coordination Function (EDCA) for arbitration of channel access. In its simplest form, a single frequency channel mesh WLAN operates as overlay to existing BSSs. The MAPs and the stations associated with the APs compete on the wireless channel. STAs and APs using EDCA are called QSTAs

Mesh Networking Technology Provides Coverage Extension



- An IEEE 802.11 device is a station (STA). In IEEE 802.11s, an STA able to associate to multiple APs, simultaneously, is called Light-weight Mesh Point (LWMP). It does not forward traffic but may route its traffic according to path selection information gathered.
- Mesh Point (MP) is the general term for a device participating in a Mesh WLAN, able to forward traffic. A Mesh AP (MAP) offers the AP functions and is an MP. A Point Coordinator (PC) or Hybrid Coordinator (HC) controls a BSS; it may be co-located with a MAP.

Figure 8.3 The basic entity in an IEEE 802.11 network is a station. In most cases, a station associates with an Access Point (AP). The AP may be co-located with a Portal that provides connectivity to non-IEEE 802.11 networks.

and QAPs, respectively. Competition among stations and their BSS serving MAP has several implications. Non-QoS stations (nQSTA) support DCF only. DCF does not support prioritization. Unlike QSTAs that support EDCA, nQSTAs have fixed backoff parameters that cannot be controlled by a QAP. When the MAP supports the IEEE 802.11e QoS mechanisms, it operates as QAP. However, with several nQSTAs in the BSS the MQAP competes with all of them on the Wireless Medium (WM). As no priority is granted by the nQSTAs to the MQAP, the downlink traffic in the BSS is affected. Furthermore, the mesh WLAN traffic cannot be segregated from the BSS traffic. Depending on the overall ESS topology, this may have a severe impact on the overall mesh performance. Bottleneck MQAP might be hindered to handle the mesh WLAN backhaul traffic, accordingly. To grant priority to the MQAP, it needs to implement Hybrid Coordinator (HC) functionality, too. Only the HC is able to apply the Hybrid Coordination Function Controlled Channel Access (HCCA), which gives full control over the WM. Hence the HC-MQAP may set up traffic streams with its QSTA and can increase its share of capacity of the WM as no backoff is used when performing the HCCA. The HC-MQAP can control its BSS in total.

The absence of backoff at the HCs, when performing HCCA, has severe impact on BSSs neighbored to each other. As other, potential QBSSs may have their own HC too, continuous mutual collisions of the HCs cannot be prevented. As all neighbored HCs access the WM after it has been identified idle for a certain time interval, their frames collide. Mutual interference of simultaneous transmissions can be expected to severely impact the BSS and the ESS mesh WLAN performance. In single frequency mesh WLAN, which needs to rely on single transceivers in each MAP, HCCA cannot be applied therefore without coexistence supporting functions. Some coexistence supporting functions are described in amendments to the Baseline Document, aiming to cover coordination of operation of MAPs.

(A) Common Channel Framework

The Common Channel Framework (CCF) is one amendment that foresees exchange of Request to Switch (RTX) and Clear to Switch (CTX) control frames. MPs may use the RTX/CTX handshake to negotiate on frequency channels for data frame transmission. MPs with a single transceiver cannot communicate if they are tuned to different channels. A Channel Coordination Window (CCW) forms a shared resource to which all MPs tune their radio at given time instants. Repetitively, MPs tune to the common channel, where they negotiate on the channel usage. To allow all MPs to make use of the CCW, synchronization among them is needed. The joint Baseline Document has optional sections that explain how to synchronize MAPs in a mesh WLAN.

MPs that have negotiated on a channel for their data frame exchange tune their radio to the new channel and sense the channel. If the WM is detected idle for a DIFS interval, frames can be exchanged. Although the availability of a channel that two MPs have agreed on cannot be guaranteed, usability of the channel is highly probable, since other MPs do not use the channel, owing to the sequential nature of the RTX/CTX handshake in the common channel. Unlike the RTS/CTS handshake, which enables immediate reservation of the channel in the time domain, RTX/CTX handshake enables reservation in the frequency domain.

(B) Mesh Deterministic Access

As a second MAC amendment, the Mesh Deterministic Access (MDA) works as a distributed WM reservation mechanism. Influenced by the Distributed Reservation Channel Access (DRCA) as defined in the Mesh Coordination Function (MCF) of the Wi-Mesh Alliance, the joint Baseline Document offers an optional reservation-based channel access mechanism that enables prediction of future channel usage. Using Information Elements (IEs) in management frames such as beacons for example, MPs negotiate with their neighbors on MDA Transmission Opportunities

(TXOPs). An MDA TXOP is called an MDA opportunity (MDAOP). An MDAOP has predefined duration and start times. At the beginning of an MDAOP, the owner has the right to access the WM using higher priority. A different set of EDCA parameters (AIFSN, CWmin, CWmax etc.) shall be used by the MP that holds the MDAOP. As the reservation is not a strict one, other stations may have grabbed the channel earlier. The MDAOP holder then defers until other transmissions end and its local Carrier Sense (CS) indicates an idle channel. Using the MDA access parameters it then competes for channel access. As with CCF, to use MDA the MPs involved in the MDAOP need to be synchronized. To further enhance the probability of successful frame reception during MDA reservations, the MDAOP information is broadcast in beacon frames and is repeated by neighboring MPs. Thus, the direct and indirect neighborhood is informed about future transmissions. The announcement of planned frame exchanges allows dealing with hidden MPs and can lower the interference. Hence, all MDA supporting MPs store information on direct and indirect MDAOP negotiations, internally. This information is used by neighbored MPs that are not involved in the MDAOP to refrain from channel access as they preset their local NAV at the start time contained in IEs referring to neighbored MDAOP. This provides priority to the MDAOP owner and lower collision probability, thus granting higher priority to the MDAOP owner. In contrast to CCF, MDA enables reservation in the time domain.

However, spatial frequency reuse better than current IEEE 802.11 cannot be achieved, because both receiver and transmitter of an MDAOP emit power in the form of data and ACK frames onto the WM. As the role of being in transmit or receive mode is interchanged (the transmitter sends data frames that are received by the receiver and the receiver sends ACK frames back to the transmitter after each successfully received data frame), interference cannot be predicted due to arbitrary data frame lengths. Fragmentation, block acknowledgments and frame aggregation may be arbitrarily used by a transmitter. The receiver replies relative to the end of a frame transmission after a specific duration. Hence, it is not predictable by other MPs when the transmitter expects feedback from the receiver. Therefore, no other MP, which may be outside of interference range to the transmitter, can reuse the frequency channel at short distance, as a transmitter may be in receive mode itself at any time.

However, MDA offers predictable channel usage that enables support for QoS in a distributed manner. Furthermore, the coordination of planned transmissions in the future allows for the usage of smart antennas that may point their beam to the transmitter at the expected point in time. MDA offers other MPs the opportunity to collaborate and cooperate. Unlike contention-based access with high probability of collision, MDA inherently works as a collision prevention mechanism. Since neighboring MPs mutually inform each other about their own, their neighbors' and the neighbors' of their neighbors intended transmissions, mutual interference can be prevented and frame transmissions have higher success probability, thus enhancing overall spectrum usage. As MPs arrange their frame transmissions, arbitration periods that are a waste of capacity can be prevented.

CTS-to-Mesh

Under discussion and currently not part of the initial draft of IEEE 802.11s is the CTS-to-Mesh concept. To overcome the impact of nQSTA whose backoff cannot be reconfigured by a QAP, the IEEE 802.11s baseline document defines several options. A simple enhancement is a CTS-to-Mesh frame. Similar to a standard CTS frame, the duration field contained sets the Network Allocation Vector (NAV) of stations that receive the frame. However, the destination address field is set to a specific value that indicates to all MPs not to set their NAV. Hence, mesh WLAN traffic can be segregated from BSS traffic and, for a period of time as announced in the CTS-to-self frame, competition on the WM is enabled among MPs. All stations refrain from channel access. This concept is similar to the ones proposed in Sections 9.2 and 8.2.2 for enabling IEEE 802.16 coexistence and 802.11e multi-hop operation.

Path selection

The IEEE 802.11s baseline document describes an "Extensible Path Selection Framework". This framework defines a single mandatory path selection algorithm that must be implemented in any MP. Other path selection methods may be vendor specific. A Protocol Identifier determines the path selection method other than the default one. The operator of a network may set this value manually. As described in Section 4.6, path selection algorithms for wireless mesh networks need additional metrics as input, different from wired networks. The TGs baseline document describes

- channel access overhead O_{ca} (depending on PHY),
- protocol overhead O_p (depending on PHY),
- number of bits B_t in a test frame (depending on PHY),
- PHY bit rate r and
- frame error rate e_{pt} for a test frame.

The Airtime Link Metric Function calculates the airtime cost c_a:

$$c_a = \left[O_{ca} + O_p + \frac{B_t}{r} \right] \frac{1}{1 - e_{pt}}$$

Airtime cost is calculated per ML. It is used as input for the Hybrid Wireless Mesh Protocol (HWMP), which is the mandatory path selection protocol. Consideration of other metrics is implementation specific. As the name indicates, HWMP combines on-demand and proactive routing protocol aspects. As an optional path selection protocol, the baseline document describes Radio Aware Optimized Link State Routing (RA-OLSR). Details of both can be found in IEEE (2006c).

Security

As mandated in the PAR, the baseline document reuses IEEE 802.11i for ML security. End-to-end security along a mesh path consisting of several MLs is beyond the scope of the baseline document. Both centralized and distributed authentication and key management are supported. With a centralized Authentication Server (AS), each MP and station authenticates with the AS. Without an AS, MPs use the distributed IEEE 802.1X authentication model, where MPs mutually authenticate. Therefore, MPs work as supplicant and authenticator. Details regarding the security concept can be found in IEEE (2006c).

8.1.2.2 Summary

The Joint TGs proposal of Wi-Mesh Alliance and SEE-Mesh covers all necessary elements for a mesh WLAN. In its basic form, IEEE 802.11s is a simple approach that has limited performance and suffers from the single-hop EDCA MAC, which has never been designed to work in multi-hop topologies. However, optional features provide the MAC with the necessary functionality to handle QoS in a multi-hop environment, to segregate BSS and mesh traffic, to prioritize mesh backhaul over local BSS frames and to make use of one or more frequency channels with one or more transceivers.

8.1.3 Mesh WPAN

Recent research enables wireless high-speed networks that provide data rates of up to 1 Gb/s. Wireless technology has become convenient and ubiquitous. However, physical limitations impact wireless high-speed networks. Coverage decreases with increasing speed, leading to reduced range communication systems. New concepts to provide service with almost full coverage are needed.

The IEEE WG 802.15 established TG5 at the beginning of 2004. TG5 shall develop a recommended practice document for mesh WPANs. Since the scope of the 802.15 WG is large, TG5 distinguishes between low- and high-rate mesh WPANs. The mesh WPANs are typically closed networks, which do not necessarily need an Internet gateway. Further, for WPAN it is very important to enhance device battery lifetime by mesh technology for low data rate mesh networks, since the mesh WPAN has a flat hierarchy with no APs.

8.1.3.1 Status of Standardization in TG 802.15.5

Currently TG 802.15.5 establishes an application scenario document reflecting the most wanted usage scenarios for mesh WPAN. Following, 802.15.5 defines the recommended practices in these usage scenarios for low and high data rate 802.15.5 will be a description independent of any PHY or MAC technology.

Proposed usage cases

There are two main usage cases defined in 802.15.5: low-rate and high-rate applications. Sensor networks, like ZigBee, and control and maintenance networks are typical examples of low-rate applications. The high-rate applications foreseen in 802.15.5 are streaming services for consumer electronics and multimedia usage. Because of the different application scenarios, TG 802.15.5 develops a single document as a recommended practice for WPAN mesh networks, which consists of separated parts where needed.

Technical requirements

The high-rate version of WPAN mesh networks supports high data rate at low latency to allow for wireless streaming services. These services include video and audio applications, real-time computer game in- and output. A high bit rate is necessary to allow for support of high QoS.

Low-rate sensor mesh WPANs must be able to support a wide coverage area. Although the amount of information exchanged in such wireless networks may be low, a sufficient, limited latency must be guaranteed to allow the usage of mesh WPANs, e.g., in house control networks comprising sensors and actuators. Since low-rate mesh devices must be cheap and easy to deploy, many devices rely on integrated batteries. Other power support may not be available, thus battery lifetime is another highly important issue. Ranging from weeks to years, sufficient battery power must be assured. As the topology of many sensor mesh networks may not change frequently, mesh routing can be based on simple algorithms. However, the number of nodes participating in the mesh might be high. Therefore computing power for routing is another important element to consider.

Streaming applications have been foreseen to dominate high-rate mesh WPANs. On the one hand, mesh technologies not only apply to well-known video or audio services, but also to a new class of interactive extremely delay-sensitive applications of online gaming. On the other hand, plain file transfer services emerge to a universal service for many high-tech devices such as camcorders, MP3-players, digital cameras and any kind of mobile hard disk. Besides the needs for high data rates, mesh network connectivity is affected by the chosen routing algorithm, since frequent topology changes will lead to a highly dynamic system. Moving users and devices, which switch on and off, demand adaptive routing strategies that are required to guarantee QoS to the most sensitive applications operated in a mesh network.

IEEE 802.15.5

IEEE 802.15.5 started in 2004. TG5 defines "Recommended Practices for Mesh Topology Capability in WPANs". Both low-rate and high-rate WPAN are considered. The outcome of IEEE 802.15.5 will be applicable for the IEEE 802.15.3 and IEEE 802.15.4 MAC.

The Call for Applications (CFA) indicated interest to "home multimedia", "interconnection among handheld devices" and "indoor location-based service" for high-rate WPAN scenarios. For the low-rate mesh WPAN, "home and office control", "security system", "medical support", "industrial/agricultural monitoring/control", "defense/homeland security" and "logistics/personnel tracking" have been identified. As low- and high-rate mesh WPANs have different backgrounds, TG5 works on different documents covering each of the application areas, separately. According to the responses to the Call for Proposals (CFP), IEEE 802.15.5 with the highest priority works on high-rate mesh WPAN.

As IEEE 802.15.3 defines a MAC protocol with central control, the Piconet Controller (PNC) has full control over all devices in its network. However, a mesh forms a totally different topology than a star topology. No central coordination instance exists. IEEE 802.15.5 develops a hierarchical approach. Several PNCs interconnect to form a mesh network. Devices associate with a PNC to participate in the WPAN. It is the PNC's responsibility to mute their associated devices during specific intervals, to let neighboring PNC and devices communicate interference-free. Thus, PNCs must coordinate access of devices in their WPAN to the WM. The approaches proposed to IEEE 802.15.5 foresee usage of beacons to mutually signal the intended channel usage.

Path selection

In its current state, IEEE 802.15.5 uses a tree-based path selection algorithm. In a tree, a single node forms the root node. Other nodes are leaves in the graph. The IEEE 802.15.5 path selection approach foresees a local view on the network. Each device forms the root node of a local tree.

8.1.4 Mesh WMAN

Mesh WMANs are used for backhaul networks that span city-wide infrastructures. Its ease of deployment makes mesh WMAN interesting to operators. Rural and city areas already use cellular technology. Central BSs coordinate access to channels in their cell and provide services to mobile devices. All BSs are interconnected through a wired backhaul network that allows terminals to roam among the BSs without session interruption. With mesh WMAN, wireless spectrum that has been allocated to a provider can be used not only to connect customers but also to interconnect BSs. Cabling or leased lines are saved and the costs of operation, i.e., OPEX, are hence reduced. For deployment of a WMAN only BS power supply is needed. Starting from a central BS connected to a wired network, remote BSs connected wirelessly may extend the range of the central BS and serve an area far away from it with high user data rate. Cheap and small devices may be used by operators to cover (shadowed) areas not covered by the central BS. As many mesh WMAN networks are assumed to be based on fixed BSs, spatial multiplexing to connect SSs to the BS within a cell and for connecting BSs to each other may be possible applying MIMO transmission and smart antenna arrays at BSs. A high antenna gain applied in the wireless backhaul network to connect BSs to each other may be useful to provide a high capacity to a remote BS, even if it is very distant from the serving BS.

8.1.4.1 802.16 Mesh Option

IEEE 802.16 mesh Networks (mesh WMANs) are intended to be used as backhaul networks. The expected ease of deployment (cost advantage of relays and scalability) makes mesh WMAN interesting for operators.

The IEEE 802.16 mesh mode is an optional feature of the standard (IEEE, 2004a). In contrast to the mandatory Point-to-Multipoint (PMP) configuration where traffic only occurs between the BS and the SSs, in the mesh mode traffic can be routed through other SSs and can occur directly between SSs as depicted in Figure 8.4. Such SSs are called mesh SSs. The BS that is connected to the backhaul network is called a mesh BS. All PDUs, i.e., data and control messages are forwarded in the time domain by the mesh SSs.

All mesh communications are in the context of a link, which is established between two nodes. Thus, the PMP frame structure composed from a downlink and an uplink subframe is replaced by a structure based on bursts scheduled for the transmission between two nodes. Figure 8.5 shows that PDUs containing user data are transmitted during bursts of the data subframe and PDUs containing scheduling control messages are transmitted during the schedule control subframe. Instead of the schedule control subframe a network control subframe is periodically included in the frame. The network control subframe provides a basic level of communication between nodes, e.g., for synchronization, initial network entry and exchange of neighborhood lists. The schedule control subframe is used for the coordination of scheduling. Access to the network and the schedule control subframe is based on a standardized scheduling algorithm so that no collisions occur (Bayer *et al.*, 2006). Depending on the configuration of the SSs, scheduling can be done on the basis of distributed scheduling, on the basis of centralized scheduling, or a combination of both.

Using centralized scheduling, the mesh BS gathers resource requests from all mesh SSs within a certain hop range. The BS determines the amount of granted resources for each link in the network both in downlink and uplink, and communicates these grants to all mesh SSs within

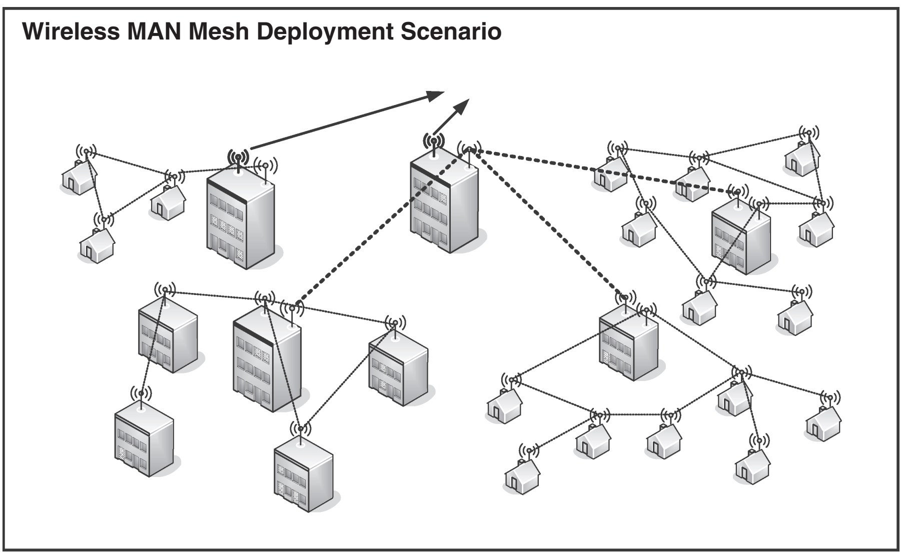


Figure 8.4 Scenario of a wireless MAN Mesh deployment.

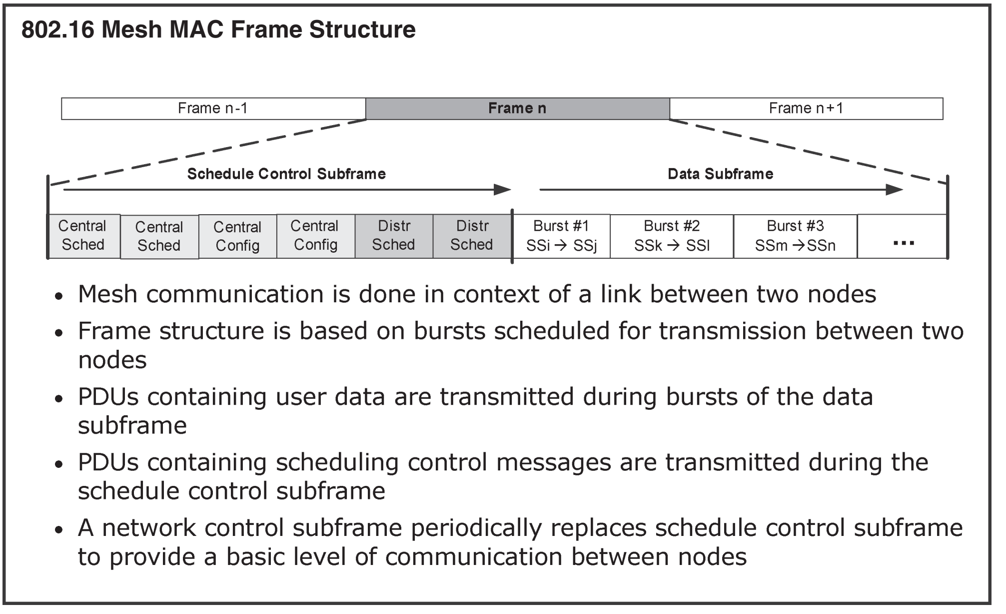


Figure 8.5 Structure of 802.16 Mesh MAC frame.

the hop range. The collection of bandwidth requests and the broadcast of the final schedule is done in the corresponding Centralized Scheduling Bursts of the Schedule Control Subframe, see Figure 8.5. However, increased overhead due to signaling overhead sent to a central device may occur.

Using distributed scheduling, all nodes including the mesh BS coordinate their transmissions in their two-hop neighborhood and broadcast their schedules (available resources, requests and grants) to all neighbors. All nodes ensure that the resulting transmissions do not cause collisions with the data and control traffic scheduled by any other node in their two-hop neighborhood. The corresponding signaling messages (three-way handshake) are transmitted either in the corresponding distributed scheduling bursts of the schedule control subframe or during a contention burst of the data subframe. If the signaling is done during the control subframe, it is called coordinated distributed scheduling. If it is done during the data subframe, it is called uncoordinated distributed scheduling. Using uncoordinated distributed scheduling collisions might affect the latency and the overhead of the signaling messages.

8.1.4.2 802.16j

The Mobile Multi-hop Relay (MMR) SG started in July 2005. According to the PAR and 5C, TGj works on an amendment for IEEE 802.16 regarding the OFDMA PHY layer described there. It enables relaying technology in the licensed band only. Such relays can be either fixed (FRS) or mobile (Nomadic Relay Station, NRS). However, in the initial deployment phase focus is on FRS. Similar to APs that are interconnected with IEEE 802.11s, TGj enables wireless communication among BSs and provides FRSs as remote BSs. The FRSs are used to enlarge coverage area of the BS. FRS and BS provide services to Subscriber and Mobile Stations (SS and MS). The fixed infrastructure helps to increase throughput, as SS and MS can use an FRS to forward data with higher SINR compared to the larger distance to be bridged without intermediate FRS. TGj expects that RS are clearly less complex than current BS. Therefore, asset and deployment cost are reduced enabling service provisioning in areas that could not be economically served before. The mesh WMAN helps to attract new customers and may enhance QoS since link quality can be improved with reduced distance. As with any mesh technology, capacity enhancements are difficult to achieve and depend on the MAC protocol and network structure. Currently, no details that allow an analytical survey are available. The concepts under development in 802.16j are similar to the relay deployment scenarios discussed in Section 10.4.

As with IEEE 802.11s, military services have also expressed interest in IEEE 802.16j mesh technology. Similar to other concepts, military services expect that most traffic is relayed over up to two hops. With three hops in total, almost all scenarios are expected to be covered. Unlike the operator-driven networks, military mesh networks are fully ad hoc and may use any frequency band. Support for Mobile Relay Stations (MRS) is mandatory and a large amount of traffic is multicast such as video and audio streaming, besides VoIP. This usage scenario has much in common with the application of MRS as studied in TGj.

MRS may be used to provide services and coverage to other devices otherwise out of range of any FRS or BS. MRSs provide connectivity to the network and enable route redundancy for devices that operate with difficult radio coverage, if roaming outside of any fixed infrastructure. However, service provided by MRS may not be reliable and fluctuation of QoS over time is likely. Furthermore, MRSs usually run on battery. Relay operation for other MS is possible but drains the battery of the MRS. As the MRS is not under control of the operator, security may be a concern. Routing tables may be affected by MRS that propagate a path to the SSs and MSs in their surroundings. Thus, end-users may have access to key elements of the mesh WMAN if MRSs are enabled. As the network operator uses a customer device to provide service for other customers, compensation to the relaying customer may be necessary. Compensation payments may be based on the amount of traffic relayed etc. However, such details may be complicated to determine. In contrast to mobile mesh WMAN for military application, usage of MRS in public networks seems to be very unlikely.

8.2 Extensions to IEEE 802 MAC Protocols – Homogeneous Multi-hop Networks

The realization of multi-hop operation is essentially simplified if the considered MAC protocol is based on Time Division Multiple Access (TDMA) with a fixed MAC frame structure such as 802.16. The contention-free medium access enables periodical and accurately timed capacity assignments to multi-hop links. Ideally, a multi-hop route can be established without requiring request of capacity for each data transmission on this route. Contrary, multi-hop operation with contention-based MAC protocols such as the DCF (EDCA) of 802.11(e) makes operation difficult, since no fixed reliable multi-hop route can be established, as the forwarding stations have to contend for access to the medium for each data frame on each hop. This makes multi-hop operation difficult in a heavily loaded network. However, in low-loaded networks contention-based medium access may lead to shorter end-to-end delay on the multi-hop route, as data incoming to an STA can be forwarded immediately, without waiting for finishing reservation of a time slot.

8.2.1 IEEE 802.16 Multi-hop Networks

Different approaches to multi-hop operation are discussed in the following using the example of IEEE 802.16. In general, the concepts introduced are standard compatible and may be also applied to similar TDMA-based MAC protocols, e.g., 802.11 and 802.15. For a detailed description of 802.16 see Chapter 7.

8.2.1.1 Multi-hop Operation in the Time and Frequency Domain

Multi-hop based on forwarding in the time and/or frequency domain is discussed in the following using the example of relaying in IEEE 802.16, see Figure 8.6. The FRSs may have a line-of-sight connection to the BS or may be served by a multipath propagation link. Gain antennas may be used to connect BS and FRS by radio. An FRS serves to increase the coverage area of the BS for serving SS.

Multi-hop in the time domain: The same frequency channel is used to connect the FRS with the BS and with its local SSs. A certain part of the MAC frame capacity is used to connect SSs and FRS and some other part is used to connect BS and FRS via a time-multiplexing channel. One transceiver only is needed at the FRS resulting in cheap, small and energy-efficient FRSs. The physical layer of the standard radio interface is not affected by this, requiring no modifications. Instead the FRS concept is realized through the MAC protocol software only.

Multi-hop in the frequency domain: This concept uses different radio channels to implement the links an FRS is operating. The two hops BS ↔ FRS and FRS ↔ SSs are operated independently of each other. This concept has a higher capacity, at the cost of increased spectrum occupancy, multiple transceivers at FRS and the frequency management. An example of such a

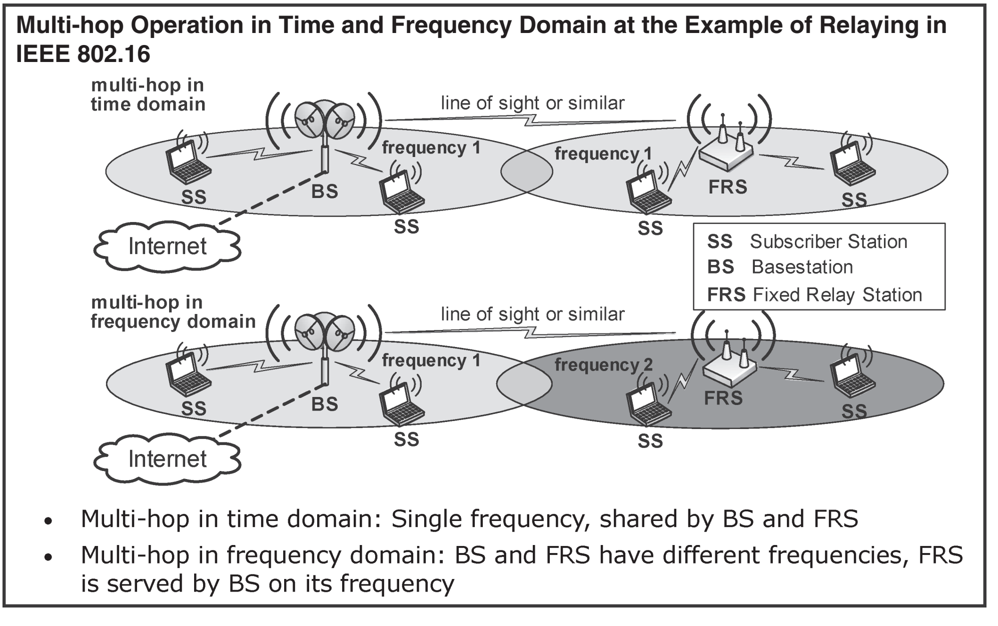


Figure 8.6 Multi-hop operation in time and in frequency domain.

network is shown in Figure 8.7 where an IEEE 802.16 mesh network with multi-hop operation in the frequency domain is depicted.

Hybrid time/frequency domain multi-hop operation: In this concept (Habetha *et al.*, 2002), the FRS needs one transceiver only and periodically switches between two frequency channels f1 and f2, allowing the BS to serve its SSs and its related FRSs using its frequency f1, while the FRS serves its terminals on frequency f2 and switches to f1 at time instants scheduled to connect to the BS. The hardware complexity is increased, since a fast channel switching must be supported.

We will focus on multi-hop operation in the time domain in this chapter.

8.2.1.2 MAC Subframe Embedding

The MAC SubFrame (SF) concept assumes one or multiple subframes embedded to the MAC frame standardized for a communication protocol. The SF concept can be applied in all MAC frame-based systems such as 802.16. The High Performance Local Area Network Type 2 (H/2) as introduced in Esseling *et al.* (2001) and Esseling (2005) has a MAC protocol similar to 802.16. The MAC subframe for one FRS embedded into the MAC frame of a BS is shown in Figure 8.8 using the example of IEEE 802.16. It can be seen that the FRS is using the same (but shortened) MAC frame structure as the BS. The FRS MAC subframe is thereby embedded as an FRS phase to the MAC frame of the BS. The SF concept enables a bidirectional Forwarding Mode (FM), where the FRS serves as intermediate node between BS and the SSs. The SS is served by the FRS for signaling and user data exchange, see bottom right-hand corner in Figure 8.8. The FM differs from the Direct Mode used for instance in H/2, where the signaling is provided by the BS and the data is directly exchanged between two user terminals (ETSI, 2001b).

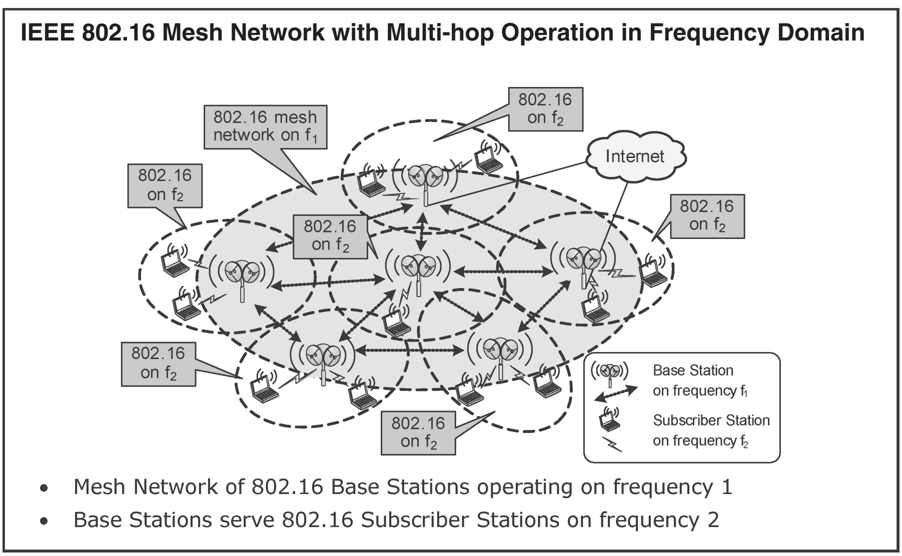


Figure 8.7 IEEE 802.16 Mesh Network with multi-hop operation in frequency domain.

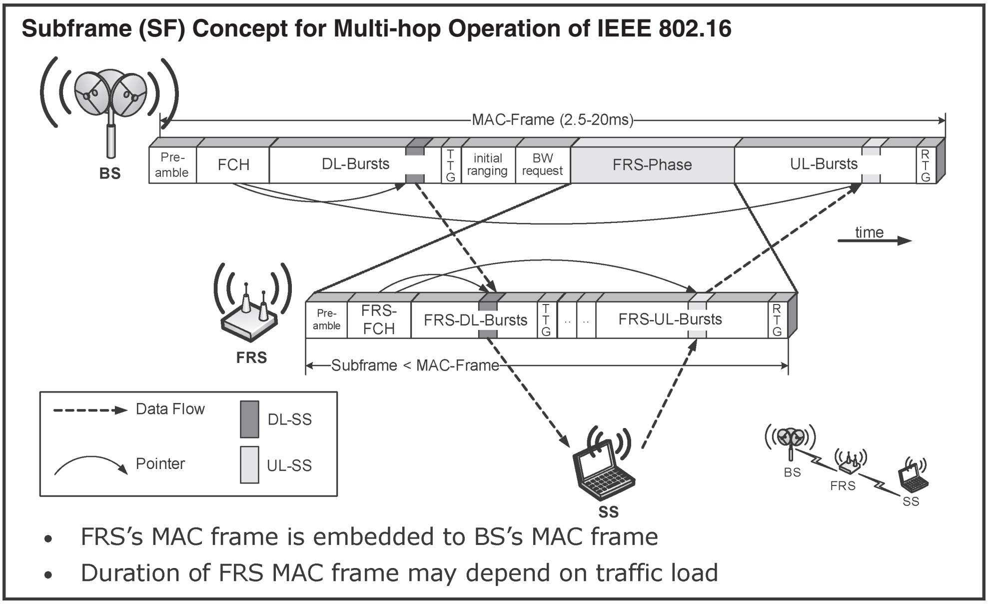


Figure 8.8 MAC Subframe concept in IEEE 802.16.

As shown in Figure 8.8, the FRS phase in the uplink (UL) phase of the BS MAC frame is reserved so that the FRS can transmit and receive the elements of its subframe without interfering the BS. Radio resources are assigned in a hierarchical way: The FRS may request radio resources from the BS by using the Bandwidth (BW) request field of the BS MAC frame, while the SS may requests bandwidth from the FRS in the same field of the MAC subframe. Besides that, the BS may assign MAC-frame-periodic slots, i.e. UL and/or DL bursts of its MAC frame, to the FRS using fixed-slot allocation, thereby establishing TDMA channels that can be detected by other BSs and FRSs and can be respected to avoid mutual interference. For a general overview on wireless resource and channel/slot allocation see Katzela and Naghshineh (1996).

The FRS MAC subframe serves to organize data transmission on the second hop. In addition, the FRS must organize communication on the first hop, behaving like a SS to the BS. Figure 8.8 illustrates a typical two-hop data flow.

As known from a single-hop 802.16 system, the BS has complete control over the composition of its MAC frame. However, after having granted a reserved space for the FRS, the MAC subframe is organized by the corresponding FRS. From the SS point of view, the FRS is a BS and from the BS point of view, the FRS acts like an SS. Apparently, this is essential for backward compatibility to legacy 802.16 systems.

The SF concept is also applicable when supporting spatial channel reuse by a BS, where multiple FRS SFs may be scheduled by a BS for simultaneous transmission.

8.2.1.3 Hierarchical Beacon with Fixed Slot Allocation

The Hierarchical Beacon with Fixed Slot Allocation (HBFSA) (Walke *et al.*, 2006; Wijaya and Zirwas, 2003; Zirwas *et al.*, 2002) in a MAC frame-based protocol such as IEEE 802.16 is illustrated

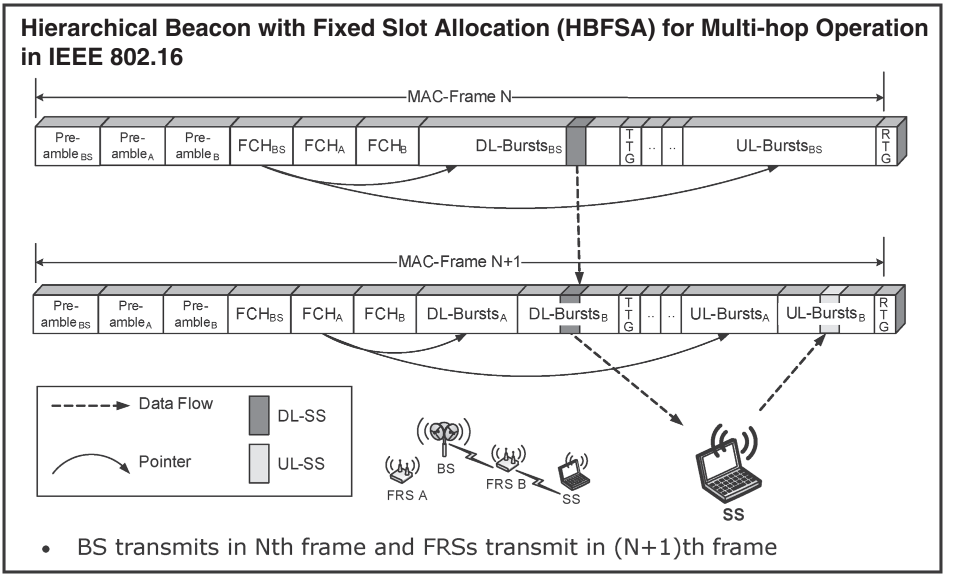


Figure 8.9 Hierarchical Beacon with Fixed Slot Allocation (HBFSA) applied to IEEE 802.16.

in Figure 8.9 where two FRS A and B are in the service area of a BS. FRS B connects an SS that is not in the radio range of the BS. If the SS (shown in the small pictogram) is an FRS C operating out of range of the BS, FRS B would serve as a relay to connect FRS C with the BS.

In the HBFSA concept, the BS and all FRSs transmit their preamble and Frame Control Header (FCH) in each MAC frame. The DL and UL bursts of N^{th}, $(N+2)^{\text{th}}$, etc. MAC frame is exclusively reserved for use by the BS, where also transmissions from and to FRS A and B are scheduled. In contrast, the $(N+1)^{\text{th}}$, $(N+3)^{\text{th}}$, etc. MAC frame is shared by the FRSs that are in one-hop distance from the BS, namely FRS A and B, to serve its SSs. FRSs with a two-hop distance from the BS, not shown here, would need a third MAC frame periodically assigned in the sequence of reserved MAC frames (Wijaya, 2005). In this way, a hop-based hierarchy of MAC frame assignment is realized.

Different from the SF concept, HBFSA requires the BS to be aware of the presence of FRSs in order to compose the MAC frame structure and sequence, adequately. From a data communication perspective, the FRSs appear to the BS as SSs and towards the SS as a BS. With increased number of FRSs overhead in the MAC frame is increased.

Since the HBFSA concept implies a fixed-slot allocation, each FRS has only a fixed amount of capacity available that makes it difficult to adapt to traffic load fluctuations.

It may be advantageous for the BS to assign multiple sequential MAC frames to one hierarchy level of FRSs following the Dynamic Slot Assignment (DSA++) concept from Herrmann (1995) and Petras (1999).

8.2.1.4 Time Sharing Wireless Router

The Time Sharing Wireless Router (TSWR) concept for homogeneous multi-hop networks (Wijaya, 2005) follows the HBFSA concept and originates from the SF concept. One MAC frame

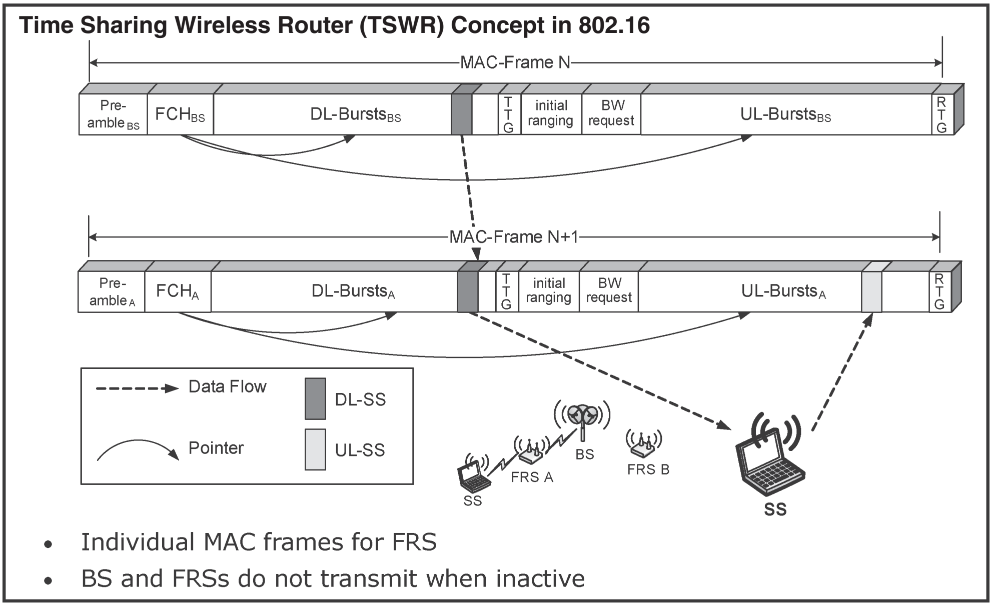


Figure 8.10 Time Sharing Wireless Router (TSWR) in an IEEE 802.16 multi-hop network.

is designated for exclusive usage by a BS or one FRS. Thus, BSs and FRSs do not transmit in periodic MAC frames not assigned to them. In the TSWR concept, the number of MAC frames allocated to different uses depends on the sum total of all active BSs and FRSs. Figure 8.10 illustrates a two-hop scenario where FRS A connects an SS with a BS as illustrated in the small pictogram. The BS is active in the first MAC frame of a periodic sequence of MAC frames, while FRS A transmits in frame $N+1$ and the second FRS B transmits in frame $N+2$. Besides a variation of the lengths of a MAC frame depending on the usage by a BS or FRS (supported by 802.16), multiple sequential MAC frames may be assigned to the same BS or FRS according to DSA++ mentioned before. The order of MAC frame assignments is repeated with every frame sequence.

8.2.1.5 Time Sharing Wireless Router with Spatial Reuse

The Time Sharing Wireless Router with Spatial Reuse (TSWR+SR) concept introduces spatial reuse to the TSWR concept, see Figure 8.11. It enables allocation of MAC frames to FRSs according to radio resource management by the BS. The concept combines an interference vector with the Dynamic Frequency Selection (DFS) *complete* measurement procedure of standard H/2 (ETSI, 2002). The TSWR + SR concept enables a multi-hop system to benefit from spatial reuse, in that MAC frames at different FRSs are being operated in parallel, if separable in the spatial channel domain (Bana and Varaiya, 2000; Kleinrock and Silvester, 1987).

In both concepts, TSWR and TSWR + SR, the BS broadcasts information on the number of MAC frames allocated per hierarchy level. Both concepts, on the one hand enable multi-hop operation with less overhead than HBFSA, but on the other hand introduce longer delay as forwarding of data packets is always delayed by at least one MAC frame.

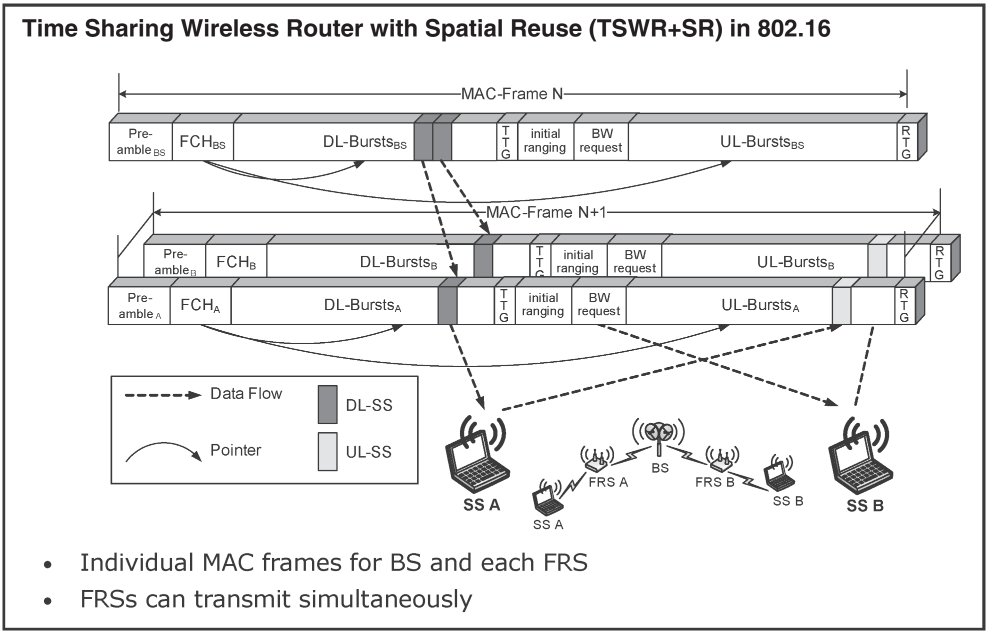


Figure 8.11 Time Sharing Wireless Router with Spatial Reuse (TSWR + SR) in an IEEE 802.16 multi-hop network.

The TSWR+SR concept requires maintaining a fixed-length interference vector, where each time slot represents the received signal strength value of the corresponding MAC frame*. The interference vector may be used by a newly deployed FRS to identify: (i) the MAC frames of the BS and (ii) candidate MAC frames for its own allocation. Thus, a spatial reuse dependent on the received signal strength is enabled. Interference awareness of signal strengths in the environment also allows using an optimal PHY mode.

8.2.2 IEEE 802.11e Multi-hop Networks

This section introduces a multi-hop concept for IEEE 802.11e that enables multi-hop operation under central control. Further approaches to multi-hop 802.11e networks can be found in Wijaya (2005). For details on the IEEE 802.11e protocol see Chapter 5.

In the following it is assumed that APs and FRSs of the 802.11e multi-hop network can operate like an 802.11e HC, able to periodically transmit beacons and to poll associated QSTAs. It is assumed further that the FRSs are immobile and always active. An FRS appears to the QSTAs associated to it as an HC granting TXOPs to it by sending *QoS-CF Poll* messages. QSTAs are assumed to be able to access the wireless medium using the EDCA. The root HC, see pictogram in Figure 8.12, assigns capacity to each FRS in its service range with the help of *QoS-CF Poll*

---

* This scheme is similar to the Mesh Distributed Coordination Function (MDCF) discussed in Section 12.2 where the signal strength of a Traffic Channel (TCH) is stored in an interference vector.

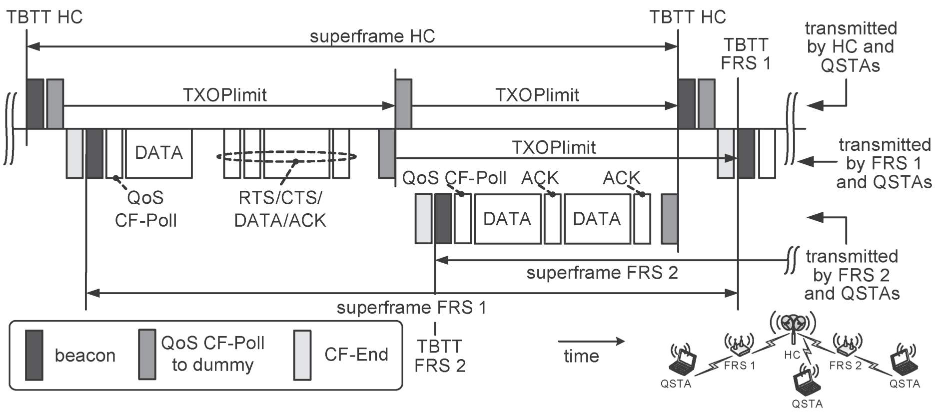
messages in preventing collisions of the FRSs' transmissions. This assigned capacity enables the FRSs to periodically broadcast their beacons to associated QSTAs. The FRS's beacon defines its superframe and restricts the duration of the EDCA TXOPs of associated QSTAs. Superframes are specified by the FRS based on the Transmission Opportunity Limit (*TXOPlimit*) value received from the root HC as part of the *QoS-CF Poll* message. This value restricts the transmission duration of any QSTA to stay within the superframe specified by the local HC/FRS. In general, an HC may poll an FRS once or several times per superframe. In the case of multiple polls, the FRS can decide on its own when to transmit its beacon to its associated QSTAs.

Two kinds of collisions may occur: (i) among QSTAs using EDCA and (ii) between QSTAs using EDCA, and the poll messages from HC/FRS. In order to prevent these collisions, algorithms for channel reservation are discussed in the following.

8.2.2.1 Collision Avoidance through Channel Reservation

Stations in the coverage area of an IEEE 802.11e multi-hop network need a common synchronization in order to support collision-free multi-hop operation. Without a common synchronization, HC and FRSs may interfere if hidden to each other, see Section 4.4.1. Channel reservation for multi-hop operation in the time domain introduced in the following is referred to as Collision Avoidance through Channel Reservation (CACR). The CACR concept (Wijaya, 2005) enables a reserved channel shared by the HC, FRSs and QSTAs in time domain. It prevents both collisions among QSTAs served by different FRSs and collisions between QSTAs served by an FRS and QSTAs served by a HC. Collision avoidance is achieved by transmitting *QoS-CF Polls* by the root HC to the FRS and by polling a dummy QSTA as depicted in Figure 8.12.

Collision Avoidance through Channel Reservation (CACR) Concept for Multi-hop IEEE 802.11e



- QoS CF-Poll to prevent QSTAs to access the medium in distributed way
- CF-End to clear medium "lock" in reception range of FRS/HC from QoS CF-Poll

Figure 8.12 Collision avoidance through Channel Reservation concept for multi-hop IEEE 802.11e networks.

Both together prevent all QSTAs in the area served by HC and FRSs from attempting a transmission on their own and allow the coordinated polling of QSTAs.

When an FRS is polled by a root HC, it activates its service area by transmitting a *CF-End* message. Thereafter, it broadcasts a beacon including the *TXOPlimit* for EDCA operation. Before expiration of its TXOP assigned by the HC, the FRS transmits a *QoS-CF Poll* message addressed to a dummy QSTA with an address not used in the service area. On reception of the *QoS-CF Poll* the associated QSTAs set their Network Allocation Vector (NAV) timer and put themselves in silent mode. The *TXOPlimit* should be large enough to keep the QSTAs in the service area of the FRS in silent mode until the FRS is polled next by the HC. The optimal value for this *TXOPlimit* is the Target Beacon Transmission Time (TBTT) of the next beacon transmitted by this FRS.

8.2.2.2 Collision Avoidance by Channel Reservation with Spatial Reuse

The CACR concept can benefit from spatial reuse similar to the TSWR + SR concept introduced in Section 8.2.1.5. Here, each slot of the interference vector, maintained locally by the HC/FRS, represents 16 μs time duration, since a TXOP has a duration of a multiple of 16 μs (Wijaya, 2005). The Collision Avoidance by Channel Reservation with Spatial Reuse (CACR + SR) concept is realized as follows.

Scan by FRSs scan for *QoS-CF Poll* messages dedicated to any FRS transmitted by the HC to assign channel capacity. If the corresponding message sequence *CF-End*/beacon/*QoS-CF Poll* is received by a FRS the respective slots of the interference vector according to the TXOP duration are marked occupied and are not considered for spatial reuse anymore. If only a *QoS-CF Poll* message dedicated to a FRS is received but no *CF-End*/beacon/*QoS-CF Poll* message sequence follows in due time, the channel may be used locally and the respective time slots in the interference vector are marked free. In the case of not receiving any *QoS-CF Poll* message or when only receiving reply messages, spatial reuse cannot be applied.

The introduced concepts to multi-hop operation in IEEE 802.11e networks are compared in the following to concepts for 802.16 multi-hop networks explained in Section 8.2.1.

8.2.3 Performance Evaluation Results

The various multi-hop concepts for homogeneous networks discussed in the previous two sections, either using the example of 802.11e or 802.16, are compared in the following through a stochastic and event-driven performance simulation. Additionally, the simulation results are verified by mean value analysis as explained in Wijaya (2005). A detailed description of the simulation environment can also be found there.

The simulation parameters are summarized in the table contained in Figure 8.13: Two channel models are used in the simulative performance comparison, namely the free-space propagation model from Mangold et al. (2001a) and the Multi Wall and Floor (MWF) model from Lott and Forkel (2001), see Chapter 2. The modem assumed to represent the PHY is as specified in 802.11a.

8.2.3.1 Scenario Description

The performance comparison is done here using the example of a two-hop scenario with one BS/HC, four FRS and four immobile SSs/QSTAs as illustrated in Figure 8.13. The BS/HC and SSs/QSTAs are separated through buildings and the FRSs have a line-of-sight communication with the BS/HC

Two Hop Scenario with Four Fixed Relay Stations

Simulation Parameters

Parameter	Value
Traffic	CBR, up to system capacity
Distance HC/BS - FRS	15 m
Distance FRS - SS/QSTA	5 m
Channel Model	free space (γ=2) and MWF
Wall Type, Thickness, Loss	concrete, 20 cm, 29 dB
MAC burst size, packet size	54 bytes, 2000 bytes
TX Power	200 mW
PHY Mode	64 QAM ¾

Simulation Scenario:
Relay Enhanced Cell

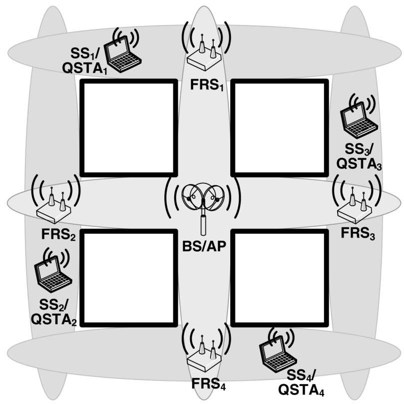


Figure 8.13 Scenario description and simulation parameters for the comparison of various multi-hop concepts in 802.11e and 802.16.

and SSs/QSTAs. The radio coverage of the BS/HC and FRSs is also highlighted in Figure 8.13. A duplex communication between BS/HC and SSs/QSTAs is assumed with a Constant Bit Rate (CBR) traffic load. The FRSs do not generate any traffic themselves and serve as relays for the BS/HC only. Resource allocation is coordinated by the BS/HC and limitedly delegated to the FRS.

In the SF concept, the MAC frame is equally shared between the BS and the four FRSs: Each gets a fifth of the complete 802.16 MAC frame for data transmission to associated SSs. In view of the BS, the FRSs are considered to be SSs. The FRSs request capacity from the BS for generating a subframe if associated SS have data to transmit. If the BS is not able to assign an adequate capacity to the FRS, the FRS is not able to generate the subframe.

Following the HBFSA concept, the BS periodically assigns one MAC frame to the FRSs that equally share this MAC frame, and BS and FRSs transmit their beacons in each MAC frame.

In the TSWR concept, the BS assigns an individual MAC frame to itself and one MAC frame to each FRS for exclusive usage. Thus, the number of assigned MAC frames equals the number of operating FRS plus the BS. Spatial reuse is realized in this scenario following the TSWR + SR concept through simultaneous transmission of FRSs. Here, the pairs $FRS_1 + FRS_2$ and $FRS_3 + FRS_4$ operate at the same time, separated through buildings. This represents operation where the FRS enhanced cell, shown in Figure 8.13, is separated in the area to implement a cellular area coverage.

The *TXOPlimit* for a FRS is assumed to be 4448 μs, for QSTAs it is 2064 μs and for the TBTT it is 8896 μs. In the CACR concept, the HC polls the FRSs and the FRSs poll associated QSTAs in order to enable user data transmission from the HC to the QSTAs. Thereby, the HC can poll within one superframe the four FRS in sequence. Spatial reuse is enabled by the HC through the sequential polling of the FRSs for simultaneous transmission. Again, the pairs $FRS_1 + FRS_2$ and $FRS_3 + FRS_4$ operate at the same time. The QSTAs are able to transmit multiple data packets within one TXOP assigned to it by the FRSs.

8.2.3.2 Mean Delay vs. Offered Traffic

The mean delay of 802.16 MAC bursts and of 802.11e data frames versus offered traffic load is shown in Figure 8.14 for the scenario depicted in Figure 8.13. The offered traffic between the HC/BS and each SS/QSTA is increased in 0.5 Mbit/s steps for both directions of each duplex connection.

In general, the characteristics of the delay graphs are similar for the different multi-hop concepts since an equal fraction of capacity is assigned to the BS/HC and FRSs. The saturation load representing the throughput capacity is therefore inversely proportional to the number of FRSs served. Before reaching saturation, the mean delay stays constant with increasing offered traffic. When approaching saturation, mean delay increases rapidly owing to more and more frames waiting in the queues.

The SF concept achieves the lowest mean delay in 802.16 as the BS assigns a fraction only of the MAC frame for multi-hop operation instead of a complete MAC frame. This enables a fast forwarding of data by the FRS received from the BS or SS. In contrast, the throughput capacity of the SF concept is the lowest among all multi-hop concepts. The reason for this is the large overhead introduced by the SF concept resulting from four complete subframes realized in one MAC frame.

The HBFSA concept leads to lower mean delay but lower capacity than TSWR, since with HBFSA the BS assigns only one MAC frame common to four FRSs in one-hop distance. While TSWR assigns one MAC frame exclusively per FRS. Thus, four MAC frames are assigned and SSs on average must wait longer.

The throughput capacity of the HBFSA and TSWR concepts is lower than that under CACR, but CACR delay under low to medium load is similar to TSWR – quite large.

When exploiting spatial reuse, throughput capacity substantially increases: throughput capacity gained under TSWR + SR for 802.16 is something higher than that under CACR + SR for 802.11e.

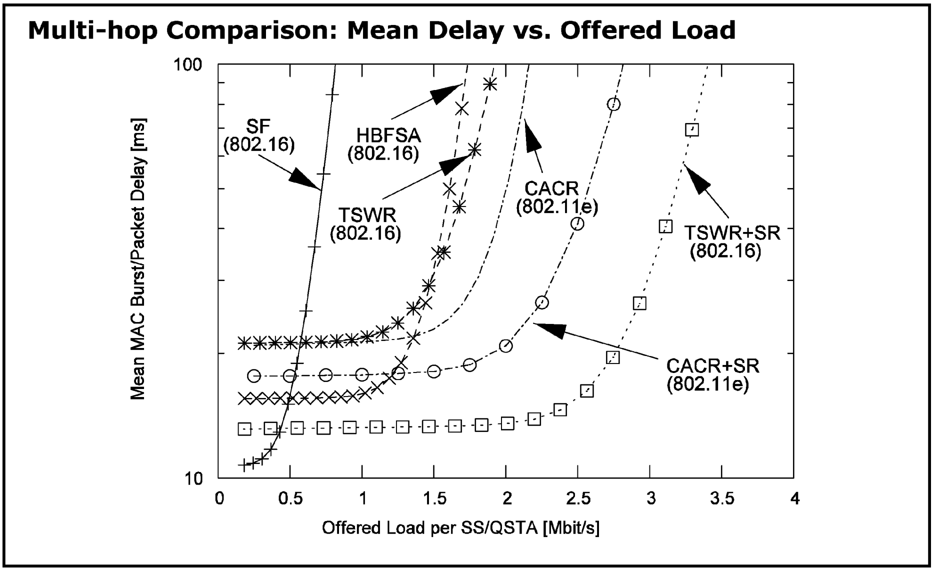


Figure 8.14 Comparison of various multi-hop concepts in 802.11e and 802.16. The mean delay vs. offered load in the downlink (BS → SS, HC → QSTA) is evaluated.

8.2.3.3 System Capacity vs. Distance between BS/HC and FRS

The system capacity in terms of maximum achievable system throughput depends on the PHY mode and the distance between the BS/HC and the FRSs. The dependency on this distance is shown in Figure 8.15 for a fixed PHY mode of 64 QAM 3/4. The analytic results are derived in Wijaya (2005).

In general, the system capacity decreases with increased distance between BS/HC and FRS due to the decreasing SNR and related BER (Khun-Jush *et al.*, 1999).

The CACR concept introduced for 802.11e has a system capacity higher than that of the SF, HBFSA and TSWR concepts proposed for 802.16. The reason is the small number of SS/QSTAs assumed lead to only a few collisions. Further, with increasing number of FRSs served by an HC, the overhead owing to *QoS-CF Poll* messages does not affect system capacity much, but the overhead in the 802.16 multi-hop concepts would affect capacity much more.

Spatial reuse according to TSWR + SR and CACR + SR concepts results in the highest system capacity in the respective systems.

8.2.4 Summary

Various multi-hop concepts for homogeneous 802.11e and 802.16 wireless networks are discussed in this section. All these concepts are applicable under software control without changing specifications of the respective standards.

The SF concept shows the lowest delays in general but is inefficient for multi-hop operation with multiple FRSs. The HBFSA approach is easy to implement since the BS and FRSs transmit their FCHs within each MAC frame allowing SSs a fast and easy association with the network.

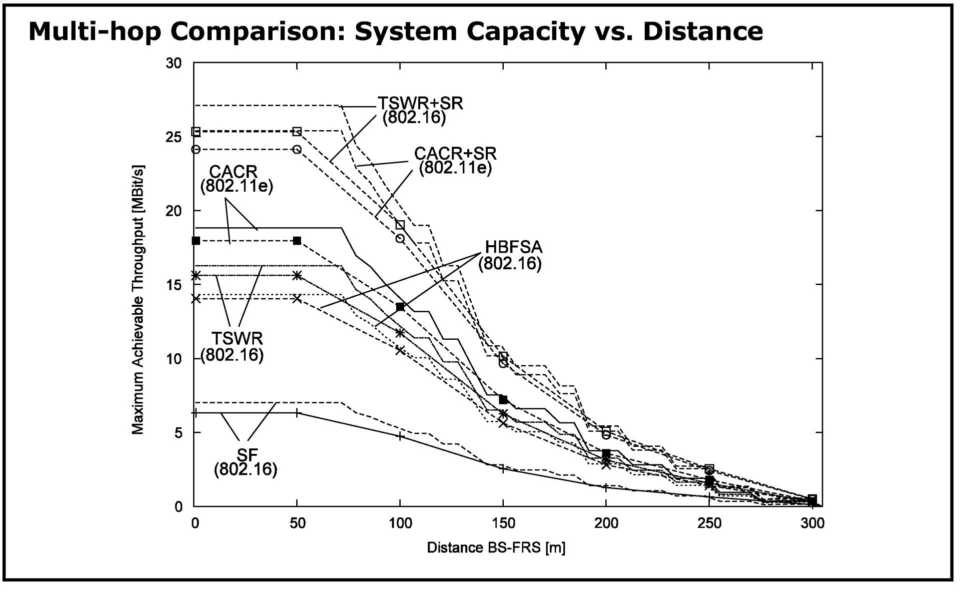


Figure 8.15 Throughput capacity versus distance between BS/HC and FRS for various multi-hop concepts applied to 802.11e and 802.16. The transmit power used is 200 mW.

However, the overhead due to the HBFSA concept is relatively large and increases proportionally with the number of FRSs. This concept ensures a good tradeoff between multi-hop and multi FRS performance (Wijaya, 2005).

Both TSWR and TSWR + SR concepts outweigh the shortcomings of the SF and HBFSA. The TSWR + SR approach benefits from spatial reuse and is a promising candidate for implementation in urban and hotspot areas with many obstacles. Its self-configuration and self-organization capabilities considerably reduce the implementation efforts compared to the subframe and MAC frame restructuring of the SF and HBFSA concept.

The CACR and CACR + SR concepts guarantee in 802.11e a collision-free multi-hop operation. The performance of these two concepts depends on the TXOP duration and packet lengths. It is therefore important to adopt packet length and TXOP duration to reduce the overhead from ACK messages owing to segmentation and to avoid partial utilization of a TXOP. The TXOP duration designated for multi-hop operation (in assigning TXOPs to FRSs) determines the performance of an 802.11e multi-hop network.

8.3 Extensions to IEEE 802 MAC Protocols for Heterogeneous Multi-hop Networks

WMANs based on the IEEE 802.16 standard are an upcoming competitor for wired last-mile access systems, since they realize a fixed wireless broadband access. Especially in rural areas, where it is too expensive to deploy wired networks due to marginal density of population, 802.16 is seen as a promising technology. Various scenarios can be thought of, where 802.16 may share spectrum with 802.11-based WLANs in office or residential deployment scenarios. Additionally, 802.16 is expected to be deployed to provide a multi-hop, relay-based wireless backhaul serving 802.11 WLAN APs operated in hotspots. Multi-mode relays supporting both operation of 802.16 and 802.11 are required, which may take advantage from offering interworking between both standards. The frame-based medium access of 802.16 requires rigorous protection against interference from WLANs in order to operate properly.

8.3.1 Overview

This section starts in Section 8.3.2 with an introduction of options for realizing heterogeneous mesh networks consisting of different radio access technologies for the mesh network and for connecting users to it. As a concrete realization example of a heterogeneous multi-hop network, an interworking concept for integrating the MAC protocols of 802.11 and 802.16 into each other, based on the proposal by Berlemann *et al.*, (2006a) is discussed thereafter in Section 8.3.3. One single communication device, capable of operating both standards, centrally controls coordination of access to the same channel and controls interworking of 802.16 and 802.11.

Operation of IEEE 802.11b and 802.16a in a shared frequency channel is evaluated in Jing and Raychaudhuri (2005). They apply a Common Spectrum Coordination Channel (CSCC) as described in Section 13.5.1 for exchanging control information on transmitter and receiver parameters in order to cooperatively adapt key parameters of the PHY. The CSCC is realized on the basis of 802.11b at an edge of the 2.4 GHz band and all 802.16a/802.11b stations must monitor the CSCC to coordinate their spectrum access. Different from that, the interworking concept proposed in the following does not require any extensions to stations or MAC protocols for establishing central coordination of ordinary 802.16/.11(e) stations for data transfer between these. This concept is able to guarantee QoS support to applications as usual with an 802.11e network under HCCA.

8.3.2 Medium Access Control in Heterogeneous Mesh Networks

Today's wireless systems are developed for specific application scenarios and have therefore different ranges of coverage. Multi-hop operation is suited to extend the radio range as motivated in Chapter 4 and described in detail in the context of cellular multi-hop networks in Section 10.4.1.1. Longer range systems may be used to mesh shorter range systems in order to increase, efficiently, the coverage area and the capacity available.

802 systems will become (and 802.11 is) part of the cellular world and vice versa. The WG 802.21 of the IEEE specifies therefore the "interworking" of 802 systems (fixed and wireless) and of cellular systems as outlined in Section 11.1.2.2. The coexistence and mutual support, beyond the scope of WLAN cellular interworking is discussed in the following sections. Different options for realizing a long-range Mesh Network (MN) that connects the APs of a short-range network are introduced. The MAC of two different radio access technologies to the wireless medium shared in the time domain is considered. All proposals have in common that a heterogeneous multi-hop operation is enabled with different radio access technologies for the wireless mesh and services provisioning links.

8.3.2.1 802.11 Mesh Network to Serve 802.11 Stations

The MAC frame composition of a heterogeneous multi-hop network operating under the 802.11 PCF mode is depicted in Figure 8.16. The MPs operate as 802.11 PCs and introduce the overlaying MAC protocol (Wijaya, 2005; Zhao *et al.*, 2006): Beacons transmitted by the PCs in the 802.11 frame format inform 802.11 stations of the Contention Free Period (CFP) and the Contention Period (CP) durations. Guard times at the end of the CP guarantee a timed initialization of the next MAC frame in preventing delay or interference. Mechanisms for realizing such a guard time by means of 802.11 and 802.11e are introduced below and in Section 9.2. The MN may be realized in using any appropriate (ideally a reservation-based) protocol, e.g. 802.11s. The MN operates during the CFP while 802.11 STAs may access the shared medium during the CP. In this way an interference-free coexistence of the MN with 802.11 stations is realized.

8.3.2.2 802.16 Mesh Network to Serve 802.11 Stations

Different from the previous section, Figure 8.17 introduces the MAC frame composition of a heterogeneous multi-hop network operating under 802.16 control. The MPs of the MN operate as 802.16 BSs and the MAC frame is based on 802.16. The BS, by transmitting an 802.11 beacon, informs 802.11 STAs of CFP and CP durations. In placing the 802.16-based operation of the MN in the CFP and 802.11 operation in the CP, too, an interference-free operation of both, 802.16 mesh and 802.11 stations can be guaranteed to implement coexistence of these systems. Guard times are required for delimitating the CP to enable in time initialization of the 802.16 MAC frame. The detailed protocol is described in (Berlemann 2006f).

8.3.2.3 New Mesh Network Protocol to Connect 802.16 BSs

The usage of a new mesh network protocol for providing meshing of APs and relays is illustrated in Figure 8.18. The new mesh protocol introduces the basic MAC frame structure and is used in the MPs and relays of the MN (Mangold *et al.*, 2001c). While the MN is realized in using any

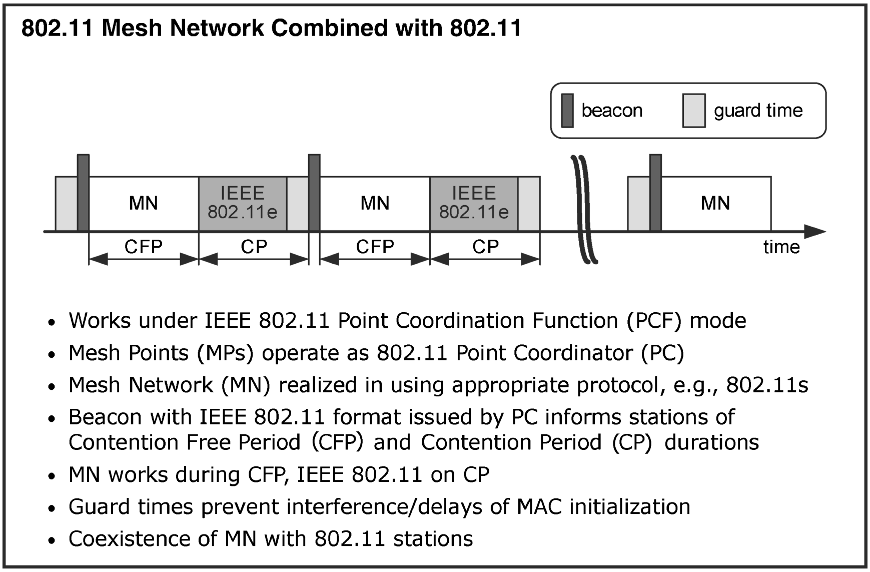


Figure 8.16 MAC frame composition for a Mesh network combined with an 802.11 system. The basic MAC frame is defined in 802.11.

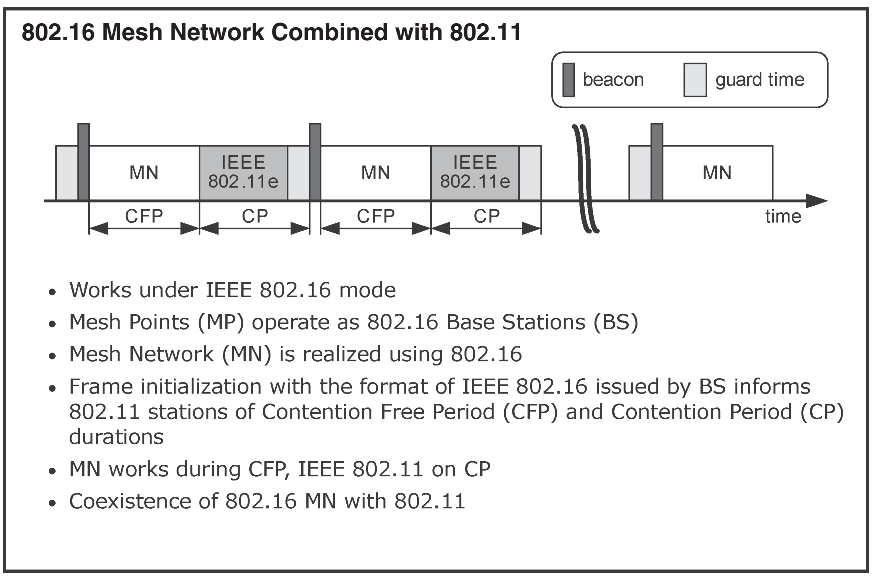


Figure 8.17 MAC frame composition for an 802.16 based Mesh network connecting 802.11 stations.

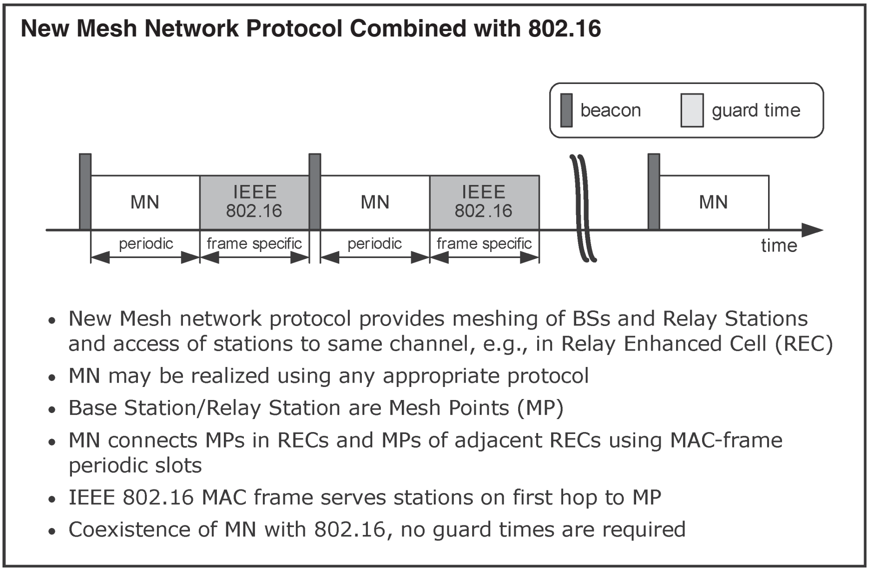


Figure 8.18 MAC frame composition based on a new Mesh network protocol combined with 802.16.

appropriate MAC protocol, 802.16 is used to serve stations on the first hop to the MP. The MN connects MPs in using periodic MAC frames easing the coexistence with 802.16. Applying 802.16 for the first hop has the additional advantage that no guard times are required as all frame transmissions are under the control of the BS. Promising candidates for the new mesh network protocol are for instance the Mesh Distributed Coordination Function (MDCF) introduced in Chapter 12 or the MAC protocols developed in the IST-project WINNER (Mohr, 2005; WINNER, 2005).

8.3.3 Interworking Control of 802.16 and 802.11[2]

The new interworking concept described in the following goes beyond the cooperation of heterogeneous systems described in Section 8.3.2.2. It allows operation of 802.11 and 802.16 at the same frequency channel and is based on the idea of the Central Controller Hybrid Coordinator (CCHC), see Mangold *et al.* (2001b) and Mangold (2003), where the H/2 Central Controller (CC) is combined with the Hybrid Coordinator (HC) of 802.11a/e to realize interworking of both. There, the H/2 MAC frames are scheduled as Controlled Access Phases (CAP) within the 802.11e superframe. The CCHC concept has been further developed by Wijaya (2004; 2005) when introducing a Wireless Router (WR) for multi-hop operation as discussed in the previous section.

IEEE 802.16 and 802.11 MAC protocols have fundamental differences. While 802.16 is a frame-based MAC protocol with central control, 802.11 applies distributed control with contention-based medium access but also may apply contention-free channel access under central control. Both 802.16 and 802.11 use a similar OFDM-based PHY facilitating interworking.

To meet the requirements of 802.16 SSs, the new concept applies a central controller device. Integration of 802.16 and 802.11 implies interworking between similar and different types of devices. The central controller device combines the BS function of 802.16 with the HC function of 802.11e (that is similar to the Point Coordination Function, PCF, of 802.11) and is thus referred to as Base Station Hybrid Coordinator (BSHC). The BSHC is capable of operating both MAC protocols, 802.16 and 802.11. Interworking is realized by declaring 802.16 MAC frame transmissions to happen during contention-free phases and by integrating 802.11 frame transmissions into the 802.16 MAC frame. Optionally, a contention period may also be included in an 802.11 superframe. From the perspective of an 802.16 SS, the BSHC is a BS transmitting its beacon (frame initialization) periodically with some time gaps in between, while QSTAs regard the BSHC an ordinary 802.11e HC.

8.3.3.1 Scenario

The interworking scenario in Figure 8.19 shows one 802.16 BS centrally controlling channel access. The BS assigns capacity (a certain period of time) to one BSHC and delegates the responsibility of coordinating channel access within this time duration to the BSHC. In doing so, the BSHC in the view of the BS is an SS, as described in Section 8.2. The BSHC itself controls channel access of 802.11e stations by assigning TXOPs or alternatively offering a limited time for contention-based access according to the 802.11e protocol. The BSHC is also serving 802.16 SSs by scheduling the DL/UL bursts provided in its 802.16 MAC frame. The BSHC is connected via a multi-hop relay link to the 802.16 BS and appears to the BS as an ordinary SS. In this way, the BSHC is a heterogeneous relay able to connect the 802.16 BS and 802.11e QSTAs. Legacy 802.11 STAs may transmit, too, if a contention period is provided by the BSHC superframe.

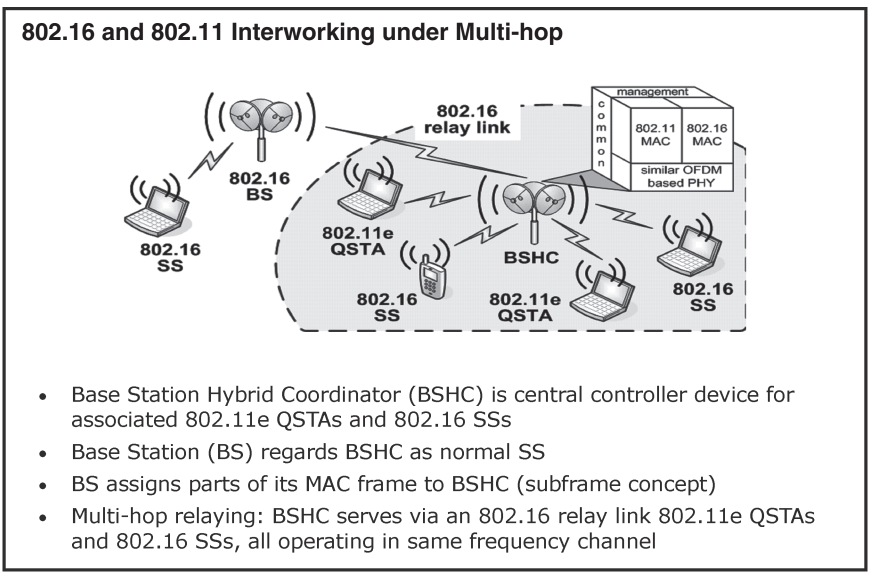


Figure 8.19 Interworking scenario of a BSHC serving both 802.11e QSTAs and 802.16 SSs, where the BSHC itself is a Fixed Relay Station served by an 802.16 Base Station. All the links are operated on a common frequency channel.

In the interworking scenario of Figure 8.19, the BSHC provides broadband access to QSTAs and SSs. For communication between 802.11(e) QSTA and 802.16 SS, the BSHC receives data during a TXOP from the 802.11(e) station and forwards it in the subsequent 802.16 MAC frame to the 802.16 SS and vice versa. A protocol stack at the BSHC combining both MAC protocols of 802.16 and 802.11(e), as illustrated in the right-hand upper corner in Figure 8.19, with a MAC bridge function makes this possible.

8.3.3.2 Medium Access Control

In order to coordinate operation of 802.16 and 802.11 networks within its subframe, the BSHC operates at one single frequency channel controlling medium access of all 802.16 SSs and 802.11 STAs. The full control of the BSHC over the channel enables it to support QoS requirements of applications. The control by the BSHC is realized through assigning the channel for a well-defined duration to one SS or STA at a time. Comparable to the polling of stations in 802.11e, the STAs/SSs decide on their own which data to transmit when they get a transmit resource allocation period assigned from the BSHC.

A MAC frame composition for realizing an interworking of 802.16 and 802.11e is shown in Figure 8.20. The MAC frame is introduced and broadcasted by the 802.16 BS. Thus the mesh network is based on 802.16 as introduced in Section 8.3.2.2. The BS assigns a part, i.e. a subframe, of its MAC frame to the BSHC and delegates the control over this period of time to the BSHC in taking up the subframe concept from Section 8.2.1.2. The BSHC's subframe is introduced and protected with the means of 802.11e following the proposal from Section 8.3.2.1. The subframe can be regarded as an 802.11e superframe consisting of a CFP and CP. The BSHC's 802.16

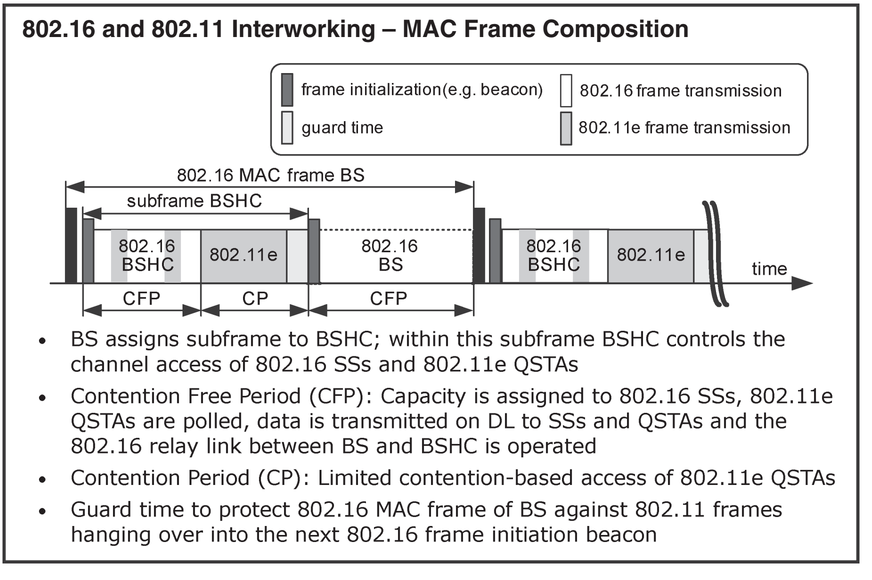


Figure 8.20 Composition of the MAC frames for realizing an interworking of 802.16 and 802.11e. 802.16 BS assigns a subframe to the BSHC that is completely scheduled by the BSHC to serve 802.16 SSs and 802.11e stations. The BSHC in view of the BS is an SS.

MAC frame is part of the CFP and 802.11e frame transmissions may be scheduled as bursts of this 802.16 MAC frame, as shown in Figure 8.20.

The MAC subframe structure controlled by the BSHC is explained in detail in Figures 8.21 and 8.22. Transmissions related to 802.16 are visible by frames filled white, while the 802.11 transmissions in frames filled gray. Transmissions by the BSHC are depicted above the timeline and transmissions by SSs/STAs are depicted below.

Contention-free access

An 802.16 MAC frame must not be interfered or delayed by any 802.11 stations. Different from 802.11, the 802.16 MAC protocol is not prepared to handle any 802.16 frame start delay. To guarantee a timely start of 802.16 MAC frames, the BSHC reserves resources for it in its CFP. The CFP starts with the 802.11 beacon transmission and ends with a *CF-End* message, leaving sufficient time for transmitting the beacon starting the next 802.16 MAC frame, if no CP follows, see Figure 8.21.

802.11 STAs see a periodic superframe marked by beacons in Target Beacon Transmission Time (TBTT) distance. A superframe comprises the CFP and a CP. Information elements in the beacon indicate the superframe duration and the time instant to start the CFP, if applicable. Further the *TXOPlimit* and additional EDCA parameters are broadcast by the BSHC in the beacon to control the EDCA's contention-based access. The 802.16 preamble and FCH are transmitted as usual.

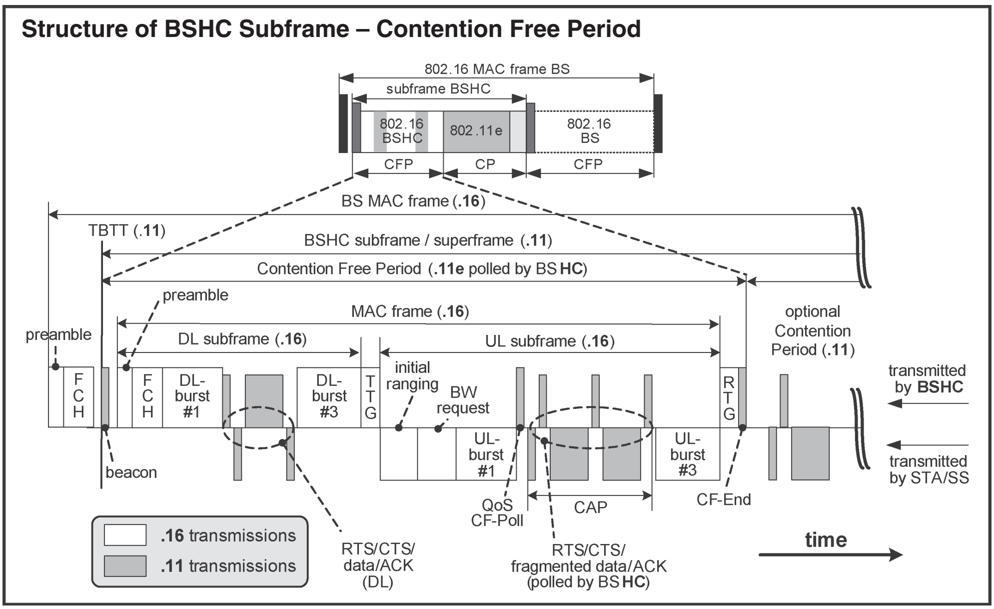


Figure 8.21 Structure of the BSHC subframe nested into the 802.16 MAC frame controlled by the BS. The BSHC subframe is an 802.11 superframe comprising a contention-free period and a contention period (shown in Figure 8.22). The CFP is regarded by 802.16 as a MAC frame, since BSHC starts the subframe by transmitting the standard 802.16 broadcast information (preamble, FCH). 802.11e TXOPs are scheduled in the DL and UL bursts of the 802.16 subframe. Reproduced by permission of © 2006 IEEE[2].

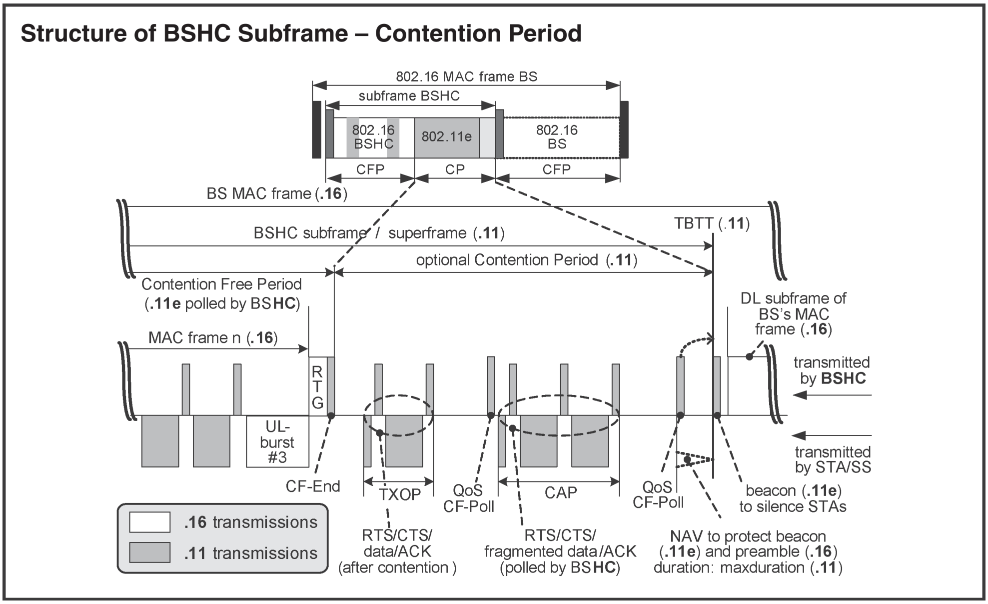


Figure 8.22 Structure of the BSHC subframe. Its duration is announced by transmitting an 802.11 beacon specifying a contention-free period and contention period. The next 802.16 MAC frame issued by the BS is protected against potential interference by 802.11 legacy stations operating in CP through a QoS CF-Poll frame transmitted by the BSHC. Reproduced by permission of © 2006 IEEE[2].

The 802.11(e) part of the interworking concept is based on the HCCA. The BSHC may schedule TXOPs in polling associated QSTAs for UL data transmission or may immediately initiate its own DL transmissions. The BSHC may even schedule 802.11e transmissions during the 802.16 MAC frame in both DL and UL bursts. This is shown in Figure 8.21 for DL burst #2 that is replaced here by an 802.11e DL transmission by the BSHC protected through Request-To-Send (RTS)/Clear-To-Send (CTS) message exchanges preceding. In order to define the DL burst adequately when specifying the 802.16 subframe, it is necessary in the preceding FCH to determine the length of the RTS/CTS/data/ACK sequence in the BSHC. Any 802.16 SSs will fail to decode the 802.11 data received and will discard it. By addressing an 802.11 DL burst in the FCH to a dummy station, no 802.16 SS will try to decode it. Duration of a polled 802.11e TXOP for UL data transmission during the 802.16 MAC frame is controlled by the BSHC. Such TXOP are referred to as Contention Access Periods (CAPs) corresponding to the 802.11e access within the CFP with a duration limited by the 802.16 burst duration, which is under control of the BSHC. The CAP is polled by the BSHC through QoS Contention-Free Poll (*QoS CF-Poll*) and its duration may differ from an EDCA TXOP: The *QoS CF-Poll* allows setting an individual maximum transmission size.

The polling of 802.11e QSTAs using *QoS CF-Poll* is shown in Figure 8.21: The UL burst #2 is replaced with an 802.11 frame transmission sequence of QoS-CF-Poll/RTS/CTS/data/ACK. Such a data transmission on the UL is regarded by the 802.16 SSs as a UL burst. Within the CAP the polled 802.11 STAs decide themselves what data to transmit. A Receive/transmit Transition Gap (RTG) is required for switching the BSHC between reception and transmission modes. This also refers to the reception of 802.16 UL-bursts and the transmission of *QoS CF-Poll* of 802.11e, as also depicted in Figure 8.21. The resources for 802.16 UL/DL bursts may be (partly) used in the opposite transmit direction for 802.11 data transmission (for instance a polled UL transmission in the DL burst) in order to meet restrictive QoS requirements of applications supported by 802.11. The RTGs and Transmit/receive Transition Gaps (TTGs) required for switching between reception and transmission, when changing operation between 802.11 and 802.16 are not shown in Figure 8.21.

Limited contention-based access
From the perspective of the 802.16 SSs, an 802.16 MAC frame is transmitted periodically with a certain period of time between two consecutive MAC frames that is not available for the 802.16 MAC protocol. The CP depicted in Figure 8.22 continues the superframe started in Figure 8.21 with the CFP. After the *CF-End* frame, the CP may start to allow access by 802.11 legacy stations. The main problem with permitting a CP in the superframe is that the next 802.16 MAC frame started by the BS must not be delayed. The BSHC therefore must organize that no 802.11 transmission is hanging over into the next 802.16 MAC frame. Such hangovers cannot be avoided with 802.11e EDCA-TXOPs or frame transmissions from legacy 802.11 STAs.

To avoid hangover, the BSHC schedules in advance the time instance where the next 802.16 MAC frame will be transmitted. In order to protect the timely transmission of the next 802.16 preamble/FCH broadcast by the BS, the BSHC must allocate the channel when a CP transmission has ended and the next 802.16 allocation is closer than the maximum possible duration of an 802.11 legacy station transmission. 802.11 and 802.11e MAC protocols both provide means for realizing such a protection: A *QoS CF-Poll* or a CTS to itself may serve for this, or a new beacon may be transmitted by the BSHC in order to announce a CFP for silencing all 802.11/802.11e stations for a sufficiently long time duration.

With the 802.11a PHY, a station may continuously transmit up to 2 ms (longest possible transmission duration with maximum packet size and most robust PHY mode). The *TXOPlimit* in 802.11e allows a better control of the duration of an EDCA-TXOP. The use of *QoS CF-Poll* is depicted at the end of the CP in Figure 8.22. *QoS CF-Poll* may be used by the BSHC to allocate a high priority TXOP during the CP. The BSHC may initiate in this way in the CP a frame exchange directly after PIFS with or without using the RTS/CTS cycle. The STAs may gain control of the channel after contention corresponding to the EDCA and are allowed to transmit with a maximum duration of *TXOPlimit*. Alternatively, *QoS CF-Poll* or CTS transmitted to the BSHC itself can be used to transmit 802.16 MAC frames within the CP in a protected way.

8.3.3.3 BSHC and Legacy 802.11 Stations

The 802.11e MAC protocol is better suited to implement interworking with 802.16 MANs than the 802.11 protocol, since the 802.11 CP must be delimited at well-defined time instants, where the 802.16 MAC frame must transmit its beacon. Therefore, the BSHC concept may allow operation of a CP as long as the 802.11 stations use EDCA but not the DCF. The EDCA limits the duration of an allocation by specifying the length of a TXOP and the BSHC can thus control the use of the WM and guarantee interference-free transmission of the 802.16 beacon. Legacy 802.11 stations may violate the *TXOPlimit* leading to fatal interference with the next 802.16 MAC frame broadcast phase. 802.11 STAs respect the CFP that is part of legacy 802.11 protocol through the PCF. Nevertheless they do not respect the TBTT as legacy devices cannot be polled.

Mangold (2003) suggests not allowing legacy 802.11 stations to associate with the BSHC and to exploit the Extended Interframe Space (EIFS), originally designed to be used under hidden station interference. This would allow the BSHC to force legacy 802.11 STAs to defer from medium access for time duration EIFS: According to 802.11, if an STA detects a preamble but is not able to receive the complete frame it defers from medium access for EIFS duration. If the Frame Check Sequence (FCS) of an 802.11 transmission fails, reception is not successful. The BHSC may take advantage of this in transmitting correct preambles and headers but wrong FCSs or may use a wrong PHY mode for the remaining frame. Legacy STAs detecting such frames will pause for EIFS instead of DIFS for channel access. A frequent transmission of preambles and Physical Layer Convergence Protocol (PLCP) headers during EIFS duration would result in manipulating the backoff algorithm for permanently setting the timers in legacy STAs to EIFS.

8.3.4 Performance Evaluation Results

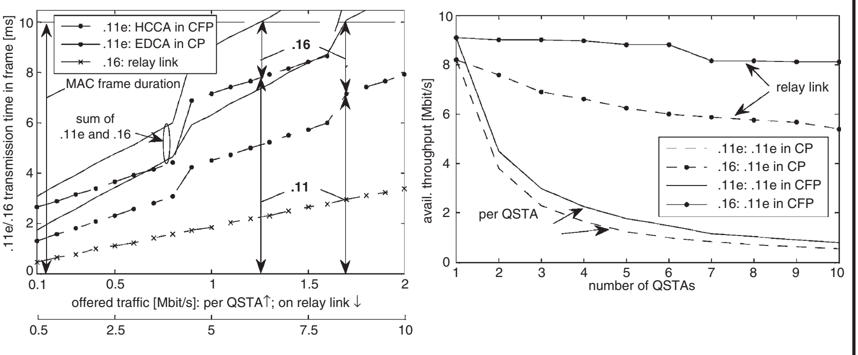
This section presents calculated results on the performance of interworking of 802.11e and 802.16 systems on the example of the scenario shown in Figure 8.19. A BS connected to the fixed network serves a BSHC by means of an 802.16 MAC frame and the BSHC (as an FRS) serves 802.11e QSTAs on the same frequency channel. For the sake of simplicity, it is assumed that the complete channel capacity is assigned by the BS to the subframe controlled by the BSHC. Thus, the BS serves only one relay (the BSHC) and no SSs on its own. Such a scenario could for instance be a village which is connected for enabling broadband access via one link to the outside world. The results presented take the MAC and PHY overhead of the respective protocols into account. The system parameters used in the simulation study are summarized in Figure 8.23 and the results are shown in Figure 8.24.

Two alternatives are being studied: The TXOPs of 802.11e are either allocated using HCCA in the CFP, embedded to the UL/DL bursts of the 802.16 MAC subframe or they are transmitted using EDCA in the CP. The optimal partitioning of the MAC frame between 802.11e and 802.16 based transmissions is used and the results shown in Figure 8.24(a). The utilization of a 10 ms MAC frame depending on the offered traffic per QSTA is shown. Five QSTAs with

common system parameters			
MAC-frame duration: 10 ms	packet size: 512 Byte	MCS: QAM 16 1/2	DL/UL ratio: 50/50
channel bandwidth: 20 MHz	all QSTAs have same traffic loads		
IEEE 802.16		**IEEE 802.11 e**	
FCH with one IE for DL	UL MAP with one IE for UL burst	802.11a PHY parameters	EDCA in CP: AC_VI
CRC-32 enabled	fragmentation disabled	no RTS/CTS	CWmax = 15
114.58 µs for BW req slot	68.75 µs for ranging slot	with WEP	CWmin = 7
cyclic prefix=1/32	TTG = RTG = 22.9 µs	retry count: 7 with Address 4	AIFSN = 2

Figure 8.23 System parameters used for evaluating the interworking scenario shown in Figure 8.19 with one IEEE 802.16 relay link and IEEE 802.11e based communication between BSHC and a variable number of QSTAs. Reproduced by permission of © 2006 IEEE[2].

Traffic Performance of Interworking of 802.16 and 802.11



(a) Frame usage of 802.11e and 802.16 in scenario of five 802.11e QSTAs served by one 802.16 relay link

(b) Throughput of the 802.16 relay link and for one 802.11e QSTA when each QSTA gets the same throughput

Figure 8.24 Performance evaluation results of interworking systems 802.11e/802.16 in scenario of Figure 8.19. Optimal partitioning of MAC frame between 802.11e/802.16 according to (a) is used in (b) to determine the maximum available throughput per QSTA and on the relay link. Reproduced by permission of © 2006 IEEE[2].

the same offered traffic loads are assumed. The 802.16 link carries the traffic of all QSTAs together and therefore has five times the offered throughput of one QSTA. The total time required for transmitting the offered traffic over 802.11e is compared to the transmission time needed in 802.16. Owing to the distributed medium access, using the same PHY mode in both networks, twice the time for transmission of the same amount of user data is required in 802.11. The time needed with EDCA in the CP is calculated using a Markov model of the backoff procedure (Mangold, 2003). The steps in the 802.11e graphs results from fragmentation necessary with increased traffic load under packet size limitation: An additional polling or RTS/CTS is required in UL and DL direction of 802.11e at each step. The sum total of 802.11e and 802.16 frame transmission times leads to a maximum value for the system throughput and implies an optimal partitioning of the MAC frame between 802.11e and 802.16 as indicated in Figure 8.24(a).

The maximum available throughput under optimal partitioning of the MAC frame that depends on the number of QSTAs served by the BSHC, is shown in Figure 8.24(b). The example scenario shown in Figure 8.19 but with a variable number of QSTAs and no 802.16 SSs served by the BSHC is studied with an optimal 802.16 MAC frame partitioning assumed. The throughput of the 802.16 link and of each QSTA is shown. With increased number of QSTAs the overall system capacity decreases as visible from the 802.16 link throughput. Contention during CP essentially increases with increasing number of QSTAs and, accordingly, the share assigned to 802.11e of the 802.16 MAC frame must be increased. Serving QSTAs in the CFP increases MAC overhead with an increasing number of QSTAs too, but gives an overall better performance.

8.3.5 Summary

The concept introduced for interworking of 802.11 and 802.16 systems under central BS/BSHC control permits operation of both networks on the same frequency channel. It solves the coexistence problem of 802.16 and 802.11 by giving local (one-hop) control to an 802.11e HC that cares for 802.16 SSs to coexist to 802.11 STAs, under control of the remote 802.16 BS. It has been shown that interworking influences the medium access of all the wireless networks involved in spectrum sharing. Restrictions to and requirements on the coexisting protocols need to be taken into account to enable QoS support under coexistence conditions. The adherence of a common frame structure can be regarded as extreme cooperation.

8.4 Conclusion

This chapter introduced the mesh standardization activities ongoing in IEEE 802. The approaches taken so far towards introducing meshing to 802.11, 802.15 and 802.16 have been discussed.

In addition, a number of candidate solutions to allow for multi-hop operation in 802.11 and 802.16 wireless networks have been discussed in this chapter that are so far not part of these standards. Since the concepts discussed are standard compliant, they might be applied and realized without involving standardization activities. Especially the concepts applying central control only require minor MAC software modification to the 802.16 BS and/or 802.11 AP so that company-specific, proprietary implementations are expected to be a promising alternative to waiting for a solution to be approved by a long-lasting standardization process. Current standardization in 802 Task Groups tends to not specify options for MAC protocols, which support multi-hop operation. Instead, TG 802.11s currently envisages, for instance, the introduction of routing protocols on top of the MAC layer (layer 2.5) to care for multi-hop operation, see Section 4.3.1.

The concepts presented demonstrate the feasibility of multi-hop operation in homogeneous as well as heterogeneous 802.11 and 802.16 wireless networks. Additionally, solutions for realizing the interworking of 802.11 and 802.16 are proposed. The frame-based MAC protocol of 802.16 and the HCCA of 802.11e apply central control, and therefore facilitate the introduction of efficient multi-hop operation and interworking.

Notes

1. Reproduced by permission of © 2006 IEEE. Source: G. R. Hiertz, S. Max, E. Weiß, L. Berlemann, D. Denteneer, S. Mangold, “Mesh Technology enabling Ubiquitous Wireless Networks - Invited paper,” in Proc. of the 2nd Annual International Internet Conference, Boston, USA, 2006. (Section 8.1 (in parts)).
2. Reproduced by permission of © 2006 IEEE. Source: L. Berlemann, C. Hoymann, G.R. Hiertz, and Mangold, “Coexistence and Interworking of IEEE 802.16 and IEEE 802.11 (e),” in Proc. of the IEEE 63rd Vehicular Technology Conference, VTC2006-Spring, Melbourne, Australia, 7–10 May 2006. (Figs. 8.21, 8.22, 8.23 and 8.24, Section 8.3.3. (in parts)).

9

Coexistence in IEEE 802 Networks

Lars Berlemann, Stefan Mangold and Bernhard H. Walke

Coexistence and spectrum sharing of licensed and unlicensed bands is becoming an increasingly significant research problem. With the proliferation of IEEE 802.11 Wireless Local Area Networks (WLANs), and other future radio systems using unlicensed bands, spectral coexistence of dissimilar radio systems has to be addressed in the near future. Wireless Metropolitan Area Networks (WMANs) such as 802.16 may also require coexistence capabilities when operating in these unlicensed bands or when simultaneously operating the same frequency channel. Depending on spectrum regulation of 802.16, WMANs belonging to different operators might have to share spectrum.

Coexistence can be achieved with the help of spectrum etiquette, or alternatively by implementing new communication protocols and defining common spectrum coordination channels (Raychandhuri 2003). In this chapter, we investigate how independent wireless networks may share spectrum without direct coordination and information exchange. The uncoordinated access to radio resources that are shared with other radio networks leads to increasingly problematic situations, especially in the context of Quality-of-Service (QoS) support. Such coexistence scenarios are not addressed in standards such as IEEE 802.11 with its extension 802.11e for QoS support and IEEE 802.16. For future radio networks, coexistence among dissimilar radio networks sharing common radio resources are under discussion in standardization groups such as the Wi-Fi Alliance and IEEE 802.19. Wireless networks operating in unlicensed spectrum, i.e., open spectrum, are typically not designed for exchanging information among dissimilar radio networks.

The realization of homogeneous coexistence, i.e., spectrum sharing of wireless networks of the same technology/standard, is introduced in Section 9.1 using the example of 802.11e WLANs. The end of this section discusses the applicability of the introduced approach on spectrum sharing 802.16 wireless networks. Section 9.2 addresses the heterogeneous coexistence in proposing methods that allow the unlicensed operation of 802.16 in a shared spectrum with 802.11. This chapter is summarized in Section 9.3 with a conclusion.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

9.1 Homogeneous Coexistence – Spectrum Sharing 802.11e Networks[1]

This section gives a tutorial description of using game theory for modeling the competition of WLANs/WMANs sharing unlicensed frequency bands (Berlemann *et al.*, 2005b). Therefore, we approach the coexistence problem with a stage-based, non-cooperative game (Fudenberg and Tirole, 1998; Osborne and Rubinstein, 1994) to analyze competition scenarios of two wireless networks. Fundamentals of game theory and its application in resource management for modeling the interaction between service provider and customers are introduced in Das *et al.* (2004). Contrary to Altman *et al.* (2002) and Srinivasan *et al.* (2003), where cooperative relaying in ad-hoc networks is considered, we focus on the decentralized coordination for distributed QoS support in unlicensed communication systems. The distributed coordination of reservations for spectrum allocation, as for instance in Wireless Personal Area Networks (WPANs) as part of the Distributed Reservation Protocol (DRP) introduced in Chapter 6 is enabled as well.

9.1.1 Coexistence Scenario

A coexistence scenario of 802.11e WLANs according to the scenario illustrated in Figure 9.1 is considered in the following sections. Two 802.11e Hybrid Coordintors (HCs), each represented by a player, compete for the exclusive access to a single, shared frequency channel. This is the classical coexistence problem known from 802.11/802.11e: The exclusive right of the central coordinator to access the wireless medium is lost and a QoS guarantee is impossible. The results presented in Figure 9.2 indicate the consequences of the coexistence on the observed QoS. The QoS of an isolated WLAN of one player (a) is compared with the QoS in a spectrum-sharing

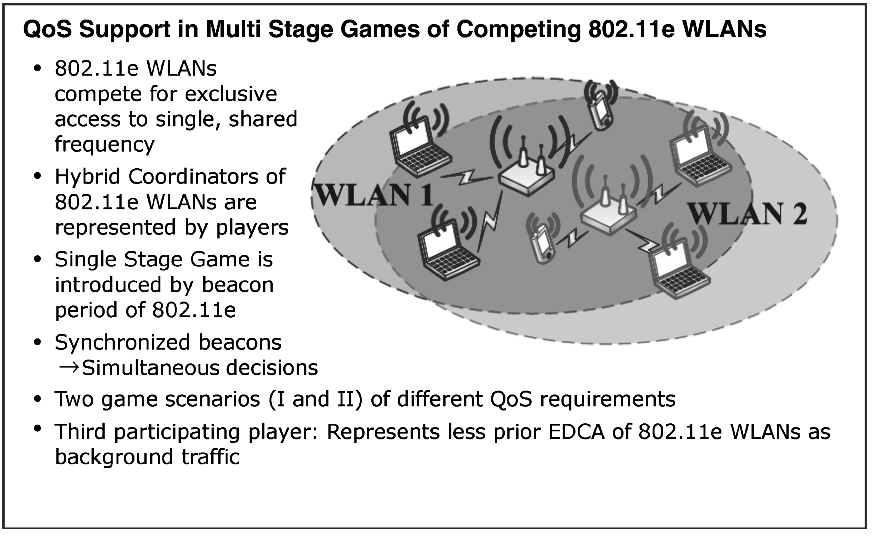


Figure 9.1 QoS support in spectrum sharing games of competing 802.11e WLANs.

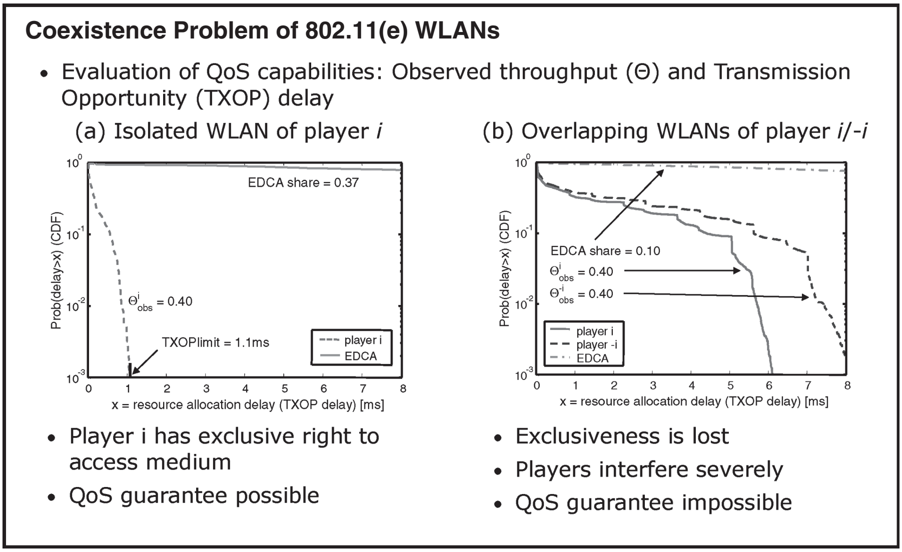


Figure 9.2 Coexistence problem of 802.11(e) WLANs.

scenario of two overlapping WLANs (b). In scenario (a) a QoS guarantee is possible due to the means of 802.11e that are described in detail in Section 5.5. As the exclusiveness of spectrum access is lost in scenario (b), QoS support is nearly impossible indicated by the high delays of both WLANs.

9.1.2 Overview

To facilitate the understanding of the terms used in this chapter we illustrate the concepts with the help of the Universal Modeling Language (UML), see Figure 9.3. Each radio system is represented by a player, which competes with another player for the control over a shared resource to support QoS. Such a player stands for all MAC entities of one coexisting wireless network. In our example of coexisting IEEE 802.11e WLANs, a player includes at least one 802.11e HC. Although radio technologies for unlicensed bands share a common resource, they are typically not designed to arrange spectrum usage with different systems. We take this into account as we assume that players cannot establish communication with each other directly. The QoS requirements imposed by services and applications define a multidimensional utility function. The utility is an abstract representation of the observed throughput and delay. It is an important part of the stage-based game model and is discussed in detail in Section 9.1.3.

Players interact repeatedly by selecting their own behavior (a selection of MAC parameters). For the sake of simplicity the behaviors of a player are limited here to cooperation and defection. After each stage of the game the players estimate their opponent's behavior. The estimated behavior of the opponent has to be classified in taking its intention into account, as discussed in detail in Section 9.1.4. This classification is necessary, as there is no communication between the dissimilar radio systems, i.e., the players, which hinders direct negotiations. Nevertheless, players are aware of their influence on the opponent's utility, which enables interaction on the

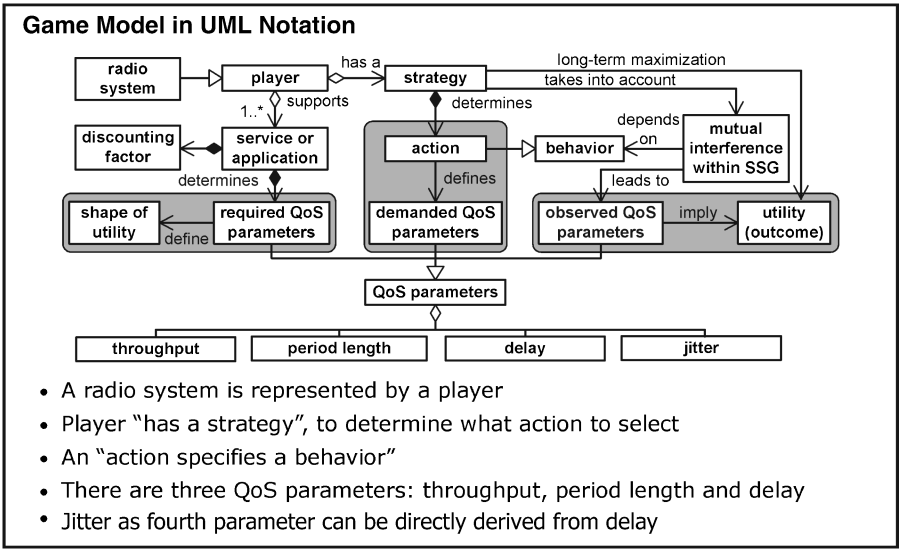


Figure 9.3 The game model in UML notation. Reproduced by permission of © 2004 IEEE[2].

basis of punishment and cooperation, i.e., a handpicked allocation of the radio resource aiming at a specific intention.

The context of game theory for judging game outcomes is outlined on two levels. The existence of equilibria, i.e., steady outcomes of interaction, in Single Stage Games (SSGs) depends on the players' behaviors as outlined in Section 9.1.5. Strategies in a Multi Stage Game (MSG), which is formed by repeated SSGs, are discussed on a second, higher level in Section 9.1.6. The players decide about their strategy in discounting expected utilities to calculate future outcomes of the MSG. The selected strategy of a player is decisive for the course of interaction. Thus, the capability to guarantee QoS depending on the chosen strategy is evaluated in Section 9.1.7.

9.1.3 Single Stage Game

An SSG is formed by an IEEE 802.11e superframe and has a fixed time duration. Such a superframe is the time between two consecutive beacon frames that are used for broadcasts. These beacons can be used to determine such an interval, which in the following is arbitrarily set to a duration of 100 ms. Spectrum allocating players always allocate the shared channel exclusively with the help of the HC, whose allocations are referred to as TXOPs. Note that each player represents all MAC entities of a single coexisting WLAN, and each network is assumed to include at least one HC. During the SSG duration, the demanded resource allocation times of all players determine which individual player can allocate a TXOP at what time.

An SSG consists of three phases, as illustrated in Figure 9.4. Together they form one single stage: (I) The players decide about their action, which means they demand resource allocation times and durations. This is an instant of time at the beginning of an SSG, and hence does not consume any time. (II) The competitive medium access of the allocation process during the SSG occurs in the second phase. It may result in resource allocation delays and collisions of allocation attempts: Resources may already be used by opponent players when an allocation attempt is

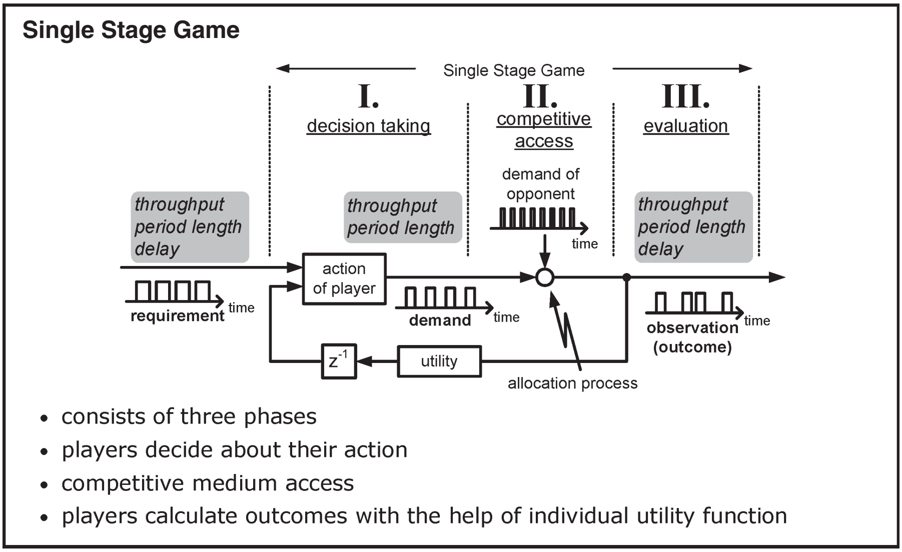


Figure 9.4 Repeated interaction on the basis of SSGs. An SSG consists of three phases.

demanded. Hence, the observed allocation points may differ from the demanded allocations, which is the reason for the difference of demand and observation. The second phase is the one that consumes the time of the SSG. (III) After the allocation process, players calculate the outcomes with the help of individually defined utility functions in the third phase of an SSG, again in an instant of time. The outcome can be regarded as the difference between what was wanted and what is *de facto* achieved. Spectrum allocations can be observed by all players, but the observed utility of a player, cannot be observed by opponent players. The utility (i.e., the outcome) is only individually known by each player. In the following stage, demands and actual observations are taken into account when the players decide which action to select next. Observed outcomes of an SSG contribute to the game history over multiple stages, as it will be explained in more detail in Section 9.1.6.

9.1.3.1 Quality-of-Service as Utility

We define three values between 0 and 1 that together represent the QoS targets for the players, as shown in Figures 9.3 and 9.4: (i) The throughput $\in [0, 1]$, (ii) the period length $\in [0, 0.1]$ and (iii) the delay $\in [0, 0.1]$. Period lengths and delays are limited to grant typical values for a stage length of 100 ms. For a detailed definition of these QoS targets in the context of an SSG see Mangold *et al.* (2003). The demanded throughputs and allocation intervals are determined by the player's demands. They are selected at the beginning of each stage. This selection is the actual decision making, or, as in the rest of this chapter called, the *action* of a player. The period length, i.e., the interval between two successive TXOPs, aims at a signaling of the players' tolerable delay: The period length demanded by one player can be observed by opponent players, which is the important characteristic in our SSG, to enable players to estimate their opponent's intentions, and respond to their behaviors. With the help of the period length, a player can signal its own intention, such as cooperation, and hence this parameter allows the establishment of cooperation, as it will be described in the following section.

Observed delay and throughput can be significantly different to the demanded parameters, because they result from the dynamics of the interactions during an SSG. This is reflected in the utility function. The multidimensional utility function represents the value of the observed QoS for a player. It is the unique objective of a player to optimize this value with respect to the utility function. The utility represents the supported QoS of a player and depends consequently on the above-introduced three QoS targets. For example, an observed utility may be zero although the radio resource is not frequently used in times when a player is unable to allocate resources as required.

There are many different thinkable approaches to reflect QoS characteristics in an utility function. We have chosen an approach based on rational functions to simplify the analytical analysis. Player i's utility U^i is defined as

$$U^i = U^i_{throughput} \cdot U^i_{period\ length} \cdot U^i_{delay}, \ U^i \in [0 \ldots 1].$$

The utility function U^i consists of three terms that are related to the throughput, the period length and the delay.

For a better understanding of the utility function, Figure 9.5 depicts in (a) the observed utility of a player, which exclusively utilizes the radio resource depending on its demanded QoS. As introduced above, the player can demand a specific throughput, i.e., share of capacity, and period length that is together referred to as an action. The maximum of the utility function is per definition given by the required QoS: (*throughput, period length, maximum delay*) = (0.4, 0.045, 0.02). Missing this requirement implies either unfulfilled restrictions to the medium access or exceeding the in fact needed resource utilization. Due to the exclusive resource allocation, the player observes this requirement when it is demanded and its allocations are not delayed. To force players not to allocate too much of the medium to the benefit of the opponent, the utility

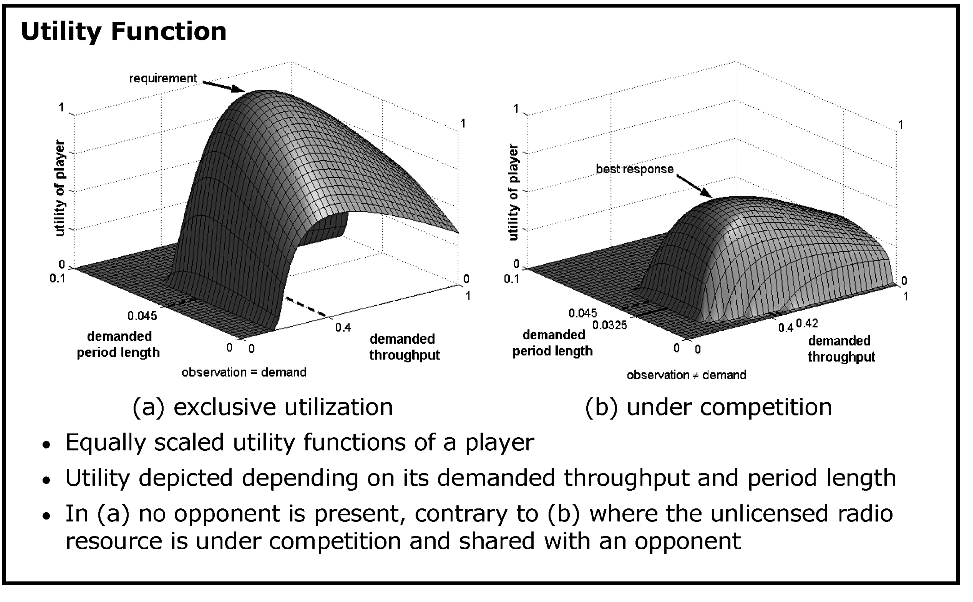


Figure 9.5 The utility function of a player.

is reduced for high demanded throughputs. Further, the utility function implies some kind of elasticity, whether a player may strictly need its requirement or is satisfied with less adequate observations. The dimensioning and definition of the utility function is described in Mangold *et al.* (2003).

9.1.3.2 Utility under Competition

In a competition scenario of coexisting WLANs the radio resource has to be shared. Under such competition, i.e., in the presence of another player, the players' allocations interfere. Thus, the players observe less QoS than demanded. This leads to a decreased utility as depicted in Figure 9.5(b): The opponent has a fixed QoS requirement of *(throughput, period length, maximum delay)* $= (0.4, 0.02, 0.02)$ while the demanded throughput and period length of the player from above is varied. The observed delay, because of the opponent's allocations of the shared radio resource, inevitably is considered in the utility function as a factor, but is not part of an action. Further, a player has to demand more restrictive QoS than needed to satisfy its QoS requirements. In this context, Figure 9.5(b) illustrates the "best response" action of a player: The optimal pair of throughput and period length, here 0.42 and 0.0325. In a "best response" action, the resources that are demanded by the opponent are considered. Players estimate their opponent's demands (Mangold, 2003), and use the history of interactions to predict the opponent's expected action for the next stage. This calculation results in an estimate for the players' expected potential utilities and thus determines the optimal "best response".

9.1.4 Behaviors in Single Stage Games

9.1.4.1 Cooperation through Predictable Behavior

In the absence of a centrally coordinating entity, a player can establish cooperation by behaving predictably. Predictable resource allocations of a player during an SSG will enable other players to understand and respond to its actions.

For this reason, we refer to predictable behavior as a contribution to cooperation in the absence of a centrally coordinating entity (Mangold, 2003). The fixed periodicity of resource allocations by a player can be observed and predicted by other players. These other players may adapt their own resource allocations with the objective to mitigate mutual interference. Such a behavior is a cooperation as response to an opponent's cooperative behavior.

Actual resource allocations correspond to the individual QoS requirements of players. A decrease of the period lengths may, however, be considered as another important contribution to establish cooperation (Mangold, 2003): While the number of equally distributed resource allocations per stage is increased, individual TXOPs (i.e., individual resource allocations) are relatively short, which can reduce the observed delays of resource allocations of opponent players.

9.1.4.2 Classification of the Opponent's Behavior

Dynamic (trigger) strategies are used by players to change behaviors from stage to stage. They consider the strategies of opponent players. An opponent's strategy needs to be understood by a player that operates with such a dynamic strategy. Due to the lack of direct information exchange, it is impossible for a player to identify an opponent's strategy in the game model, unless behaviors are altered by the opponent in an intelligent way. Players therefore have to classify the opponent's behavior by differentiating between two possible intentions: cooperation and defection. Players are aware of the influence of their demanded allocations on the opponent and vice versa. Consequently, they are able to identify a specific intended behavior. Cooperation (C) is defined as intending a behavior that aims for a fair share of resources. The second intention is defection (D), which can be motivated differently for two different reasons. The action that corresponds to defection is the known "best response" and is identical for the different motivations. On the one hand, defection can be an intended act of ending an established game-wide cooperation for the purpose of increasing the own outcome. On the other hand, defection can be the reaction to an opponent's deviation from game-wide cooperation with the aim of punishing the opponent in response. Such a punishment can be implemented in different ways. All actions of a player that reduce the utility of the opponent can be considered as punishment, because the player is aware of its influence on the opponent. To demand more restrictive QoS targets than needed, sending a busy tone or transmitting empty data packets are examples of punishing the opponent.

Defective behavior implies neglecting future payoffs motivated for instance in the players' support of applications on a "best-effort" basis such as email. Applications with restrictive QoS requirements, as for instance video-conferencing services, give one reason for cooperative behavior.

9.1.5 Equilibrium Analysis of Single Stage Game

The outcome of an SSG, namely the utility as defined above, depends on the observed QoS parameter. They can be determined by the players through an analytic model (Mangold, 2003) based on the expected opponent's action. It is interesting to analyze whether these outcomes from repeated, simultaneous interactions are steady and/or utility maximizing. Therefore, in an SSG of rational acting players the existence of a "best response" action on the expected opponent's action has to be considered. In addition, the uniqueness and stability of such an action is of interest for the players' decision-making process, which action to take. A commonly used solution concept for the question which action should be selected in an SSG of rational players is the Nash Equilibrium (NE) solution concept (Fudenberg and Tirole, 1998; Osborne and Rubinstein, 1994). This solution concept in the context of our game model is outlined in the following.

In general, "an NE is a profile of strategies such that each player's strategy is a best response to the other players' strategy" (Fudenberg and Tirole, 1998). Here, in the context of the SSG the players' strategy consists of a single specific action. This action leads to an observed utility, as the outcome of the SSG. Neither player has the incentive to leave the NE, as a deviating action would imply a reduction of the own observed utility. Thus, NEs are consistent predictions of how the game will be played. In the sense, if all players predict that a particular NE will occur, then no player has the incentive to play differently. Thus, an NE, and only an NE, can have the property that the players can predict it; predict that their opponents predict it, and so on. The NE is a value for the game's (in-)stability. Hence, it can be seen as a lower limit for the QoS that can be guaranteed in a competition scenario of rational players.

A microeconomic concept to judge outcomes of a game is the *Pareto efficiency* (Fudenberg and Tirole, 1998): An SSG outcome is called Pareto efficient if neither player can gain a higher utility without decreasing the utility of at least one other player. A non-Pareto efficient situation is not a preferable outcome of a stage because a rational player could improve its utility without changing the game and its outcome for the other player.

The *bargaining domain* of Figure 9.6 contains a subset of all possible SSG outcomes, by means of players' utilities, corresponding to an action pair of player i and $-i$. Here, the actions belong

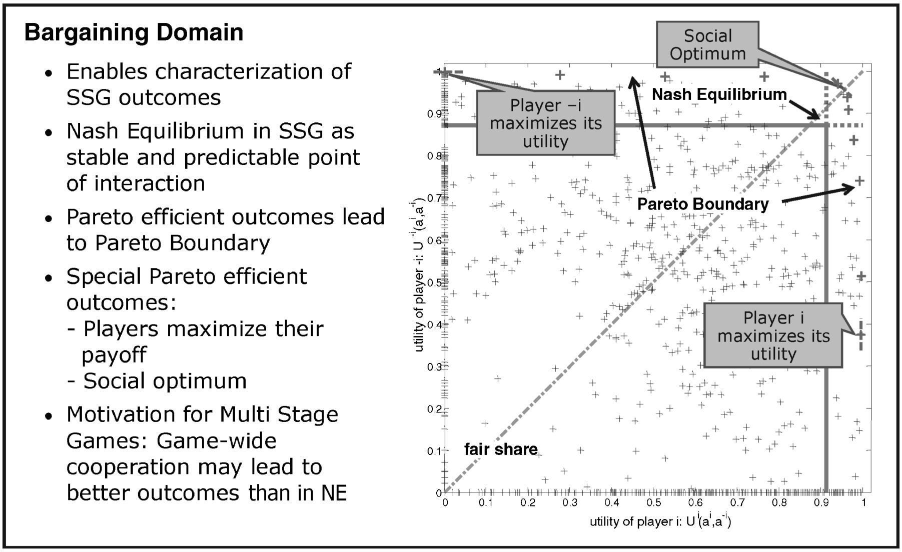


Figure 9.6 The bargaining domain of a game scenario. Each cross marks the players' utilities of an SSG and the players' actions are varied.

to a discrete set. The corresponding SSG outcomes, i.e., utility pairs, are calculated with the help of an analytic model (Mangold, 2003). The bargaining domain supports the judgment of potential SSG outcomes. Depending on the players' QoS requirements, none, a unique NE or several NEs can be found. Here, we have a game scenario of unique NE with the actions player i demanding a throughput of 0.5 and a period length of 0.032 while the opponent $-i$ demands 0.54 and 0.03 respectively. The corresponding utilities are 0.913 and 0.872, as depicted in Figure 9.6. This NE, which is not Pareto efficient, can be considered as a minimum for the reachable utilities of both players. In this way a lower but nevertheless predictable limit for the support of QoS is given. A further analysis indicates that this unique NE is reached as a steady outcome of an MSG if both players follow the "best response" behavior.

The single NE enables a definition of a *Pareto boundary*, which is introduced by the Pareto efficient SSG outcomes at the right and upper corner of the bargaining domain. The game scenario that is illustrated in Figure 9.6 has three prominent Pareto efficient outcomes: The utility pairs with a maximum utility for each player located for player $-i$ in the upper left area of the bargaining domain and for player i in the lower right area. The Pareto efficient social optimum is located in the upper right corner where the sum of both players' utilities has its maximum value. There, both players gain, contrary to the other two Pareto efficient outcomes, a higher utility than in the NE. As a result, both players can improve their utilities through interaction, compared to utilities in the NE. This interaction is referred to as cooperation to the benefit of all players and is the motivation for our focus on MSGs.

9.1.6 Multi Stage Game

The above-introduced SSG and behavior of a player allows us to introduce another degree of interaction: The dynamic interaction in repeated SSGs, coordinated through strategies. Technical restrictions, as for instance the battery power, limit the duration of a MSG. We nevertheless assume legitimately that MSGs have no limited time horizon: An MSG can be regarded as infinite as "a model with infinite time horizon is appropriate if after each period the players believe that the game will continue for an additional period" (Osborne and Rubinstein, 1994).

When selecting how to access the medium, players take into account the expected results (the expected utilities) of the instantaneous stage, but should also take into account the effects of their decisions on the utilities of future stages. This is usually expressed through weighting the stages. Players give present utilities a higher weight than potential utilities in the future, because of the uncertainty of those future results. A known approach to model this weighting of the future is to discount the utilities for each future stage of a game. Therefore, a discounting factor δ, $0 < \delta < 1$, is defined in Berlemann *et al.* (2004). δ reflects in the present stage the worth of future utilities of following stages. Player i's utilities U^i of an infinite MSG is defined as the sum over its utilities U_n^i of stage n discounted with player i's discount factor δ^i:

$$U^i = \sum_{n=0}^{\infty} \left(\delta^i\right)^n U_n^i = \frac{1}{1-\delta^i} U_n^i,\ if\ U_n^i = const.$$

A δ^i near one implies that future utilities are considered similarly to the utility of the current stage. Thus, the player tends to cooperate to enable a high long-term utility. Contrary, a player with a δ^i near zero only has its focus on the present utility and completely neglects potential future utilities resulting into non-cooperating defection (Berlemann *et al.*, 2004).

9.1.7 Strategies in Multi Stage Games

In MSGs, strategies determine the behaviors for each individual SSG. Players try to optimize their utility in applying adequate strategies. Using a state model as defined in Figure 9.7(a), a strategy describes the alternatives of a player. Each state represents a certain behavior. A strategy also models under which circumstances a transition from one state to another happens; hence, it models the decision making. We only allow what is in game theory referred to as a "pure" strategy (Fudenberg and Tirole, 1998): Players have to choose one specific behavior for each stage, and cannot perform soft decisions by assigning probabilities to different state transitions. Strategies can be interpreted as social norms in repeated interaction: "Social norms are isolated types of strategies that support in any game mutually desirable and thus stable utilities" (Osborne and Rubinstein, 1994). In other words, strategies enable QoS support independently from the opponent's strategy and QoS requirements. We distinguish in the following between static and dynamic (trigger) strategies.

9.1.7.1 Static Strategies

Static strategies are the continuous application of one behavior without regarding the opponent's strategy. In static strategies, there is no state transition, and the state model contains one single state. In our approach, the set of available static strategies is reduced to two: The cooperation strategy (COOP) is characterized through cooperating every time, independently from the opponent's influence on the player's utility. The COOP strategy is to the benefit of a player if the opponent cooperates as well. Figure 9.7(b) illustrates this simple strategy of following a cooperative behavior ©. Equivalent to the COOP strategy, the defection strategy (DEF) consists of

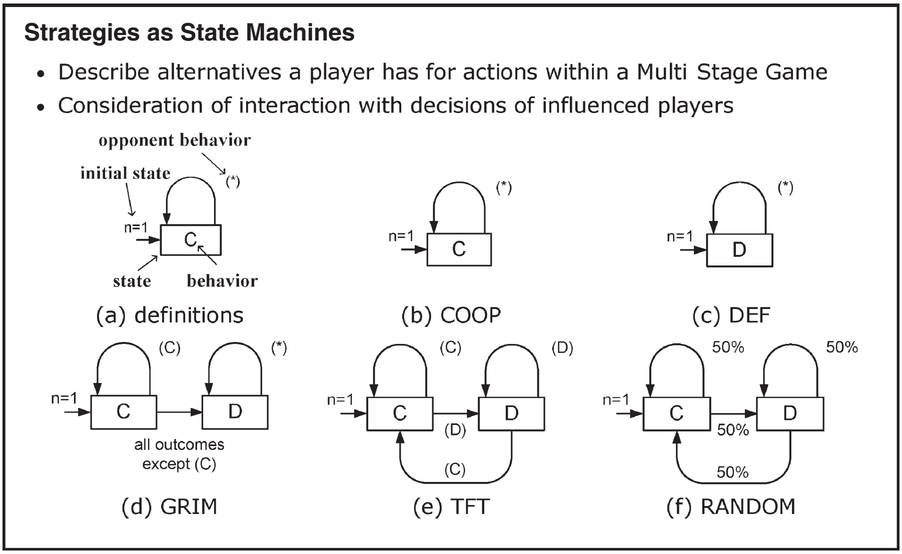


Figure 9.7 State machines for modeling strategies as defined in (a). Static strategies: Strategy of cooperation (b) and defection (c). Trigger strategies: GRIM strategy (d), TitForTat strategy (e) and RANDOM strategy (f). Reproduced by permission of © 2005 IEEE[2].

a permanently chosen behavior of defection (D). Figure 9.7(c) illustrates the DEF strategy as a state machine.

9.1.7.2 Dynamic (Trigger) Strategies Grim and TitForTat

Trigger strategies are well known in game theory (Fudenberg and Tirole, 1998; Osborne and Rubinstein, 1994). A trigger strategy is a dynamic strategy where the transition from one state to another state is event-driven; an observed event triggers a behavior change of a player. Depending on the number of states (the number of behaviors a player may select), a large number of trigger strategies is possible. For the sake of simplicity, the familiar Grim (GRIM) and TitForTat (TFT) trigger strategies are applied in the following. A player with a GRIM strategy punishes the opponent for a single deviation from cooperation with a defection forever. A player applying this strategy may be referred to as an unforgiving player. The initial state of the GRIM strategy, selected at the first stage of the MSG, is however the cooperation. The player cooperates as long as the opponent cooperates, and the transition to defection is triggered by the opponent's defection. See Figure 9.7(d) for an illustration of the state machine of the GRIM strategy. The TFT strategy selects cooperation as long as the opponent is cooperating, similar to the GRIM strategy, also with cooperation in the initial stage. An opponent's defection in stage N triggers a state transition and is punished by defection in the following stage $N+1$, as illustrated in Figure 9.7(e). However, in contrast to the GRIM strategy, TFT changes back as soon as the opponent cooperates again. The TFT strategy is well known in game theory and social science. The advantage of the TFT strategy is on the one hand that it motivates opponent players to cooperate (because of the potential punishment), and on the other hand the robustness when applied in non-cooperative environments, where opponent players often defect.

9.1.7.3 RANDOM Strategy

We also want to analyze how the different strategies perform when applied against purely random behavior. To analyze whether a random play, or alternatively a deterministic, predictable play that usually results in a stable course of the game is to the advantage of a player, we introduce the dynamic strategy RANDOM. This strategy, as shown in Figure 9.7(f), results in uniformly distributed behaviors, 50 % cooperation (C) and 50 % defection (D), regardless what behavior the opponent player may select.

9.1.7.4 QoS Support in Multi Stage Games of Competing WLANs

MSGs can be evaluated based on the observed utilities, to decide which strategy is optimal as it is done in Berlemann *et al.* (2004). This section continues this evaluation in considering the level of QoS support during an MSG. This reverses the abstraction done when introducing the utility. Our game model and the basic IEEE 802.11e access mechanisms to a shared resource are evaluated with the help of our Matlab-based event-driven simulator YouShi (Mangold, 2003).

The QoS capabilities of the strategies introduced in Figure 9.7 are evaluated in the following. Multiple MSGs with varied strategies for both players are evaluated and summarized. Each strategy pair of player i and $-i$ has a corresponding course in the MSG. Such a strategy pair results in specific combinations of behaviors in the SSGs of an MSG. Such a behavior combination is noted as (behavior of player i, behavior of player $-i$), for example (C,C). The QoS outcomes in MSGs of various strategies are summarized in detail in Berlemann *et al.* (2005c). The analyzed observed QoS of a player considers the achievable throughput, which is given as fraction of total capacity, as well as the probability of an observed TXOP allocation delay. The observed QoS of player is evaluated over the outcomes of 400 stages of two-player MSGs. The players have normalized QoS requirements, as introduced and defined in Mangold (2003) of (throughput, period length, maximum delay) = (0.4, 0.05, 0.02) for player i and (throughput, period length, maximum delay) = (0.4, 0.031, 0.02) for player $-i$, referred to as game scenario I. A third participating player, which is not evaluated here, represents the background traffic and the contention-based medium access.

For comparing the success of different strategies of a player, we need a summarizing value for each strategy. As the opponent's strategy is unknown, we focus on the QoS values resulting from a strategy, against the generality of all opponent strategies. The success against a specific single strategy of the opponent is therefore second rated. We define a weighted Strategy Comparison Index (SCI) as a summarizing value for the capability of a strategy to support QoS (Berlemann *et al.*, 2005c): The SCI contains the 98 % percentile of the resource allocation delay (TXOP delay) distribution function together with the fulfillment of the required throughput of a player in applying the considered strategy against the opponent's strategy comparable to the QoS results depicted in Figure 9.2. We choose the strategies of COOP, DEF and RANDOM as representative for all strategies available to the opponent. The SCIs for both players resulting from MSGs are summarized in Figure 9.8. Game scenario I leads therefore to the graph marked with crosses and the other values are formed respectively. The smallest SCI values indicate the strategy that is most adequate against all opponent strategies: It is the best strategy if the opponent's strategy is unknown to the player.

Corresponding to game scenario I, the GRIM strategy is the most adequate one for both players to successfully support QoS. The "best response" behavior of the defection is adequate against a non-cooperating opponent. Nevertheless, MSGs of game-wide cooperation for player i lead to shorter delays and thus the GRIM strategy is the most suitable one. This is contrary to the results

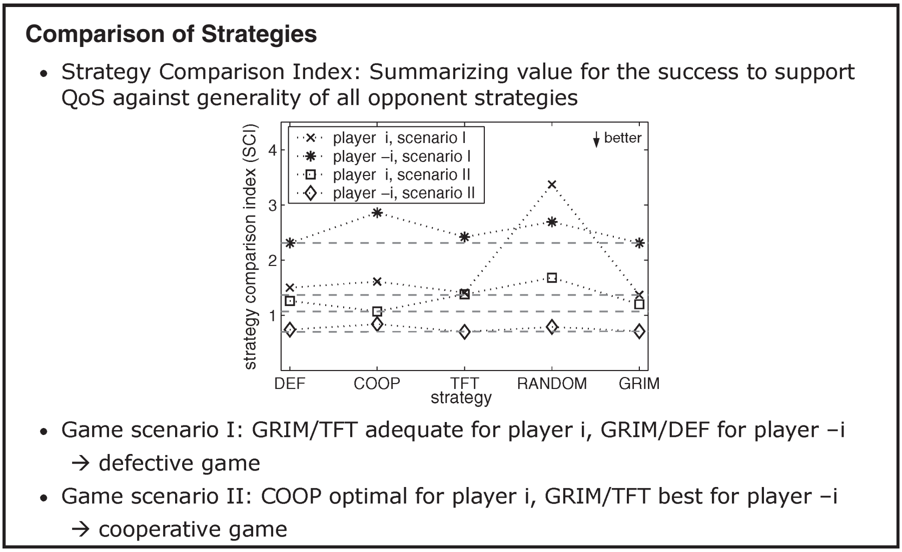


Figure 9.8 Strategy comparison of the success for QoS support with the help of a Strategy Comparison Index. Reproduced by permission of © 2005 IEEE[3].

of player $-i$, which show the same SCI value for the DEF strategy. The difference in the course of the MSG between the GRIM and DEF strategy is the behavior in the case of a cooperating opponent: Game-wide cooperation (C,C) leads to shorter delays for player i than (D,C). This is not the case for player $-i$ and thus the DEF and GRIM strategies are both preferable. In summary, defective strategies are to be favored in this game scenario by both players.

Now, an additional game scenario II with slightly different QoS requirements for player i of (throughput, period length, maximum delay) = (0.1, 0.05, 0.02) and for player $-i$ of (throughput, period length, maximum delay) = (0.4, 0.031, 0.02) is considered. The results are summarized in Figure 9.8 with the graphs marked through squares and diamonds. Here, the strategy of COOP is the best for player i and the TFT strategy for player $-i$. As player i has a required throughput of 0.1, the competition for the medium is less severe, compared to the previous scenario, leading to an advantage in strategies with cooperative game outcomes of (C,C). Here, the "best response" optimizes the players' utility with less destructive interference, i.e., blocking of the medium, for the opponent: Player $-i$ with a required throughput of 0.4 is more sensitive to the opponent's behavior and has a better QoS with the TFT trigger strategy.

9.1.8 Coexistence Among 802.16 Systems[4]

The realization of coexistence among 802.16 systems beyond the scope of 802.16h as introduced in Section 7.3.7 is briefly discussed in this section.

In the next section, mechanisms for solving the coexistence of 802.16 with 802.11a in blocking it out of the shared frequency channel are discussed. These mechanisms guarantee an exclusive spectrum usage right to the BS. This exclusiveness is lost if multiple BSs compete for spectrum access. This coexistence problem is similar to the coexistence problem of several 802.11e HCs sharing a single frequency channel in the time domain as discussed above. Modified 802.11e HCs are able to support QoS when sharing spectrum without requiring direct information exchange in observing past spectrum utilization. The proposed solution concept can be generalized to any MAC protocol that realizes spectrum access based on Time Division Multiplex (TDM).

The application of game theory results in periodic resource allocations that enable a distributed coordination based on past spectrum utilization. An application of these approaches for mitigating the coexistence problem of 802.16 benefits essentially from this periodicity. Nested MAC frame transmissions of multiple 802.16 BS can be compared to these decentralized coordinated periodic resource allocations.

9.2 Heterogeneous Coexistence – Unlicensed Operation of 802.16[4]

The coexistence of IEEE 802.16 (WiMAX) and IEEE 802.11 (Wi-Fi) in shared radio spectrum is an acute problem. The frame-based medium access of 802.16 requires rigorous protection against interference from WLANs in order to operate properly when sharing spectrum.

Various scenarios will arise where 802.16 might have to share spectrum with already deployed and operating WLANs of 802.11 such as in office or residential deployment scenarios. The U-NII frequency band at 5 GHz is one example of spectrum which might be shared between 802.16 and 802.11a (IEEE, 2003a).

WLANs of the 802.11 standard are able to coexist (Mangold *et al.*, 2001c), i.e., operate at the same time and location without harmful interference in using DFS and Transmit Power Control (TPC). More complex strategies are required, when QoS support is demanded: Successful, deterministic control of access to the radio resource is necessary for all coexisting wireless systems in order to guarantee QoS. The information exchange between spectrum-sharing networks enables an interworking but is not required for coexistence. Approaches without information exchange based on the observation of spectrum utilization are discussed in the context of horizontal spectrum sharing in Section 13.3.3. With interworking, wireless networks are able to coordinate spectrum usage among each other. A central coordinating device that combines the central BS of 802.16 with the Hybrid Coordinator (HC) of 802.11e is proposed in Section 8.3. This central device, called Base Station Hybrid Coordinator (BSHC) requires an operation in both protocol modes, 802.16 and 802.11(e), in order to realize interworking and coexistence.

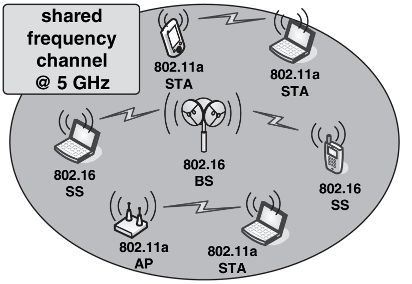
Contrary, we propose in this section software upgrades to the MAC of the 802.16 BS. Coexistence between 802.16 and 802.11 is enabled (without any data exchange) between both standards. Thereby, no 802.11 frame transmissions are required by an 802.16 system. We expect 802.16 systems to be available in laptops soon and then to provide wireless VoIP services that 802.11 cannot support satisfactorily well. As shown in this section, 802.16 is able to control access to a radio channel such that competing 802.11 systems only get access when permitted by an 802.16 BS. It is even possible for an 802.16 BS to push away any 802.11 system from a frequency channel at 5 GHz, if necessary.

For the regulatory restrictions of unlicensed operation in the frequency bands at 5 GHz the interested reader is referred to Chapter 3.

9.2.1 Coexistence Scenario

Approaches for enabling the coexistence of a single 802.16 system with multiple 802.11a devices (APs and STAs using the DCF for medium access) are introduced in the following. The basic

Coexistence Scenario of 802.16 and 802.11a



Basic OFDM Parameters and Resulting Time Values

802.11a + 802.16 - Frequency Band	@ 5 GHz
802.11a + 802.16 - Bandwidth	20 MHz
802.16 - OFDM Symbol Time	13.89 μs
802.16 - Contention Slot Duration for BW Requests	2 OFDM Symbols = 27.78 μs
802.16 - Contention Slot Duration for Initial Ranging	10 OFDM Symbols = 138.9 μs
802.11a - DCF access after idle time of DIFS with a duration of	34 μs

Figure 9.9 Coexistence scenario of 802.16 and 802.11a sharing the same frequency channel in the 5 GHz band. Basic OFDM parameters and resulting time values are depicted. Reproduced by permission of © 2006 IEEE[4].

idea is to prevent medium access of 802.11a before and during MAC frame transmission of 802.16. The proposed solutions target at the avoidance of an idle medium with a duration equal or longer than DIFS.

The coexistence scenario considered in this section is illustrated in Figure 9.9: One BS and controlled SS are depicted. They are operating at a frequency channel at 5 GHz that is shared with multiple 802.11a devices. Details of the PHY of 802.11a are described in Section 5.3, while the 802.16 PHY is introduced in Section 7.4. We assume that the 802.16 system has selected this frequency channel in using DFS according to the regulatory restrictions for the 5 GHz band. In the following, three risk points are identified that imply a danger for the interference-free transmission of the 802.16 MAC frame. These dangers and their handling are illustrated in the timing diagram of Figure 9.10 and are marked with (1.), (2.) and (3.), respectively. It is worth mentioning that all mechanisms proposed in this section have in common that no multi-mode device capable of operating according to 802.16 and 802.11 is required. A manipulation of the 802.11 devices' NAVs in order to prevent unwanted allocation attempts requires for instance such a multi-mode device. This and similar concepts are discussed and evaluated in Section 8.3 under the umbrella of heterogeneous multi-hop networks.

The MAC frame duration of 802.16 can be varied between 2.5 to 20 ms, while the beacon interval (often referred to as a superframe) of 802.11 typically has a value of 100 ms. Thus, multiple 802.16 MAC frames are nested into one 802.11 superframe. We assume in the following that the 802.16 system allocates only a fraction of the shared frequency channel for its own operation in dependency on its current traffic load. The time interval between two consecutive 802.16 MAC frames is accessed by the coexisting 802.11 devices in using the DCF, as also illustrated in Figure 9.10. We further assume that TTG and RTG are shorter than the DIFS duration interval of 802.11a.

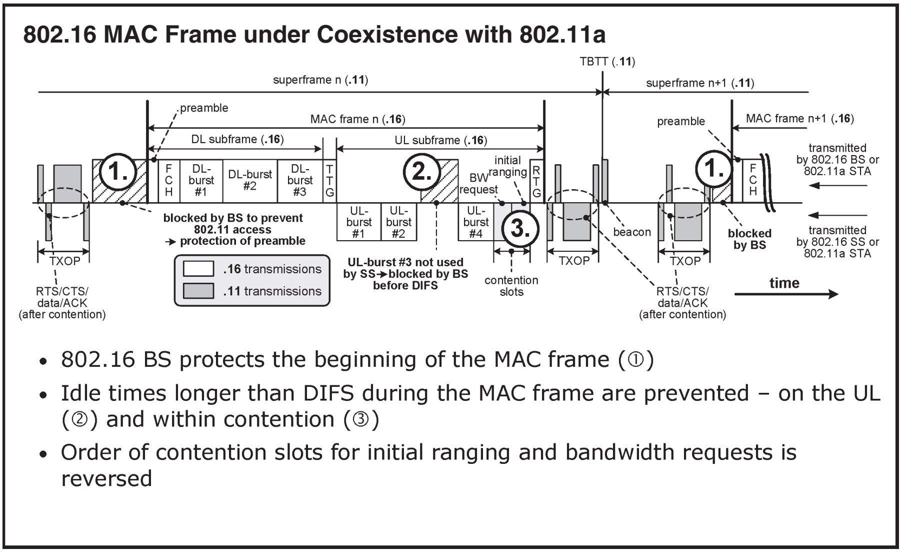


Figure 9.10 Timing diagram of an 802.16 MAC frame transmitted on a shared frequency channel. Reproduced by permission of © 2006 IEEE[4].

9.2.2 Protecting the Beginning of 802.16 MAC Frame

Contrary to 802.11, 802.16 is not able to tolerate a delayed beginning of its MAC frame. Therefore, no 802.11a transmission is allowed to be ongoing when the 802.16 MAC frame is scheduled for beginning (with a preamble and the FCH). The BS therefore blocks the medium before the intended frame initialization in order to prevent an access from 802.11. The medium has to be blocked as soon as the time instance of the next 802.16 MAC frame is closer than the maximum duration of an 802.11 transmission. With the most robust PHY mode and the largest data packet size (2346 byte) the maximum duration of 802.11a transmission is approximately 2 ms (Mangold, 2003). Thus, the blocked time interval has in the worst case a duration of this 2 ms. An intended interfering of an ongoing 802.11 transmission that reaches into this blocking interval is not required. Only a new allocation attempt after this transmission has to be prevented. Consequently, the effective duration of the blocked time interval may differ from one MAC frame to the next one, which is also illustrated in Figure 9.10 ((1.)). As the concrete beginning point of time of the blocking is not under the control of the BS, the blocked time interval is difficult to use for 802.16 user data transmissions. The BS is not able to notify the SSs in the previous MAC frame about the exact beginning of the transmission used for blocking. Nevertheless, the BS may broadcast data on a best-effort basis as a DL burst, although this data has only an unreliable chance of being scheduled for transmission.

In total, 20 % of the transmission time/capacity is wasted in the worst case for guaranteeing a timely beginning of the MAC frame. This can be seen as cost or effort for operating 802.16 and thus supporting QoS in unlicensed frequency bands shared with 802.11.

9.2.3 Protecting the 802.16 UL Subframe

The 802.16 system transmissions in the DL subframe are completely under the control of the BS and no 802.11 device has the opportunity to access the medium: The medium is never idle for a duration equal or longer than DIFS. In the case of the DL subframe this control is lost: A BS may not allocate an assigned UL burst and the medium may become idle again. In this case, the BS has to block the unallocated UL burst with an own transmission before an idle time duration of DIFS, as depicted in Figure 9.10(2.). This is done to prevent a medium access of 802.11 and to protect the following UL bursts of the ongoing UL subframe.

9.2.4 Shifting the Contention Slots

The contention slots of 802.16 lead to a vulnerability of the 802.16 MAC frame in the face of potential 802.11a transmissions. These slots are used in the UL subframe for initial ranging and bandwidth requests by SSs. The random access to these slots follows the slotted ALOHA principle described in Section 2.6.1. Unallocated 802.16 contention slots may lead to an idle medium with a duration equal or longer than DIFS. In this case, 802.11 devices might access the frequency channel and destroy the ongoing 802.16 MAC frame. The duration of contention slots in 802.16 depends on the used frequency band. Here at 5 GHz with a bandwidth of 20 MHz, the contention slots used for bandwidth requests have a duration of 27.78 μs. The slots used for initial ranging are essentially longer and have a duration of 138.9 μs as summarized in the table depicted in Figure 9.9. Consequently, the medium access of the 802.11a DCF after 34 μs implies a danger for the 802.16 MAC frame.

802.16 allows a rearrangement of the MAC frame structure. The contention slots are therefore scheduled after at the end of the UL subframe in order to protect the UL bursts as depicted in Figure 9.10 ((3.)). In this way, an interference of UL bursts due to an 802.11a medium access in the contention phase of 802.16 is prevented. The contention slots for bandwidth requests are shorter than DIFS. At least the first two slots may be used for requests by SSs without interference from 802.11a STAs. The contention slots used for initial ranging are essentially longer than DIFS. Thus an access of 802.11a STAs might only be prevented by the BS in blocking each unallocated contention slot if it is idle, similar to the blocking of unused UL-bursts as described above. The order of the contention slots from Figure 7.6 is therefore reversed here: First the slots for bandwidth requests are scheduled and thereafter the slots for initial ranging as also illustrated in Figure 9.10((3.)).

Ideally, associated SS do not need bandwidth requests in contention slots to change the capacity assigned to them by the BS. Usually this done in piggybacking these requests to scheduled UL transmissions in order to avoid the contention with other SSs in the bandwidth request phase.

9.3 Summary and Conclusion

The transfer of solutions concepts from game theory and social science to the competition of radio resource sharing wireless networks enriches research with a new interdisciplinary aspect. Especially the consideration of multiple QoS parameters in the players' coordination efforts are a decisive step towards a realization as extension of QoS-supporting wireless communication protocols.

The application of game models enables an aimed interaction and provides an analysis of the competition for the utilization of a shared radio spectrum. Our analysis and simulation results indicate that cooperation is an achievable equilibrium that often improves the overall spectrum efficiency. Traffic requirements that are imposed by services and applications determine whether the selected strategies should pursue cooperation, or ignore other radio systems leading to games of defection. In such defective environments, a regulating intervention, as for instance a specification of certain MAC parameters, would be advantageous to enable nevertheless a guarantee of QoS in all coexisting wireless networks. The learning in games to facilitate overcoming insufficient information about the opponents and game models of multiple players are next steps to mitigate the mutual interference of radio resource sharing wireless networks.

In the second part of this chapter mechanisms for enabling an operation of IEEE 802.16 in spectrum shared with 802.11(a) have been proposed. Coexistence is enabled in partly blocking 802.11(a) out of the medium. This enables a guarantee of QoS in 802.16. A drawback of operating 802.16 in unlicensed frequency bands shared with 802.11 is identified: In the worst case, approximately 20% of the available transmission time is wasted leading to a reduced efficiency of channel/spectrum usage.

The coexistence of multiple 802.16 BSs introduces an additional level of complexity. Its solution requires intelligent algorithms for distributed coordination that imply intensive impacts on the medium access control. Contrary, the solution proposed in this section requires only minimal modifications of the medium access in the BS.

The practical realization of the proposed solution is challenging but nevertheless feasible related to the required transceiver turnaround times in the 802.16 BS.

From the perspective of 802.11a, the proposed method can be regarded as unfair. Unfortunately, 802.16 requires such a rigorous protection against interference of its MAC frame from other communication systems because of its inability to coexist. A fundamental regulatory recommendation for operation in unlicensed frequencies shared in the time domain can be derived from our proposed solution: A limitation of spectrum access in its duration and the usage of deterministic spectrum allocation patterns allowing a distributed coordination on the basis of spectrum observation. The fulfillment of this recommendation would allow QoS support (Berlemann, 2006), mitigate the coexistence problem of multiple 802.16 systems and would introduce fairness, when sharing spectrum with dissimilar systems.

Notes

1. Reproduced by permission of © 2005 IEEE. Source: L. Berlemann, S. Mangold, G. R. Hiertz, and B. Walke, "Radio Resource Sharing Games: Enabling QoS Support in Unlicensed Bands," *IEEE Network, Special Issue on "Wireless Local Area Networking: QoS Provision and Resource Management"*, vol. 19, no. 4, pp. 59–65, July/August 2005. (Section 9.1, 9.1.1–9.1.7).
2. Reproduced by permission of © 2004 IEEE. Source: L. Berlemann, B. Walke, and S. Mangold, "Behavior Based Strategies in Radio Resource Sharing Games," in Proc. of *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, 5–8 September 2004. (Fig. 9.3 and 9.7).
3. Reproduced by permission of © 2005 IEEE. Source: L. Berlemann, G. R. Hiertz, B. Walke, and S. Mangold, "Strategies for Distributed QoS Support in Radio Spectrum Sharing," in Proc. of *IEEE International Conference on Communications, ICC 2005*, Seoul, Korea, 16–20 May 2005. (Fig. 9.8).
4. Reproduced by permission of © 2006 IEEE. Source: L. Berlemann, C. Hoymann, G. R. Hiertz, and B. H. Walke, "Unlicensed Operation of IEEE 802.16: Coexistence with 802.11(a) in Shared Frequency Bands," in Proc. of *17th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2006*, Helsinki, Finland, 11–14 September 2006. (Section 9.1.8, Section 9.2, Fig. 9.9 and 9.10).

10

Broadband Cellular Multi-hop Networks

Bernhard H. Walke, Ralf Pabst and Daniel C. Schultz

In recent years the importance of multi-hop networks has been growing: their areas of application are expected to substantially expand in the near future. Originating in indoor WLANs, multi-hop networks are applied in various scenarios of wireless communication, for example in long-range, outdoor communication with a fixed multi-hop cellular network. Activities of standardization in IEEE 802 touching on multi-hop and meshed networks have been discussed in the previous chapter. The focus of this chapter lies on extensions and possible enhancements beyond the current state-of-the-art IEEE 802 and targets at giving impulses on standardization activities in the Working Groups (WGs) of 802.

After a definition of terminology in Section 10.1, this chapter proceeds with a motivation for the introduction of multi-hop concepts in next-generation wireless networks in Section 10.2. Thereafter, Section 10.3 discusses related work in the area of relay-based networks and distinguishes them from other work in the area of multi-hop networks. Section 10.4 shows how multi-hop concepts based on fixed relays can be applied to improve network coverage and capacity of wireless broadband networks. This chapter is summarized with a conclusion in Section 10.5.

10.1 Definitions

In a multi-hop network no direct link exists for communication between data source and data sink. Therefore, data is forwarded by relay stations in order to reach its destination. A relay station may also be referred to as a forwarding or routing station and may be a mobile or fixed terminal. The link between two neighbored relay stations on the route of data packets traveling through a multi-hop network is referred to as a “hop” and the route, accordingly, is called a multi-hop route. Multi-hop operation implies interworking and interdependencies of communication between neighbored stations and links, respectively. Different from the decoupled data reception

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

and forwarding if the hops can be regarded independent from each other, e.g., an Access Point (AP) connected by a microwave link to the wired backbone network, in multi-hop networks packet transmission on sequential links is affecting each other.

Depending on the Medium Access Control (MAC) protocol applied, multi-hop communication may happen under central or distributed control. Under central control, the network element coordinating communication may be a Base Station (BS), an AP or even some mobile station. Since spectrum access is coordinated by this element, collisions of packets transmitted by stations within the range of the central controller in a single-hop network are avoided. Under distributed control, collision avoidance is more difficult to realize and requires intelligent algorithms for distributed coordinated channel access. With multi-hop, collisions are hard to avoid under both central and distributed control.

In the literature, the term BS is usually applied in cellular systems, while the term AP is more common in the context of WLAN-type systems. Both have in common a direct interface to the wired backbone network. A Relay Station (RS) differs from this in that it has interfaces to wireless links, only.

Multi-hop cellular is a promising candidate technology for next-generation wireless networks, where a cell is served by a BS supported by multiple Fixed Relay Stations (FRSs). In these networks, even some mobile stations may serve as relay to forward data in order to increase the coverage area or improve performance of the multi-hop enabled cell.

Multi-hop networks that apply one radio interface technology throughout are referred to as homogeneous, while networks applying different radio interfaces on a multi-hop route are called heterogeneous multi-hop networks.

In a mesh network, more than one multi-hop route may exist to route packets from data source to sink. A more general view of a mesh network implies that all network elements may communicate with each other. We differentiate mesh networks into ad-hoc and infrastructure based mesh networks depending on the relay station mobility and resulting stationarity of the network topology. Infrastructure mesh networks may be flexible and extendable, in general, but change their topology much less frequently compared to ad-hoc mesh networks.

This chapter's focus is on multi-hop operation of ISO/OSI/IEEE 802 Reference Model layer 2 protocols, more precisely MAC protocols referred to as L2 relaying. Physical layer relaying (repeater) based multi-hop networks are discussed in the context of cooperative relaying below.

10.2 Rationale

The transmission range of a broadband radio interface such as the one envisaged by advanced wireless and 4G systems is limited by the high attenuation of radio waves at high carrier frequencies (beyond 3.4 GHz), a limited transmission power (EIRP) owing to regulatory constraints and unfavorable radio propagation conditions, e.g., indoors and in densely populated areas. Conventional cellular radio network deployments would require a very high density of base stations to achieve sufficient radio coverage under these conditions. As a consequence, the system deployment cost in terms of Capital Expenses (CAPEX) and Operational Expenses (OPEX) for broadband radio would increase dramatically if traditional deployment concepts are applied, resulting in a high cost per bit transmitted.

It is well known that an increased data rate (for a given power and carrier frequency) leads to a reduced radio range and that the available data rate decreases with increased distance from an AP as illustrated in Figure 10.1. In general, the service quality in terms of data rate, delay, outage probability, etc. seen by the user must not depend on its location or distance from the AP in a cell.

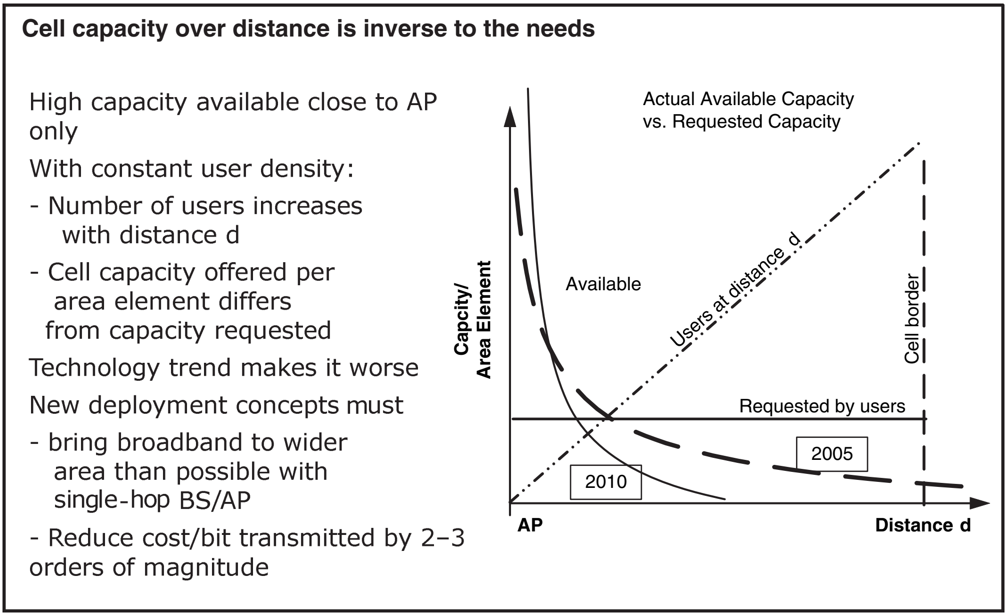


Figure 10.1 Motivation for introducing relay-based technologies.

Assuming a constant number of users per area element in a cell, the number of users increases linearly with the distance *d* from the BS/AP. It appears reasonable that the requirements on future broadband radio systems in terms of capacity, delay, user-experienced data rate and deployment cost cannot be met using conventional cellular deployment concepts. Instead, a novel disruptive deployment concept is needed.

The basic property exploited in relay-enhanced deployments is a link budget saving on the two consecutive relay links connecting AP to RS and RS to User Terminal (UT), compared to the budget for direct communication between AP and UT, which must be traded against the need to twice transmit the same data over the channel.

An RS introduced to the communication path induces two hops. Under the assumption that the RS is positioned approximately halfway between AP and UT, the two hops each reach half the distance of the single-hop link. Since the pathloss is a power law of distance (see Section 2.1.1) with a pathloss coefficient 3 or larger, in most real-world environments a total pathloss reduction by at least 25% can be expected – at the cost of transmitting data twice.

Virtually all modern wireless systems apply Link Adaptation (LA) based on combinations of Modulation and Coding Schemes (MCS) and do supplement this by Automatic Repeat Request (ARQ) protocols, see Chapter 2. LA leads to a highly nonlinear throughput-versus-distance relationship that may result in the link budget savings mentioned. An increased effective throughput on a two-hop link may be achieved in two ways: Reduced pathloss at each hop under a given MCS will reduce the probability for retransmission. Even more important, reduced pathloss enables use of a less robust but higher data rate MCS. It is important to note that the parameters of the radio channel between AP and FRS can be estimated safely, since FRSs do not move. In addition MIMO techniques may be applied to exploit the spatial dimension of the channel, leading to an increased throughput performance.

If carefully engineered, the benefits gained from higher throughput per relay link may outweigh the effort of transmitting data twice, resulting in a higher spectrum efficiency of multi-hop compared to single-hop, as reported in Pabst *et al.* (2005).

Hence, to meet the goal of low-cost radio network deployment for both metropolitan and ubiquitous (wide-area) coverage, deployment concepts based on fixed layer-2 relay stations appear to be an interesting technology. Relay stations do not need a wired (fiber) backbone access but only access to the mains or a solar-panel fed battery. This reduces deployment costs (CAPEX and OPEX) and introduces high flexibility in relay positioning, allowing a fast network rollout and adaptive traffic capacity engineering. Relays in cellular may also be used to provide indoor coverage from outdoor BSs.

10.3 Related Work

Relaying for cellular following a survey paper (Yanikomeroglu, 2002) may be based on: (i) mesh network; (ii) multi-hop network; (iii) Opportunity Driven Multiple Access (ODMA); (iv) user cooperative diversity; and (v) multi-user diversity. All papers cited in Yanikomeroglu (2002) either focus on layer-1 relays (repeater) or on a combination of cellular and ad-hoc networks with mobile relay stations. Mesh networks (i) to connect fixed nodes in tandem to a BS, as specified in IEEE 802.16, have not found acceptance so far. Mobile Multi-hop Relaying (MMR) is under standardization in a new Task Group 802.16j as introduced in Section 8.1.4.2. Cellular multi-hop networks (ii) are represented by Lin and Hsu (2000) "where the service infrastructure is constructed by fixed bases, and it also incorporates the flexibility of ad-hoc networks where wireless transmission through Mobile Stations (MS) in multiple hops is allowed". Cellular combined with ad-hoc results in high complexity of organization and random QoS supported (Aggelou and Tafazolli, 2001; Walke and Briechle, 1985; Zadeh and Jabbari, 2001). The first suggesting this (Harrold and Nix, 2000) name power reduction and coverage extension as main benefits. According to Harrold and Nix (2000) the capacity could be enhanced compared to single-hop cellular. No real network uses ad-hoc multi-hop relaying. 3GPP adopted from Salbu (1978) a method (iii) and named it ODMA (3GPP, 2004). ODMA is an option that has not been deployed so far, although recommended by Rouse *et al.* (2002), "ODMA remains attractive …due to advantages offered by a reduction in transmission power, potentially enhanced coverage and with a greater trade-off possible between QoS and capacity in the extended coverage region, and under certain circumstances may show increased capacity". "An economic system deployment based on a new broadband radio interface requires a new network architecture based, e.g., on multi-hop or repeater concepts" is stated in Mohr *et al.* (2002).

User cooperative diversity (iv) and multi-user diversity (v) involve layer-1 repeating that is combined with macro-diversity to gain from receive signals combining, and is called cooperative relaying, see Figure 10.2 from Pabst *et al.* (2004) and Zimmermann *et al.* (2003). Cooperative relaying may be used to support L2-relaying.

The difference between the fixed multi-hop concept introduced in this chapter to enable a low-cost broadband radio network for short-range, metropolitan and wide-area (Esseling *et al.*, 2000; Esseling *et al.*, 2004; Pabst *et al.*, 2004; Pabst *et al.*, 2005; Walke *et al.*, 2001; Wijaya and Zirwas, 2003) compared to all other designs is that fixed L2-relay stations are part of the cellular infrastructure and appear to their local environment like a BS. This allows the engineering of the radio network infrastructure such that QoS can be guaranteed and subscribers to a network can be attracted. In this way, the Relay Enhanced Cell (REC) is introduced. Many industry research groups have taken up the REC concept for current and future cellular, recently, e.g., Samsung,

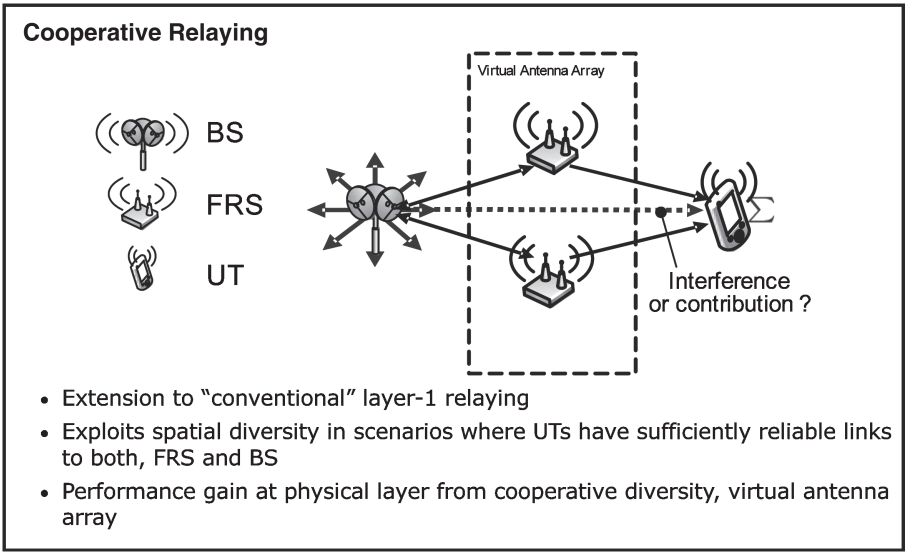


Figure 10.2 Combination of source signal and repeated signal in cooperative relaying (Pabst *et al.*, 2004; Zimmermann *et al.*, 2003).

Sony (Kusuda *et al.*, 2005), NTTDoCoMo, Ericsson, Nokia and Siemens (Mohr, 2005; WINNER, 2005).

10.4 Relay-based Deployment Concept for Cellular Broadband Networks

The proposed concept of using layer-2 relays as fixed infrastructure elements in an infrastructure based cellular deployment (Esseling *et al.*, 2000; Pabst *et al.*, 2004; Walke, 2000) differs from other concepts that assume the relays as mobile nodes, which randomly enhance a fixed infrastructure in an ad-hoc manner (Lin and Hsu, 2000). In the following the relays are denoted FRS, although they could also be movable, since they are deployed temporarily fixed, in order to increase the capacity in certain service areas with a temporary hotspot characteristic, such as exhibitions, events or disaster areas.

Figure 10.3 aims to illustrate how mobile broadband radio coverage can be provided by means of a hierarchical layered system architecture combining different relaying concepts. The dotted circles around APs and FRSs denote the area where broadband radio coverage is available from these Radio Access Points (RAPs) to connect UTs. APs feed multi-hop routes across FRSs based on a broadband radio interface that is used both for the feeding links and the radio access to RAPs. The APs together with the tiers of FRSs placed around them together feed what is called an REC, outlined by a thick black REC border line in the middle of Figure 10.3. The shape of the REC is formed from its RAPs connected multi-hop and the buildings in the service area. Indoor coverage is provided from outdoor RAPs positioned close to building APs are assumed to have fixed network connectivity, either by fiber or point-to-multipoint microwave radio links as shown in the figure. APs connected to a microwave radio link could be considered as heterogeneous relays, since they operate different radio interface technologies in tandem. The

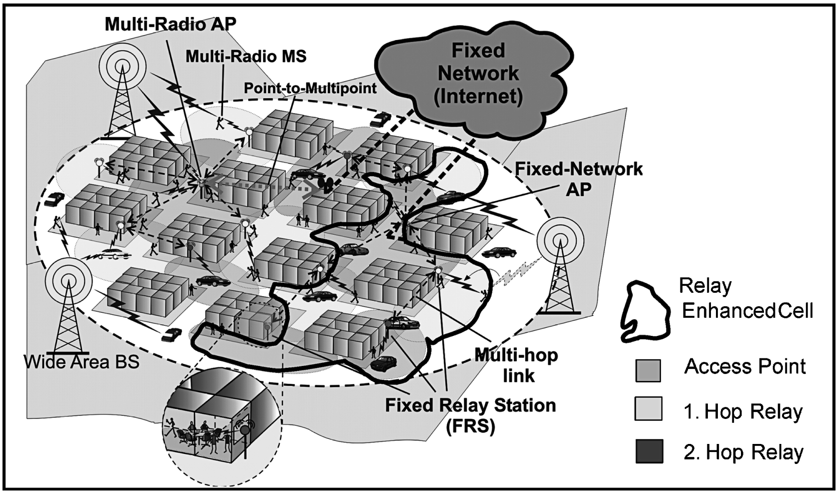


Figure 10.3 Relay-based mobile broadband system for densely populated areas.

wide-area radio network (shown by means of hexagons in the background, served by wide-area BSs) serves to provide continuous radio coverage for terminals roaming at high mobility.

10.4.1 Relaying Use Cases

As mentioned in Section 10.2, in infrastructure-based deployment concepts FRSs may be applied in different cases, which are described in detail in the following.

10.4.1.1 Relay to Increase Coverage Range

FRSs introduced to a cell (to become a Relay Enhanced Cell – REC) may be used to enlarge the coverage area of the BS or AP as shown in Figure 10.4. If an FRS is placed outside the coverage area of the BS/AP, antenna gain is needed to connect BS/AP and FRS. The higher the antenna gain used on the BS–FRS link, the larger is the capacity transferred from BS/AP to the FRS. As FRSs are placed outside the coverage area of a BS/AP, UTs served by FRSs cannot receive broadcast information from the BS/AP, so the broadcast channel must be forwarded by the FRS.

The concept of FRS also enables installing temporary coverage to an area not needing permanent coverage (e.g. construction sites, conference and meeting rooms) and enables a fast and cost-efficient initial network roll-out.

FRSs are considered as a main trend for providing radio coverage in wireless broadband networks and extend the service range, since there is a clear trend towards increased transmission rate of radio transceiver modems resulting from technology progress, see Figure 10.1. Modems in the future will tend to provide over-capacity for the more and more shrinking local service area of an AP, since higher data rate under given power limits always means a reduced radio range. Relay stations may substantially increase the size of the service area of an AP/BS and may increase thereby the probability that the high capacity offered by an AP/BS will be used effectively.

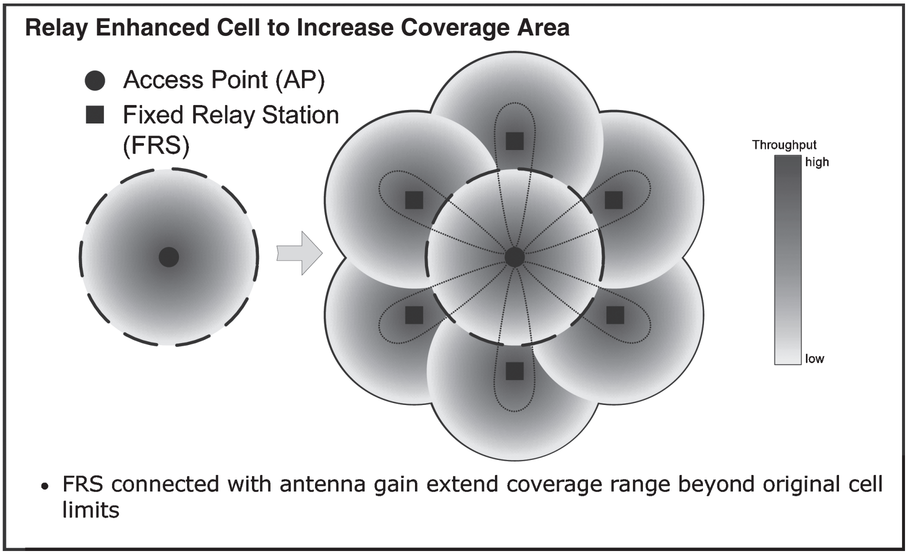


Figure 10.4 Left: Conventional cell. Right: REC using FRS to enlarge the cell coverage area. Reproduced by permission of © 2006 IEEE[1].

10.4.1.2 Relay to Increase Cell Capacity

FRSs may be used in order to increase the capacity at outbound cell regions of a given cell as shown in Figure 10.5. In both scenarios shown in Figure 10.4 (right) and Figure 10.5 (right) the capacity per area element in the REC approximates the requested capacity (see Figure 10.1) more closely than it would be possible with a conventional (single-hop) cell, as visible from the dark shaded areas in Figures 10.4 and 10.5. For a cellular radio deployment the channel reuse distance is minimized when receive antenna gain instead of transmit antenna gain is used.

Relays to increase capacity inside a cell, as shown in Figure 10.5, can also be used to minimize the transmission power needed by UT, BS and FRS. It is referred to as the Power Minimizing concept (Lin and Hsu, 2000), allowing UTs to benefit from reduced energy consumption, while the reduced output power at BS and FRS leads to reduced hardware costs. Different from the area optimization concept as presented in Figure 10.4, all UTs in the REC may be able to receive the broadcast channel from the BS, so that it must not be relayed.

10.4.1.3 Relay to Cover Locations Heavily Shadowed from Access Point

A capability of FRS not available from traditional radio network deployment concepts is to provide radio coverage to locations otherwise shadowed from the AP/BS. In current technology, repeaters (layer-1 relay) are being applied but have been found to contribute much to intercell interference. Figure 10.6 gives an example by means of the Manhattan Grid Scenario (3GPP, 1998) demonstrating how two hops across FRSs may be used to serve an UT that otherwise would be shadowed from the AP. A single-hop path to the UT that is obstructed heavily is replaced by a three-hop multi-hop path, each served under Line-of-Sight (LOS) propagation conditions. One aspect worth mentioning is the reuse of the radio channel for the multi-hop link, e.g., for hops 1 and 3, as discussed in Section 8.2.

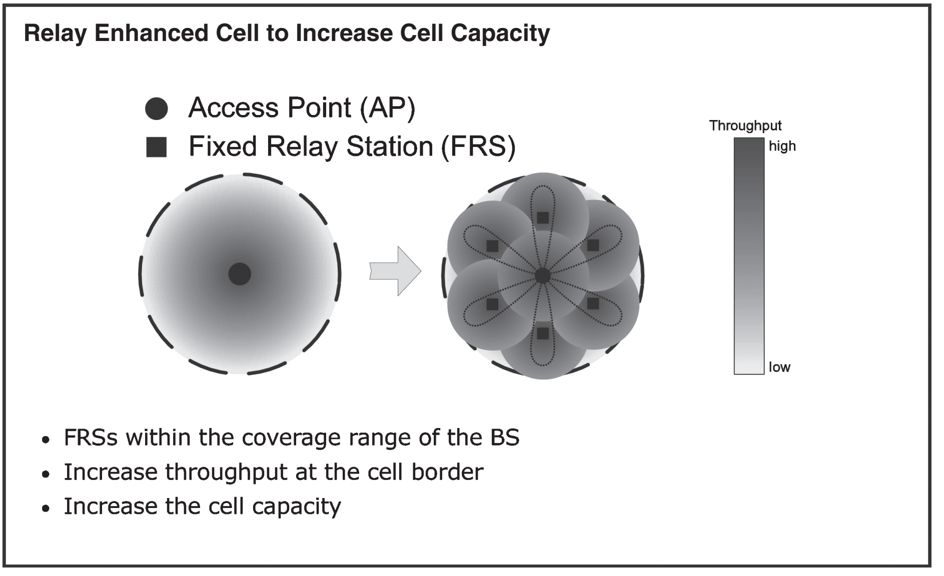


Figure 10.5 Left: Single AP cell. Right: REC with FRS to increase the capacity at the cell border and balance the capacity per area element. Reproduced by permission of © 2006 IEEE[1].

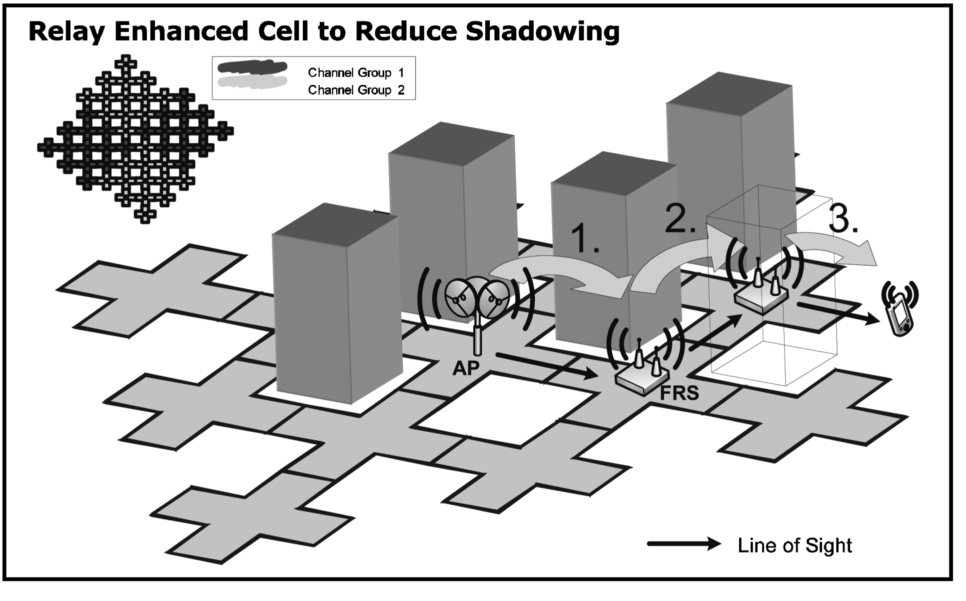


Figure 10.6 Increasing the network coverage quality and enabling spatial reuse of spectrum through introduction of relays in a highly shadowed environment.

Figure 10.6, in the left-hand upper part, also suggests how a continuous multi-cellular radio coverage can be provided to the Manhattan Scenario, e.g., applying a cell cluster of order 2, based on two channel groups, each comprising a number of disjunct frequency channels used in an REC. The cluster order suitable results from the acceptable amount of mutual intercell interference of adjacent RECs.

10.4.1.4 Exploiting Spatial Separation of Subcells in REC

So far it has been explained that FRSs may suit to serve an area in a cell heavily shadowed from an AP, thereby introducing a subcell around an FRS. Different from traditional cellular network deployment, a relay enhanced cell may contain many subcells either served by the AP or FRSs, and may benefit from it, if subcells are shadowed mutually by reusing the radio channel. Obviously, scheduling the channel resources available in an REC for sequential use by a single subcell is sub-optimal, only, since one subcell can be active at any time, possibly wasting the capacity of subcells. As shown in the following, the overall capacity of a two-hop cell can be enhanced by exploiting the spatial separation of some of the subcells. Spatial separation means that the subcell areas served by two or more FRSs are sufficiently shadowed to each other, as shown in Figure 10.7 for the pairs FRS#1 and FRS#2, FRS#2 and FRS#3, etc. With spatially separated subcells it may be that neither FRS#1 nor any UT in the subcell served by FRS#1 will cause harmful interference to another subcell served by FRS#2 and vice versa. This may be exploited for spatial reuse operation of both subcells resulting in a sequence of time allocations to subcells as shown in at the bottom of Figure 10.7. As visible, a full service cycle starts with consecutively serving FRS-based subcells #1 to #4, then, two out of four FRSs serve their UTs simultaneously, namely FRS#1 and #2, followed by FRS#3 and #4, leading to a parallel allocation of the channel. The cycle ends with the AP serving its subcell. This approach obviously is feasible

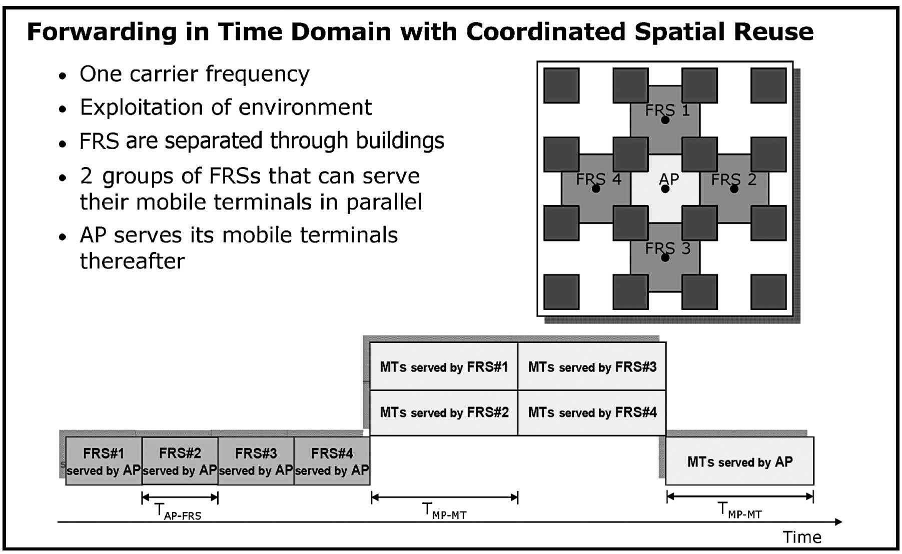


Figure 10.7 Relaying in the time-domain at a single-carrier frequency which is reused in a coordinated way.

only in an infrastructure engineered for spatial channel reuse and can most easily be applied with fixed RSs. The principle explained may be applied to schedule channel allocations of adjacent RECs such that mutual interference remains acceptably low and multi-RECs wide area coverage is provided, as studied in (Schultz *et al.*, 2003).

10.4.2 Estimation of Subcell Capacity in a Relay Enhanced Cell

In order to establish the feasibility of relay-based cellular concepts it needs to be shown that sufficient traffic capacity can be provided even at the end of a multi-hop route to serve UT roaming there. We therefore present analytical estimates of basic parameters of a REC, namely:

- multi-hop throughput in cellular deployment
- subcell capacity served by an FRS
- capacity of multi-hop links under delay constraint.

The multi-hop system studied is assumed to use the OFDM modem specified by IEEE 802.11a (IEEE, 2003a) as introduced in Section 5.3. Link level simulation results (Khun-Jush *et al.*, 1999) for the Packet Error Rate (PER) as a function of the Signal-to-Noise Ratio (SNR) from a 5 GHz H/2 system with 20 MHz channel bandwidth have been used as a basis for the analysis. The throughput of the proposed multi-hop system has been calculated as physical layer throughput without considering protocol header overhead. Retransmissions from a Selective-Reject ARQ protocol, as described in Section 2.5.2.3, are taken into account. Further we assume an ideal link adaptation.

10.4.2.1 Multi-hop throughput in Cellular Deployment

For an 802.11a like modem, assuming a hexagon-based cellular deployment, the maximum end-to-end throughput achieved by UTs in a REC is evaluated in (Esseling *et al.*, 2004). Figure 10.8 compares the results: Figure 10.8(a) shows the throughput over distance from AP in a single-hop cellular deployment for cluster order $C = 7$. The end-to-end throughput for a four-subcell REC in cluster order $C = 7$ is shown in Figure 10.8(b) with an AP serving three FRSs, as visible from the small pictogram. For fairness of comparison, the radius of the single-hop cell is set to 346 m, while the radius of the subcells in the REC is set to 200 m, resulting in equal service areas for the single-hop cell and the REC. Interference from the first ring of six co-channel cells/RECs is taken into account for calculating throughput, assuming the worst case, that all co-channel cells are fully loaded and simultaneously transmit. Coordination of scheduled time to transmit is only applied within the central cell/REC analyzed. The analysis takes into account that data designated to UTs in subcells served by the FRS need to be transmitted twice. Further, a protocol overhead of 25% for signaling and a receive antenna gain of 12 dB of the AP-FRS link have been assumed.

As a result, Figure 10.8 shows a slight reduction of the achievable throughput on the first hop in the REC, compared to that of the single-hop cell owing to the signaling overhead required for relaying. In return for introducing FRSs, the maximum throughput in the outbound regions of the REC is considerably increased compared to the conventional cell and capacity per area element is more fairly distributed than possible without FRSs.

10.4.2.2 Subcell Capacity served by an FRS

In the following we study the traffic capacity of subcells in an REC assuming a homogeneous distribution of UTs in the REC. The aim is to find the maximum FRS capacity possible if the

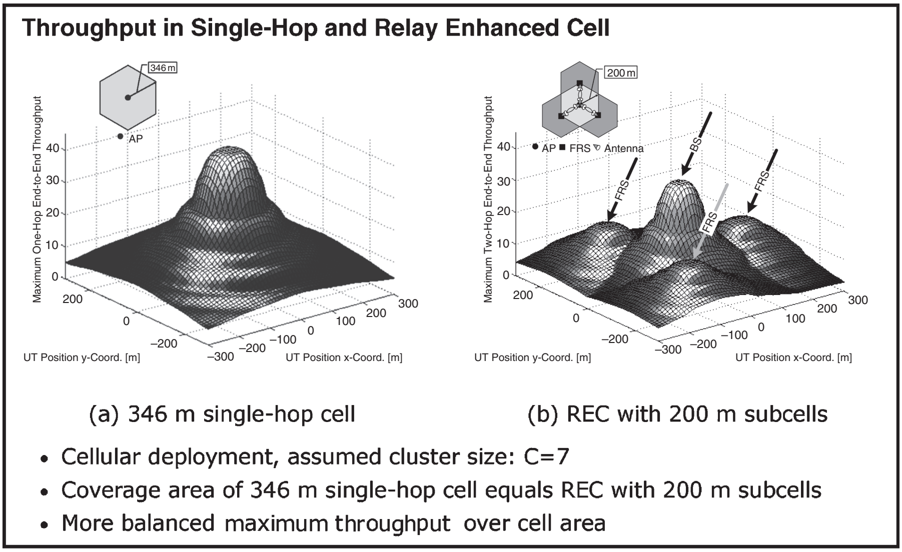


Figure 10.8 Comparison of throughput in single-hop and REC with three FRS.

AP only serves that FRS, with a varying antenna gain on the AP-FRS link. The left-hand side of Figure 10.9 shows the Manhattan Scenario considered. The right-hand graph in the figure shows the capacity of the subcell served by the AP and of subcell #1 served by FRS#1, both versus the receive antenna gain at FRS#1. The capacity in the subcell served by the AP in single-hop operation is independent of the antenna gain and amounts to 22.5 Mbit/s. The capacity available in the subcell served by FRS#1, when the whole capacity of the AP is transferred to it, grows with increased antenna gain from values of 2.7 MBit/s (no gain) until 15.9 Mbit/s at 30 dBi gain. An amount of 6,7 Mbit/s of the AP capacity, marked in Figure 10.9, is spent for transferring capacity from the AP to FRS#1 subcell by relaying.

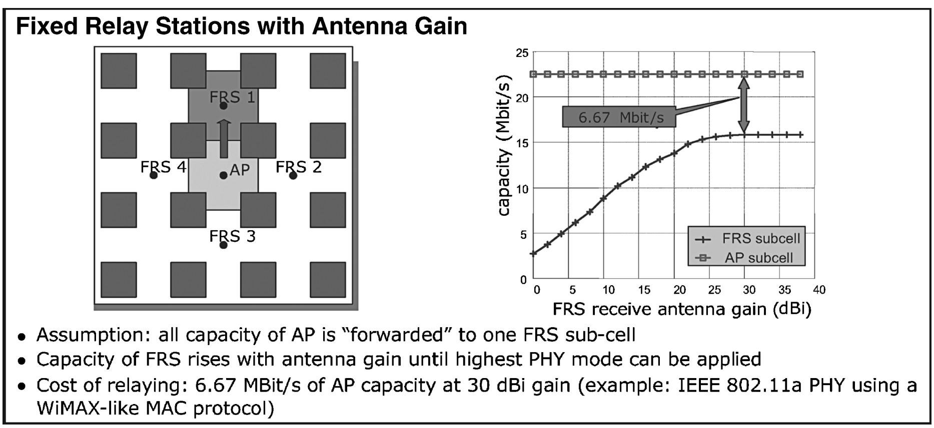


Figure 10.9 Capacity at fixed relay station with antenna gain.

An antenna gain higher than 28 dBi does not increase the capacity of the FRS subcell anymore, since at this value the highest MCS of the assumed modem is being used on the link without any more packet errors. For the scenario considered and parameters chosen approximately 29 % of the original single-hop cell capacity is spent to transfer 71 % of the AP capacity to one FRS.

10.4.2.3 Capacity of Multi-hop Links under Delay Constraint

One concern is the additional delay introduced by relaying packets and the throughput possible on a multi-hop link under delay constraint. The question arises under what circumstances two-hop relaying would be preferential to one-hop communication. Figure 10.10(a) compares the end-to-end throughput achieved with one-hop and two-hop transmission for the two scenarios depicted in the upper right corner of the figure under LOS radio propagation conditions. To account for the delay, an upper bound of 10 ms for end-to-end delay has been introduced as a constraint for both single- and two-hop link.

Without loss of generality it is assumed that the FRS is placed at equal distance from the AP and the UT. More detailed evaluations, also taking into account varying ratios of AP-FRS and FRS-UT distances, can be found in (Pabst *et al.*, 2005).

It turns out that the two-hop link from a distance of 240 m on can deliver a slightly higher throughput than the one-hop link, as visible from the light shaded area, keeping the 10 ms end-to-end delay bound. It is worth noting, too, that the range the one-hop link can cover is limited to about 380 m, while the two-hop link allows a range of 500 m, still fulfilling the 10 ms delay constraint. Under a somewhat weaker constraint of 10 ms delay per hop, the range of the two-hop link even reaches 750 m, visualized by the dark shaded area.

Relay based two-hop communication provides another considerable benefit mentioned in Section 10.4.1.3: It is able to eliminate radio signal shadowing caused by obstacles that obstruct

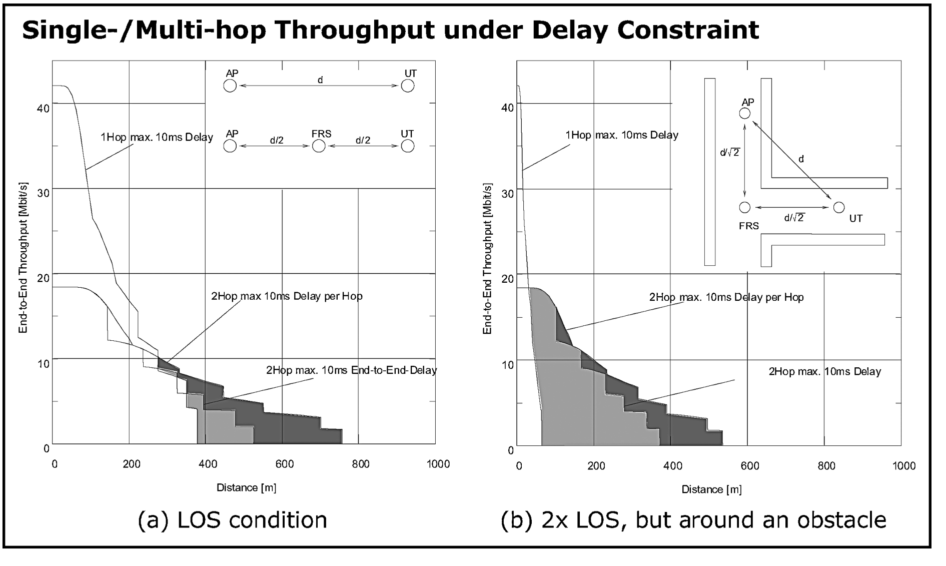


Figure 10.10 Analytical results for maximum achievable end-to-end throughput of a UT versus distance for one- and two-hop links under ARQ error correction with 10 ms maximum end-to-end delay restriction.

the radio path from an AP. An example of this is given for the scenario shown in Figure 10.10(b), see right-hand upper part, where the throughput gain of the two-hop over single-hop link is highlighted by shading (light-shaded: 10 ms end-to-end delay constraint, dark-shaded: 10 ms per-hop delay constraint). In this scenario, AP and UT are shadowed from each other by two walls that form a rectangular corner, e.g., a street corner. A modified COST259 propagation model (Correia, 2001) is assumed with walls having an attenuation of 12 dB each. The distance is running on the direct way between AP and UT. The shaded area highlights that the two-hop link is superior to single-hop, starting at a distance of about 50 m and extends the range of the AP up to about 380 m. Both examples shown in Figure 10.10 clearly support the statements made in Section 10.4.1, indicating that relaying may be of advantage for both increasing the throughput close to the cell border of an AP and for providing radio coverage to areas otherwise shadowed from the AP.

10.5 Conclusions

In this chapter, deployment concepts using FRS have been shown to be highly beneficial for extending the service range of an AP, thereby reducing the number of APs required in a given area and the cost of connecting APs to the wired network. Relays have been demonstrated to be able to substantially extend the radio coverage of an AP, especially in highly obstructed service areas.

The deep impact of relay-based deployment concepts is also identified by 802.16 WG: As introduced in Section 8.1.4.2, TG 802.16j has taken up the ideas presented in this chapter. The MAC protocol of 802.16 operating under central control is well suited to apply the principles of cellular networks, especially when using TDD mode, but also suitable to FDD mode. These principles help to increase network capacity and improve the spectral efficiency.

Note

1. Reproduced by permission of © 2006 IEEE. Source: B. H. Walke, H. Wijaya, and D. C. Schultz (2006) “Layer-2 Relays in Cellular Mobile Radio Networks,” in Proc. of IEEE 63rd Vehicular Technology Conference, VTC 2006-Spring, Melbourne, Australia, 7–10 May 2006.

11

Mutual Integration and Cooperation of Radio Access Networks

Matthias Siebert and Bernhard H. Walke

It is widely recognized that next-generation wireless systems will need to feature incorporative aspects. The reasons for this are manifold.

First, the perception that dedicated systems are more suitable to provide specific services to the user will result in complementary system design. Even today, different systems for ultra short-range, short-distance and wide-area communication exist. This leads to specific deployments of personal body networks, wireless local area networks and cellular networks.

Second, cooperation of systems in the same or adjacent channels will further increase. Due to spectrum scarceness, deregulation efforts for increased market competition and political reasons, the number of systems operating in direct impact with each other will further increase. To allow for proper incorporation, 'live and let live' doctrines are necessary. Research in this field includes listen before talk approaches, frequency-sharing rules (Hettich *et al.*, 1997) for dynamic and static sharing and game theoretical approaches as discussed in Section 13.5.4.

Third, the migration from current 2.5/3G systems to systems beyond 3G will proceed stepwise. This means that future systems need to entail a certain degree of backwards compatibility. Coexistence with antecessor generations thus is an essential requirement. This is also indicated in Figure 11.1. Although next-generation mobile networks pave their way, coexistence with previous generations will require cooperation. Multiband terminals and adaptive multiband terminals therefore will serve as enablers for complementary network usage.

For all the above-mentioned reasons, system cooperation and integration is of fundamental interest. This is also expressed by ongoing research.

Although systems generally are conceptualized in a complementary manner, this does not mean that there are no overlapping domains. The term 'overlapping' addresses different areas,

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

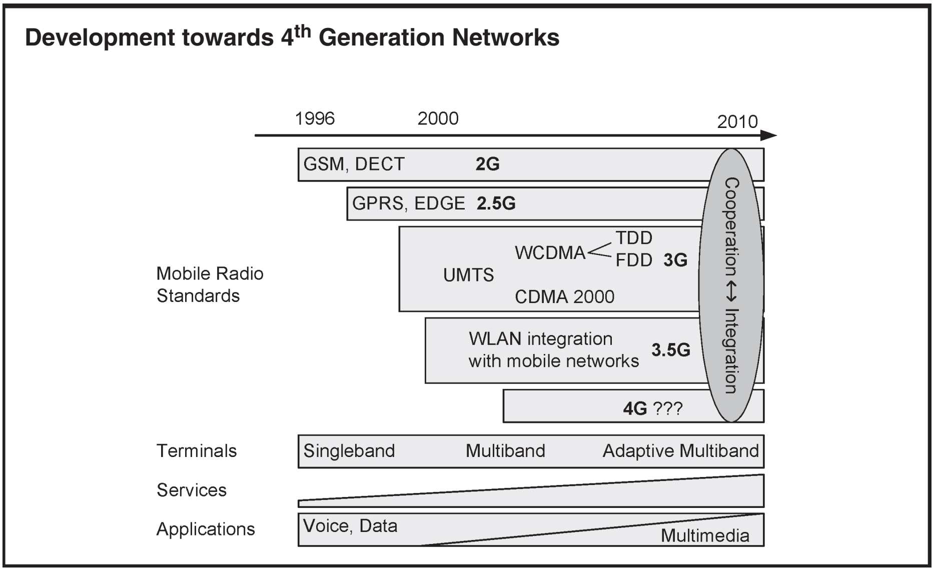


Figure 11.1 Development towards fourth-generation networks.

in particular coverage (space domain) and QoS support (service domain). As such, overlapping network properties are the elementary foundation based on which the dogma of Always Best Connected becomes feasible.

11.1 State-of-the-Art Overview

The following subsections reflect the importance of system cooperation by presenting a comprehensive overview of related standardization forums and research projects with respect to mutual system support. The description addresses ongoing research and upcoming activities, which are at an initial state of development. All these activities emphasize the worldwide necessity for further research into the field of system cooperation.

11.1.1 ETSI BRAN/3GPP

In its first technical report on "Requirements and Architectures for Interworking between High Performance Local Area Network Type 2 (H/2) and 3rd Generation Cellular Systems" (ETSI, 2001a), two approaches, loose coupling and tight coupling, were taken in ETSI BRAN for the provision of system integration between H/2 and the Universal Mobile Telecommunications System (UMTS) depending on the requirements and the feasibility of deployment. Figure 11.2 depicts different levels of Wireless Local Area Network (WLAN) coupling with UMTS.

The *loose* coupling approach is simple to implement without major modifications to the systems. It allows centralized authentication and signaling information related to a user, independent of the radio access network. However, it does not allow for seamless handover between the systems since the local Internet Protocol (IP) address needs to be changed and the QoS for each connection has to be renegotiated.

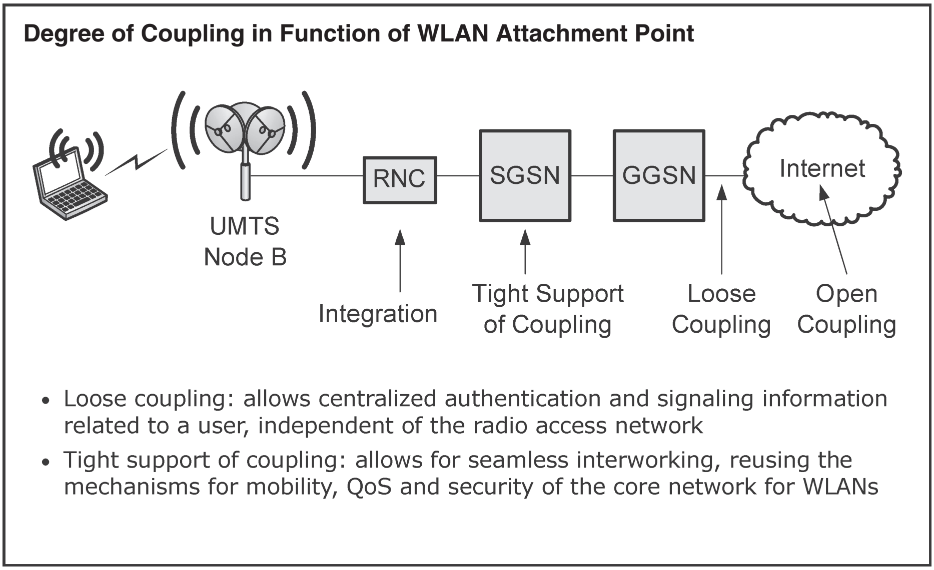


Figure 11.2 Degree of coupling in function of WLAN attachment point (Samarasinghe *et al.*, 2003).

The *tight* approach, on the other hand, allows for seamless interworking, reusing the mechanisms for support of mobility, QoS and security of the UMTS core network for H/2. Furthermore, certain addresses and identifiers of UMTS are used by H/2. However, this increases the system complexity and signaling. The standardization work carried out in ETSI BRAN/3GPP provides a good overview of the variety of approaches that can be adopted, but several aspects in mobility and handover are still open.

Based on the results of ETSI (2001a), ETSI BRAN originally started further work on a Technical Specification TS 101 961, to specify the architectures and protocols of an H/2 network that interworks with 3G networks. However, work on this item was stopped in 2003.

Similar to the ETSI BRAN standardizations efforts and to a certain extent based on it, 3GPP in TSG SA1 (Services) considers coupling for WLAN/3GPP interworking. Rather than distinguishing between loose and tight coupling only, 3GPP considers different levels of interworking in its technical report on "Feasibility Study on UMTS-WLAN Interworking" (3GPP, 2003). In addition, the considered WLAN technology is no longer restricted to H/2. IEEE 802.11a,b as well as Bluetooth have been added as candidate WLAN technologies.

The study covers service use cases in the scenarios as well as the impact on the specification on the standard that will be required for the inclusion of the specific type of coupling and level of interworking. The investigated scenarios consider systematic increase of network integration, beginning with simple 3G-WLAN interworking using common billing and customer care for seamless intersystem operation to allowing access to services provided by entities of the 3GPP Circuit Switched (CS) core network over WLAN.

The six scenarios that were investigated comprise:

- **Common billing and customer care**: This scenario is the simplest interworking scheme with association between systems by a common customer relationship i.e., single bill from the mobile operator and an integrated customer care.

- **3GPP system-based access control and charging**: Here, authentication, authorization and accounting are provided by the 3GPP system. The security level of these functions applied to WLAN is in line with that of the 3GPP system, providing the means for the operator to charge access in a consistent manner over both platforms.
- **Access to 3GPP system packet-switched services**: This is to allow the operator to extend 3GPP system packet-switched services to the WLAN, allowing services such as location-based services, Short Message Service (SMS), Multimedia Messaging Service (MMS), etc.
- **Service continuity**: This scenario allows services to be maintained after vertical handover between the WLAN and a 3G system. Here, any change of access may be noticeable to the user.
- **Seamless services**: This scenario provides seamless service continuity minimizing data loss and break time during the handover.
- **Access to 3GPP circuit-switched services**: This scenario allows access to services provided by the entities of the 3GPP circuit-switched core network over WLAN, providing seamless and user-transparent handover.

These different scenarios provide a path for evolution from entry-level services (loose coupling) to full availability of 3GPP services (very tight coupling) to the WLAN access user. With a stepwise increasing level of service, a deployment initially directed at one service scenario provides the basis for further development. The main characteristics and capabilities of each scenario are summarized in Figure 11.3.

Due to increased interworking requirements, further specifications were defined within 3GPP. Requirements for 3GPP systems to WLAN interworking have been specified in 3GPP (2005d). A "System description" and a "Functional and architectural definition" are provided in 3GPP (2002, 2005b), respectively. Additional important aspects such as telecommunication and charging management (3GPP, 2005c) as well as interworking security aspects (3GPP, 2005f) complement the 3GPP–WLAN system integration framework.

Handover capabilities for combinational services between Wideband Code Division Multiple Access (WCDMA) and WLAN networks are currently being considered for inclusion in 3GPP Release 8.

11.1.2 IEEE

Within IEEE, there are different standardization activities each dealing with a specific topic. While in the beginning only aspects of homogeneous system management have been addressed, a couple of new groups have been founded in the last 2–3 years covering interoperability and integration aspects. The most important ones are presented in the following.

11.1.2.1 IEEE 802.11u: Interworking with External Networks

IEEE 802.11u was formerly known as the Wireless Interworking with External Networks (WIEN) Study Group. Approval for a full task group was given in December 2004. By providing amendments to the IEEE 802.11 Physical (PHY)/Medium Access Control (MAC) layers, interworking with other (external) networks shall be enabled. Similar to external bodies such as 3GPP that define interworking capabilities from the cellular perspective, IEEE 802.11u shall specify logical counterparts from 802.11 point's of view, which can only be addressed within the IEEE 802.11 project. This shall allow external networks to interwork with IEEE 802.11 equipment in a common, harmonized and standardized manner. Two main areas will be addressed: enhanced protocol exchanges across the air interface and provision of primitives to support required interactions with higher layers for interworking. This includes specific interfaces to support external authentication, authorization and accounting, together with network selection, encryption, policy enforcement and resource management. In addition, interfaces to existing IEEE 802.11 functions, e.g., 802.11i, shall be considered.

There is another project within IEEE 802 with similar scope as 802.11u. The respective project is referred to as IEEE 802.21 and will be introduced in the next section. In order to avoid overlaps in the scopes of the two projects, an agreement for ongoing formal coordination has been made between IEEE 802.21 and IEEE 802.11 WIEN SG, the predecessor of IEEE 802.11u.

11.1.2.2 802.21 Media Independent Handoff Working Group

IEEE 802.21's aim is to develop standards to enable handover and interoperability between heterogeneous network types including both 802 and non-802 networks. After meeting as an Executive Committee Study Group (ECSG) for one year, 802.21 was approved by the IEEE as a full IEEE WG in February 2004 (IEEE, 2005d).

WLAN/3G Interworking Scenarios defined within 3GPP

Scenarios:	loose	loose	loose	tight	very tight	very tight
Service & operational capabilities:	Scen. 1: Common Billing & Customer Care	Scen. 2: 3GPP system based Access Control & Charging	Scen. 3: Access to 3GPP system PS based services	Scen. 4: Service continuity	Scen. 5: Seamless services	Scen. 6: Access to 3GPP system CS based services
Common billing	X	X	X	X	X	X
Common customer care	X	X	X	X	X	X
3GPP system based Access Control		X	X	X	X	X
3GPP system based Access Charging		X	X	X	X	X
Access to 3GPP system PS based services from WLAN			X	X	X	X
Service Continuity				X	X	X
Seamless Service Continuity					X	X
Access to 3GPP system CS based Services with seamless mobility						X

Figure 11.3 WLAN/3G interworking scenarios defined within 3GPP (3GPP, 2003).

A key role within 802.21 is played by the so-called Link Layer triggers (L2 triggers) (Johnston, 2004), see Section 11.3, since their availability and evaluation is essential for handover optimization. Further on, general handover information definitions and transport are addressed by 802.21, which also incorporates cellular coupling methods. Accordingly, the Project Authorization (PAR) at the IEEE was given for "Media Independent Handover Services". Future tasks of this working group include definitions of useful L2 network detection mechanisms. For this, a close cooperation with the Internet Engineering Task Force (IETF) Detecting Network Attachment (DNA) working group and the IRTF Mobility Optimizations (MobOpts) working group is aimed. Further (in)formal liaison efforts with IEEE 802.16 as well as WGs of 802.11 are on the way.

11.1.3 IETF

Contrary to the previously mentioned IEEE working groups, the focus of IETF in the context of system integration is on Network Layer (L3) connectivity and above. According to the ISO/OSI model, radio connectivity is subject to L2 while L3 should be decoupled from any radio-specific impacts. However, radio propagation related signal fading effects have significant impact on higher layer protocols as known from e.g., Transmission Control Protocol (TCP) when facing wireless interface related problems (Lefevre and Vivier, 2000). This is why dedicated working groups within IETF have taken up the topic of link layer triggers (L2 triggers) to be used by higher layer protocols.

The purpose of the IETF DNA Working Group (IETF, 2005c) is to define standards that allow hosts to detect their IP layer configuration and connectivity status quickly and would allow a host to reconfigure its Internet Protocol Version 6 (IPv6) layer faster than today. Indications currently available from a subset of wireless link layer technologies therefore may be exploited.

System integration in the sense of macro mobility support across (radio) network boundaries is the scope of the IETF Working Groups Mobility for IPv4 (mip4) (IETF, 2005a) and Mobility for IPv6 (mip6) (IETF, 2005b), respectively. IP mobility support, better known as Mobile IP, allows a node to continue using its permanent home address as it moves. The Mobile IP protocols support transparency above the IP layer, including maintenance of active TCP connections and User Datagram Protocol (UDP) port bindings. Therefore, the Mobile IP procedure is also referred to as L3 handover. Besides the basic Mobile IP protocols, several other drafts deal with concerns such as optimization, security, extensions, Authentication, Authorization, Accounting (AAA) support and deployment issues.

11.1.4 ITU-T

Within the Telecommunication Standardization Sector of the International Telecommunication Union (ITU-T), standardization work is carried out by ITU-T Study Groups. The Study Group 13 'Next Generation Networks' (ITU, 2005b) is responsible for studies relating to the architecture, evolution and convergence of future generation networks. Among others, this includes frameworks and functional architectures, signaling requirements and interoperability. Convergence as addressed by the study group implies architecture and service convergence, interoperability between fixed and mobile networks as well as integration of satellite with terrestrial and Next Generation Networks (NGNs).

11.1.5 WWRF

The Wireless World Research Forum (WWRF, 2005) is a global organization, which brings together experts from different domains such as manufacturers, network operators/service providers, Research and Development (R&D) centers, universities as well as small and medium enterprises. Founded in August 2001, WWRF addresses research issues relevant to future mobile and wireless communications, including pre-regulatory impact assessments. Accordingly, WWRF tends to cooperate closely with other fora and standardization bodies such as the UMTS Forum, ETSI, 3GPP, IETF, ITU, and other relevant panels.

Main objectives are formulation of strategic visions on future research directions in the mobile and wireless area; generation, identification and promotion of technical trends for mobile and wireless systems technologies; and enabling of global R&D collaboration. A common vision of future mobile and wireless technologies has been published in the *Book of Visions* (Tafazolli, 2006).

Current research topics with respect to system integration comprise among other system concepts and high-level architectures, requirements on future mobile and wireless systems, cooperative and ambient networks, inter-cell coordination and reconfiguration aspects.

11.2 Mobility and Handover

Mobile radio systems cover two basic human needs: communication and mobility. At first glance, both actions seem to be well defined; however, inherent manifestations are manifold.

Communication addresses any kind of information exchange including circuit/packet switching, unicast/multicast/broadcast and man/man, man/machine, machine/machine dialog. Similar diverse manifestations apply with respect to mobility to be introduced in Section 11.2.1. Handover is the enabling technology applied by mobile radio systems to bring together communication and mobility. Accordingly, handover-related properties are discussed in Section 11.2.2. A special focus will be put on Vertical HandOver (VHO) serving as an enabling scheme for system integration in the scope of this work.

To understand the various mechanisms related to mobility, some main components being part of each communication process need to be introduced first. Figure 11.4 depicts involved domains and interfaces:

- **User:** A physical person in a certain role that employs services by using network resources. One single person can have different user identities in different roles, e.g., private and business.
- **Terminal:** A device that provides access to the network facilities. In addition, it serves as a platform for services and sessions. A built-in user interface allows for input/output of various kinds of information. Using this terminal the user is able to access the network, whereby each information transfer is realized over one or several links.
- **Link:** The communication facilities between corresponding entities. A link can have several manifestations, e.g., wired or wireless, depending on the considered network. It is used as a description for physical (layer 1) or logical connections ($\geq$layer 2) between entities of a network.
- **Point of Access (Attachment):** A device or interface for connecting a terminal to the network. The connection between terminal and network is realized over a Point of Access (PoA) that may change, e.g., due to movement of the user/terminal. Its realization depends on the used technology. It may be a plug for a network cable, as well as a base station of a wireless network.

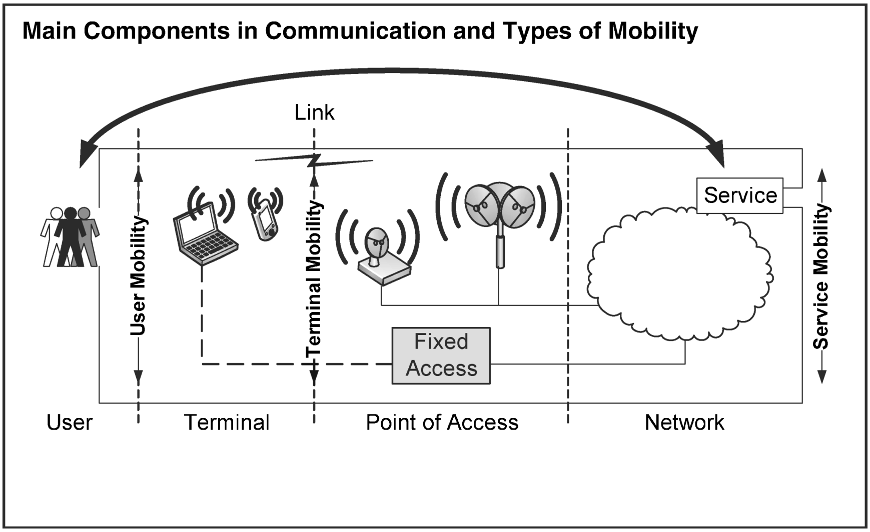


Figure 11.4 Main components in communication and mobility types.

- **Network:** Infrastructure providing connectivity and transport-related services between PoAs, control entities and (service) databases. Uplink data arriving at the PoA has to be delivered to the destination by the network and its facilities.
- **Service:** Type of information and its characteristics. A key driver of mobile communication has been the circuit-switched speech service. Future communication supposedly will be dominated by packet-switched data services with value-added services established on top.
- **Session:** Set of one or more associations between two or more entities. A session is a realization of a certain service.

The above terms are necessary for understanding the different manifestations of mobility as explained in the following.

11.2.1 General Aspects of Mobility

There are many manifestations of mobility provided by different communications systems. Mobility Management (MM) is the substantial functionality to support changes of the PoA. In the following, some main forms of mobility will be presented. The respective definitions are in line with those of fora such as VDE/ITG Task Group 5.2.4 ("Mobility in IP-based Networks") (VDE/ITG, 2005) and ITU-T Study Group 16 ("Multimedia Terminals, Systems, and Applications") (ITU, 2005a):

- **User (personal) mobility**: The ability of a user to maintain the same user identity irrespective of the terminal used and its network point of attachment. Terminals used may be of different types.
- **Service mobility:** The ability of a user to use a particular (subscribed) service irrespective of the location of the user and the terminal that is used for that purpose.
- **Session mobility:** The capability to move an active session between terminals.

- **Terminal (Host) mobility:** The ability of a terminal to change its location, i.e., network point of attachment, and still be able to communicate.

User mobility refers to a related group of functions, which enable a user to obtain access to services provided via a network, independent of which terminal the user is currently working on. Typically, authentication and authorization procedures need to be applied first. An example of user mobility is use of different mobile phones for which users may carry the same identity from one physical device to another. A Personalized Subscriber Identity Module (SIM) as a smartcard applies to map person and identity.

Service mobility is closely related to user mobility. Ideally, a user changes the device used to access services and will face exactly the same service on another device. As an example, users may access email from their home provider via a web interface or using a mobile phone with Wireless Application Protocol (WAP) capability. However, if there is a large multimedia attachment to an email, it might be impossible to access it by the mobile phone. In this case, user mobility from the web interface to the mobile is supported, but service mobility is restricted.

Session mobility is also related to user mobility in the sense that the user initially starts a session on one device and wants to continue this session on another one. For instance, a user starts a VoIP call from his PC connected by DSL wireline but then decides to leave home and to handover the session to a VoIP-capable mobile device with, e.g., wireless UMTS connection.

Terminal mobility is always related to physical movement. It refers to the function of allowing a user's terminal to change its PoA without interrupting service to the user(s) on that terminal. Note that this does not necessarily mean seamless service continuity. Terminal mobility is logically independent of user mobility, although in real networks at least the address management functions are often required to attach the terminal to the network after switch on.

Beside the mentioned mobility aspects, some further important characteristics will be introduced, namely:

- **Seamless mobility:** The user/terminal is able to change the network access point, as he/it moves, without interrupting the current service session, i.e., handovers are possible.
- **Nomadic mobility** (roaming): The ability of a user/terminal to change the network access point as he/it moves while the service session is completely stopped and started again i.e., discrete switching between locations is done.
- **Continuous mobility:** Movement over a wide area without disconnecting from the network, e.g., 2/3G.
- **Discrete mobility:** Connectivity is provided only within certain areas, e.g., WLAN hotspots.

11.2.2 Handover Aspects

Within literature, many different aspects of handover* (HO) are described and consequently, various terminologies and definitions can be found. Depending on architecture, procedure, velocity, initiating party and others, most of them are reasonable and appropriate. The following section takes up the main objectives of HO and explains respective properties in detail where necessary.

* In the literature, the terms *handover* and *handoff* are used synonymously. Usually, *handoff* is more popular in America while *handover* is the preferred notation in Europe.

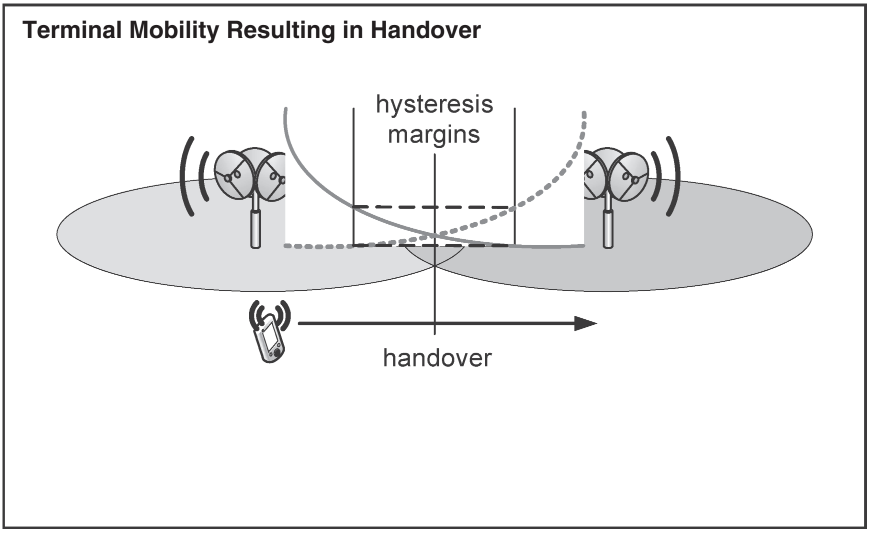


Figure 11.5 Handover of a mobile terminal between two base stations.

11.2.2.1 Definition

In telecommunication, the term handover refers to the process of transferring an ongoing call or data session from one physical resource to another, whereby new parties may be involved and old parties may be released. Accordingly, ETSI and 3GPP define handover as "The transfer of a user's connection from one radio channel to another (can be the same or different cell)" (3GPP, 2005g).

11.2.2.2 Reasons for Handover

Probably the most important reason for handover is degrading signal quality due to physical movement in the context of terminal mobility. The underlying cellular concept adopted by all modern mobile communication systems is based on frequency reuse to enhance spectral efficiency. Transmission power is limited such that only a well-defined area is covered as shown in Figure 11.5. To allow for handover, adjacent cells are designed to overlap each other, constituting areas with reception of signals from two or more base stations. Orthogonal transmission resources in time, frequency or code domain ensure minimal inter-cell interference. By leaving the original coverage area and entering a new cell during ongoing communication, handover is performed to enable (seamless) service mobility.

Including mobility as the main reason for handover, further metrics apply summarized as follows:

- Better cell handover: Due to mobility, the user reaches an area in which signal reception from a neighboring cell is better than from the current one.
- Reference sensitivity level handover: Reception level of the serving cell drops under a specific required minimum, referred to as the required "reference sensitivity level". The reference sensitivity is the minimum receiver input power measured at the antenna port at which the Bit Error Ratio (BER) does not exceed a specific value (usually required BER $< 10^{-3}$). Note that

different systems apply different transmission technologies for which respective thresholds may not be comparable:

  - For different types of Global System for Mobile Communications (GSM) transceivers this limit was defined in 3GPP (2005a), e.g., handheld GSM mobiles require a minimum reception level of $-102\,\text{dBm}$.
  - UMTS UEs as specified in 3GPP (2005e) require $-106.7\,\text{dBm}$ (Common Control Physical Channel, etc.) and $-117.0\,\text{dBm}$ (Dedicated Physical Channel for Frequency Division Duplex).
  - WLAN IEEE 802.11a/g claims a minimum reception level of $-82\,\text{dBm}$ (BPSK 1/2) (IEEE, 2003a).

- Reception quality handover: The reception quality, usually expressed as BER or Packet Error Ratio (PER), drops under a specific required minimum. Usually, BER and PER are directly related to the Carrier-to-Interference (C/I) ratio.
- Traffic reason handover: A handover may be triggered due to capacity and traffic reasons, e.g., a cell is heavily overloaded.
- Speed-based handover: Connections of fast-moving users will be shifted from small (hotspot) cells to large cells to avoid frequent handovers.
- Service-based handover: Dedicated services will preferentially be supported by dedicated cells.

11.2.2.3 Types of Handover

Different HO types are considered according to the PoA before and after handover. The following categories as shown in Figure 11.6 are distinguished:

- **Intra-cell handover**: Switching to another frequency, slot or code within the same supplying cell due to potential quality degradation or capacity reasons. Intra-cell handover comprises also the switching between different sectors of the same cell. Network connections do not need to be altered (3GPP, 2005g).
- **Inter-cell handover:** Handover to a neighbor cell. Contrary to the intra-cell handover, this type requires network connections to be altered. Further distinctions are made with respect to the logical attachment of the two involved cells:
  - Intra-Base Station Controller (BSC)/Radio Network Controller (RNC) handover: Old and new Base Transceiver Station (BTS), respectively Node B are controlled by the same BSC (GERAN), respectively RNC (UTRAN).
  - Inter-BSC/RNC handover: Handover to an adjacent cell being attached to a different BSC (GERAN), respectively RNC (UTRAN).
- **Inter-system handover:** Handover between networks using different radio systems, e.g., between UMTS and WLAN (as shown in Figure 11.6) or UMTS and WLAN. Inter-system handover is the most challenging handover type due to different properties of the involved networks. Another notation for inter-system handover hence is vertical handover to illustrate the heterogeneity of the involved networks. Inter-system handover optimization thus is a big challenge and one of the topics to be addressed in this book. A detailed discussion on vertical handover is given in Section 11.2.2.7.

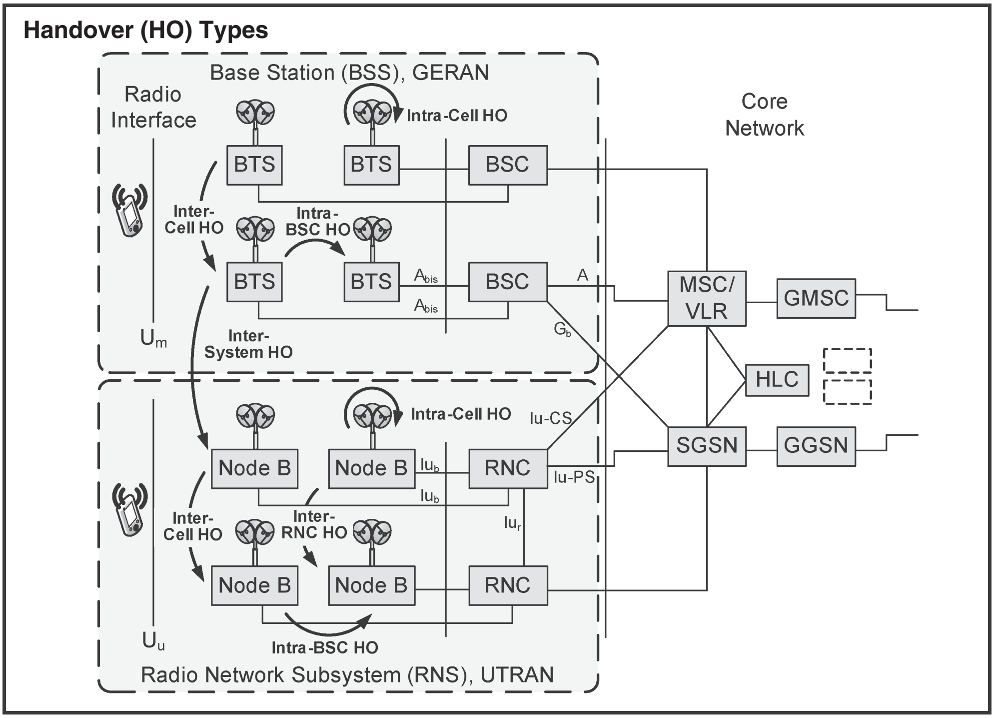


Figure 11.6 Different handover types; exemplary for GERAN and UTRAN.

Smooth, fast, seamless, ideal handover

Moving while being covered by different networks without active connection but still being capable of receiving requests is more related to roaming than to handover. As mentioned, the definition of handover incorporates an active (data) transmission. Smooth handover denotes minimization of packet loss, while fast handover addresses the requirement for minimum packet delays during the actual HO process. If handover is both smooth and fast, such that the time span needed to redirect the link is short enough that the application or the user is (almost) unaware of this process, the handover is referred to as seamless.

Accordingly, 3GPP TR 21.905 defines in its vocabulary for 3GPP Specifications that "a seamless handover is a handover without perceptible interruption of the radio connection" (3GPP, 2005g). While this definition primarily yields to the radio connection, IEEE 802.21 focuses on the end-to-end connection. It is stated in IEEE (2005d), "A seamless handover is defined as one that is either an ideal handover or is a handover with some degradation in service parameters that are mutually acceptable to both the user at a terminal and the network that is providing the service". In contrast to an ideal handover, the given definitions for seamless HO allow little tolerances. Consequently, the requirements for ideal HO are the strictest ones demanding that "An ideal handover across heterogeneous interfaces is where there is no change in service quality, security, and capability as a terminal moves from a source L2 network to a target L2 network" (IEEE, 2005d).

An important issue in all these contexts is that no data will be lost during the handover execution. As long as no path/route update is performed, data to be directed to the mobile is still misrouted to the old serving base station. If the old base station is aware of the new point of attachment of the mobile, seamless handover execution may be supported by tunneling of misrouted data. If tunneling is not supported, all the data will be lost. In spite of dedicated (higher layer) mechanisms to request for retransmission of lost packets, the time span (delay) might be too big to satisfy the challenges of seamlessness.

Another approach to overcome the problems of potentially misrouted and lost packets and to provide seamless HO is to apply soft and semi-soft HO procedures.

Hard handover (break-before-make)

Moving from one cell to another, the handover process requires the connection to be switched between the two cells. If all the old radio links are removed before new radio links are established, the corresponding handover is referred to as hard handover, see Figure 11.7.

This case requires special attention to handover latency since the mobile is effectively disconnected during handover. Fast handover signaling mechanisms (fast re-establishment) are required to provide seamless handover. In practice, a handover that requires a change of the carrier frequency (inter-frequency handover) is always performed as hard handover. The same holds for the inter-mode handover from UMTS Terrestrial Radio Access Network Time Division Duplex (UTRA-TDD) to UMTS Terrestrial Radio Access Network Frequency Division Duplex (UTRA-FDD) and vice versa.

Soft handover (make-before-break)

In 3G Code Division Multiple Access (CDMA) based systems, the user can be connected to several base stations simultaneously, combining the data from all transmitters in range into one signal using a Rake receiver. The set of base stations the terminal is currently connected to is referred to as the active set. Mobile terminals maintaining multiple connections to different cells can perform soft handover. A soft handover happens when there are several base stations in the active set and the terminal drops one of these to add a new one. Accordingly, 3GPP TR

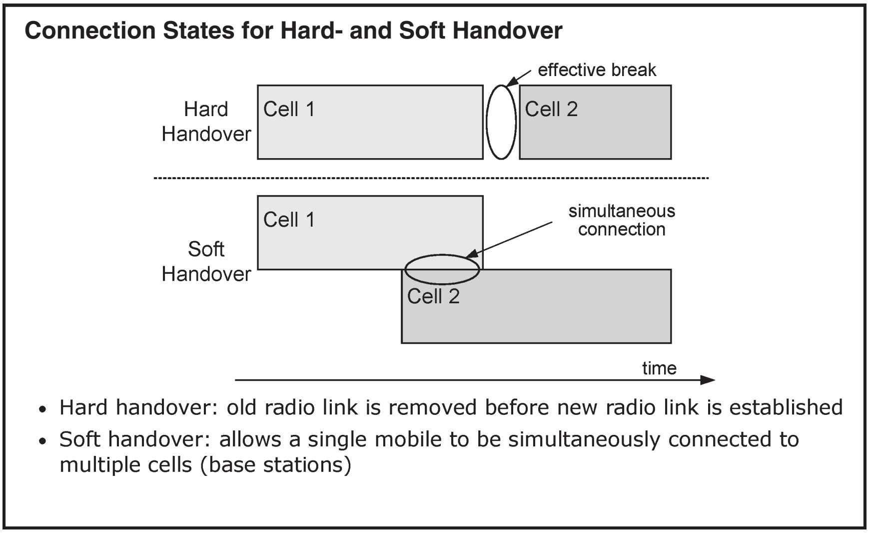


Figure 11.7 Connection states for hard and soft handover.

21.905 defines in its "Vocabulary for 3GPP Specifications" that a "soft handover is a category of handover procedures where the radio links are added and abandoned in such manner that the UE always keeps at least one radio link to the UTRAN" (3GPP, 2005g). Some definitions as used by the IETF Network Working Group (Manner and Kojo, 2004) further distinguish between soft handover and "make-before-break". However, since both terms define related diversity techniques, no further distinction is applied here.

Soft handover, which allows a single mobile to be simultaneously connected to multiple cells, renders the link more robust to fading and thereby enlarges the cell coverage area. The condition that several radio links are active at the same time is also referred to as macro diversity. Normally soft handover can be used when cells operated on the same frequency are changed.

Semi-soft handover is a special case of soft handover. It is used to describe a special micro mobility handover used in UMTS but also IP networks. During handover within IP networks routing information needs to be updated in order to provide connectivity to the mobile host. IP packets are sent to the old base station until the route is updated. Then all incoming packets are sent to the new base station. There can be a short time interval when the mobile host receives packets both from the new and from the old base station. This is called semi-soft handover.

11.2.2.4 Handover Control

The decision mechanism for handover control can be located in a network entity or in the terminal itself. The corresponding cases are called Network-Controlled HandOver (NCHO) and Mobile-Controlled HandOver (MCHO), respectively. In most cases, HO is initiated by the controlling instance. In analogy to NCHO and MCHO, it is distinguished between Network-Initiated HandOver (NIHO) and Mobile-Initiated HandOver (MIHO).

If an HO can be planned and is not due to sudden link deterioration, it is possible to trigger a forward HO, where the old BS processes information for potential new BSs. Forward signaling allows for information pushing from the old to potential new serving BSs by building a temporary tunnel. Another notation is proactive (expected) handover.

If an immediate HO becomes necessary (unplanned, unexpected, reactive), no anticipation is possible beforehand. However, the new BS can notify the old BS by means of backbone signaling in order to request for tunneling of misrouted packets (backward HO).

Performing handover control is a critical task. Considering movement of the terminal as shown in Figure 11.5 ensures that triggering (compare Section 11.3) applies early enough to achieve seamlessness. Due to fluctuations in the received signal strengths from both base stations, it is possible that repeated handover between the same two BSs occur – an effect referred to as ping-pong handover. To overcome this problem, a couple of mechanisms are applied: Hysteresis with respect to the received signal strengths of the two base stations ensures HO triggering only if the new cell's signal is significantly stronger than the current cell's one. In addition, dwell time settings ensure a minimum amount of time for which a call is maintained within a particular cell prior to executing another handover.

Obviously, additional parameters such as the speed of the moving terminal have a significant impact on the decision process for handover control. Due to link state maps as administered by a Hybrid Information System (Siebert *et al.*, 2004) combined with tracking schemes, it is possible to determine and recommend optimal handover locations. Additional benefit arises from the fact that the Hybrid Information System may not only report on existence of complementary radio systems, but also on local characteristics. Together with a (vertical) handover recommendation, the Hybrid Information System may provide additional context information such as system-specific scrambling codes or link adaptation support. Especially hard handovers will benefit by this support, resulting in decreased synchronization and adaptation times and hence, a minimum effective connection's break.

11.2.2.5 Layer 2 Handover

The movement of terminals between PoAs of the same subnet is usually managed by L2 protocols. A subnet is defined as "a logical group of connected network nodes" (Manner and Kojo, 2004) sharing a common network mask (in IPv4) or a network prefix (in IPv6). In practice, all kinds of intra-cell handovers and some kinds of inter-cell handovers are handled exclusively by L2 for which L2 HO is also referred to as "radio handover" or "cellular handover". It is either completely transparent to the routing at the IP layer or it appears simply as a link layer reconfiguration without any mobility implications. Since the entire HO execution can be executed on L2 basis all relevant information about ongoing user connections can be maintained. This includes aspects such as authentication or security parameters making renegotiation obsolete.

Within this chapter, L2 handover is of particular interest. A special challenge thereby derives from the fact that vertical handover involves the incorporation of different air interfaces. Especially synchronization and detection mechanisms play an important role. Investigations as in Montavont and Noël (2002) have shown that L2 handover execution contributes significantly to the overall HO latency. The impact was shown to get even stronger the more users contend for access. Considering the IEEE 802.11 air interface, it is further stated in Montavont and Noël (2003) "the principle overhead is due to L2 properties".

Classical handover execution applies L2 handover followed by L3 handover afterwards. Newer proposals focus on quasi-parallel L2 and L3 handover to reduce the resulting delay. If handover can be anticipated, L3 handover can already be triggered or prepared prior to L2 handover. Concepts as proposed in Lee *et al.* (2005) define pre-warming zones being candidate future APs/BSs for handover. Context transfer by means of backbone signaling hence reduces the overall handover delay to the time span needed for L2 HO execution.

11.2.2.6 Higher Layer Handover

The network layer (L3) is the highest network dependent layer being responsible for setting up, operation and termination of network connections in general, and routing in particular. During a network handover (L3 handover), the terminal associates to a new subnet and a new routing path needs to be established. Further, authentication and security procedures need to be triggered as well as renegotiation of QoS parameters.

Though higher layer handover does only play a minor role in the scope of this chapter, some mechanisms and protocols are briefly mentioned in the following for the sake of completeness. L3 handover in IP-based networks is often handled by Mobile IP (MIP). In order to reduce handover latency, MIP managed macro mobility can be supplemented by micro mobility protocols such as Cellular IP (Valko, 1999) or HAWAII (Ramjee *et al.*, 2002) that provide fast handover processing within smaller domains or subnets transparently to MIP. Mobile IP (for IPv4) (Perkins, 2002) introduces new entities such as Home Agent or Foreign Agent. Standard IP routing is applied to support mobility such that no other changes need to be implemented. A temporary Care-of-Address serves as connection end-point for the terminal. Datagrams arriving at the original home address can be tunneled to the Care-of-Address allowing for transparent connectivity of roaming users. Due to limitations of MIP based on IPv4, enhancements became necessary resulting in MIP based on IPv6 (Deering and Hinden, 1998; Johnson *et al.*, 2004). Terminals apply address auto-configuration to acquire their Care-of-Address mitigating the need for Foreign Agents. A comprehensive description of MIPv4/v6 related topics is given in Perkins (1998).

In addition to L3 handover, further mechanisms exist that could be subsumed as “higher layer handover”. From the ISO/OSI model perspective, handover can be considered as a reconfiguration or “exchange” of specific layers. L2 handover hence is interpreted as a reconfiguration of the physical layer and the data link layer, while L3 handover further affects reconfiguration of the network layer. Ideally, higher layers do not recognize replacement of lower layer functions meaning that the handover process is transparent. From this point of view, session mobility as introduced in Section 11.3.1 can also be interpreted as higher layer handover, i.e., “session handover”. A respective enabler thereby could be the Session Initiation Protocol (SIP) (Rosenberg *et al.*, 2002).

11.2.2.7 Horizontal and Vertical Handover

Performing handover from any source to a target system applying the same technologies and relying on the same specifications is referred to as Horizontal HandOver (HHO). If handover is triggered to another system using a different technology, the notation Vertical HandOver (VHO) applies.

Vertical Handover hence always comprises the involvement of at least two different systems. Accordingly, one refers to this kind of handover as inter-system handover to express that it is a “handover between networks using different radio systems, e.g., UMTS and GSM” (3GPP, 2005g).

According to the supported coverage of involved systems, it is distinguished between upward and downward VHO. Upward VHO denotes the switching from a system with smaller cell sizes and usually higher bandwidth to a wireless overlay with larger cell sizes and usually lower bandwidth per unit area. Downward VHO is the handover in the other direction, respectively. Figure 11.8 depicts both handover types.

Compared to horizontal handover, vertical handover introduces new degrees of freedom. For example, it is possible that a decision unit triggers VHO execution due to QoS aspects, though the actual link quality in the current cell is excellent. If another vertical system with a multiple of offered data rate is available, the decision space is no longer restricted to sole link parameters.

Beyond this background, origins for handover may be further distinguished into radio-related handovers and service-related handovers. While better cell handover, sensitivity level handover and reception quality handover are radio-related handover reasons, traffic reason handover, speed-based handover and service-based handover are service-related handovers. As a rough classification, horizontal handovers fall mainly in the category of radio-related handovers, while vertical handovers subsume service-related handovers. For more information on HHO and VHO, please refer to Pahlavan *et al.* (2000) and Stemm and Katz (1999).

It is worth noting that one needs to distinguish between horizontal/vertical and L2/L3 handover. While the first couple refers to involved technologies, the second couple involves layers and logical affiliations. Due to the capsulated service principle of the ISO/OSI model (ITU, 1994), the IP layer sees network interfaces and IP addresses rather than specific technologies. Thus, horizontal and vertical handovers may or may not be noticed at the IP layer.

Particular vertical handover properties

Since VHO decisions derive from an enhanced decision space, some particular properties are briefly discussed in the following section.

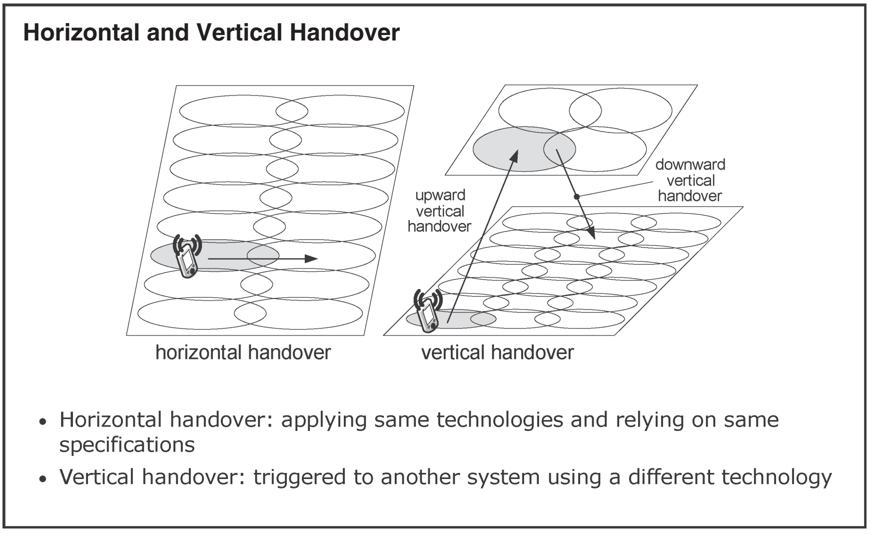


Figure 11.8 Handover from technology perspective.

Time sensitivity

Considering mobility and timing aspects, the downward VHO is much less time critical, since the terminal may long rely on its old connection during the handover process. The upward VHO instead needs to take place in time, otherwise the user may have moved out of the coverage area of its old serving (WLAN) AP before the procedure has been completed. For specific requirements, e.g., throughput optimization, downward VHO turns out to be time critical as well.

Transparency

Signal reception quality is likely to change due to various reasons. Mobile radio systems encounter this problem by applying sophisticated mechanisms such as power control, link adaptation and handover. All of these measures have in common that dedicated boundaries, e.g., minimum throughput requirements, are to be kept. The overall aim is to provide transparent service provision to higher layers in terms of e.g., constant supported bit rate. However, unlike for horizontal handover, vertical handover comprises switching between heterogeneous systems with completely different framework conditions. As such, the immense discrepancy e.g., with respect to supported bandwidth can no longer be balanced by lower layer adaptation mechanisms. In addition, it is not possible to guarantee the same levels of QoS across different systems, for which renegotiation becomes necessary. All this results in decreasing transparency for VHO compared to HHO.

Decision space

Since horizontal handovers fall mainly in the category of radio-related handovers, it is comparatively easy to define metrics for handover decision. Respective triggers are presented in Section 11.3. Vertical handover, however, introduces additional degrees of choice, such as QoS, cost or service availability. Being of a very different nature, objective comparison is difficult. Algorithms applied for HO decision hence need to apply ratings of input parameters. User-defined policies will play an important role in this.

Business aspects

Besides technical innovations, application of heterogeneous transmission technologies requires new business models, too. New cost structures in terms of Capital Expenditures (CAPEX) for investments and Operational Expenditures (OPEX) for operation, administration and maintenance need to be considered. Billing among several involved systems and operators becomes a rather complex task. Classical time- or volume-based billing schemes cannot easily be overtaken, for which questions like the following ones need to be answered (Tirla *et al.*, 2002):

- Where to meter the different traffic measures?
- How to accumulate them in a technology independent way?
- How to combine the costs caused in the networks of the different network technologies?
- How to achieve one bill when different network operators contribute?
- In which of the domains will this combination be achieved?

11.3 Trigger

The widespread need for link layer information to upper layers is clearly expressed by a number of (recently) established standardization groups, especially within IETF and IEEE. The key driver here is the intention to speed up (vertical) handover execution since link layer events are expected to anticipate user (mobile terminal) movement and to prepare the mobile terminal and network in advance. Existing L3 protocols such as Fast MIPv6 (Koodli, 2003) have been designed to incorporate link layer notifications.

11.3.1 Definition and Classification

Whenever handover decisions are to be taken, respective algorithms need to rely on particular data referred to as triggers. Depending on the origin and destination of triggers, the following distinction is herewith introduced: Triggers having been fired by a logical peer entity (different unit, same layer) usually serve for remote information provision. As such, context transfer is intended. Due to the logical information flow between two entities of the same layer, these kinds of triggers are labeled as horizontal triggers.

In contrast to this, notification messages from lower to higher layers are labeled as vertical triggers. In fact, many specifications inherently refer to vertical triggers when addressing the issue of triggering. Vertical triggers usually are unidirectional, being fired from the lower to the higher of two adjacent layers. However, some information flows may also skip intermediate layers when reporting changes. An example of this is the information flow between PHY (layer 1) and Radio Resource Control (RRC) (layer 3) in UMTS.

In general, (vertical) triggers are pieces of information that indicate changes of setup or surrounding conditions. A commonly used definition within IETF says that "An L2 trigger is an abstraction of a notification from L2 (potentially including parameter information) that a certain event has happened or is about to happen" (Perkins, 2002). Many discussions have taken place to define a commonly agreed picture of triggers, especially within IETF (TRIGTRAN, SEAMOBY, PILC, ALIAS), IRTF (MOBOPTS) and IEEE (802.21 Media Independent Handoff Working Group).

Particular benefit with respect to handover latency is given if L2 trigger allows for initiation of L3 handover before L2 handover has completed, e.g., movement anticipation for Fast Handover Protocol for MIPv6 is based on L2 triggers. While the question of the delivery mechanism, e.g., via a standardized Application Programming Interface (API), remote procedure call, or others, is commonly seen secondary, the question of what information should be passed between L2 and L3 is intensively discussed. The predominant opinion is that L2 triggers should turn out to be generic and abstract, and not specific to any particular link layer. They should rather represent generalizations of link layer information available from a wide variety of link layer protocols (Manner and Kojo, 2004). Proposals for this as favored within IETF and IEEE are Link_up, Link_down, Link_going_up, Link_going_down, Link_Quality_Crosses_Threshold and others.

Triggers subsequently only indicate ongoing changes. Specific radio characteristics herewith are hidden to higher layers providing an ISO/OSI conform and transparent operation while simultaneously allowing for anticipation of higher layers.

11.3.2 Decision Criteria

To allow for generation of triggers, L2 is dependent on measurement reports provided by the physical layer. Within the IST project MIND (MIND, 2005), the term physical layer based triggers was introduced to refer to L2 triggers generated as response to L1 reports. In addition, MIND defined algorithm-based triggers to refer to L2 triggers generated due to algorithms running in the data link layer, see table depicted in Figure 11.9.

It has been stated already in the previous section that vertical handover triggering requires special treatment due to the expanded decision space, meaning that it is pointless to compare total values of physical layer based triggers of different systems, see definition of sensitivity level handover in Section 11.2.2.3. Hence, prior to firing a HO recommendation to L3, L2 should implement sophisticated evaluation logic.

L2 Trigger Generation Due to Different Origins

Physical layer based trigger	Algorithm based trigger
Signal Strength (RSSI, RSCP, RPI)	HO Ping-Pong avoidance based on
Interference level	hysteresis
Carrier-to-Interference Ratio (C/I)	QoS violation (e.g., in/-decreasing PER)
Bit Error Ratio (BER) /	Connection Admission Control status
Packet Error Ratio (PER)	Connection Forwarding (CAC & CF)
#HO, #Retransmissions (ACK/NACK)	Location based trigger
# Dropped calls or packets	Velocity based trigger
Delay, HO-latency	A priori-knowledge based trigger
Current window size (ARQ)	Service availability trigger
	Grade of Service

Figure 11.9 L2 trigger generated in response to L1 reports.

Investigations on Radio Resource Management (RRM) principles that are more general have been undertaken by Furuskär (2003). Since the capability to handle services typically differs between systems, the allocation of services affects the overall capacity. As a result, either service allocations should be extremes, i.e., isolated services in different systems, or they should be equal to all systems characterized by the relative efficiency. Anyhow, by also considering radio resource costs, initial attempts have shown that performance improves the more information is available.

12

Future Mesh Technologies

Rui Zhao, Ole Klein, Bernhard H. Walke and Lars Berlemann

IEEE 802.11 infrastructure Basic Service Set (BSS) WLANs are widely used, each comprising an Access Point (AP) and its associated stations. There is a need to interconnect BSSs wirelessly to create an Extended Service Set (ESS) mesh network. This chapter proposes a solution for the architecture and protocols of a Mesh Distributed Coordination Function (MDCF) to interconnect a large number of APs in order to form an efficient ESS mesh network under distributed control. MDCF applies TDMA to share the radio medium and is able to run on a single frequency channel on top of the IEEE 802.11 a/b/g physical layers, concurrently to legacy stations. MDCF is capable to efficiently exploit channel capacity, fairly distribute bandwidth among the Mesh Points (MPs) and support multi-hop relaying of a large number of concurrent traffic flows strictly observing specific QoS requirement. Simulations results prove the outstanding performance of the new concept proposed.

After discussing facts and limitations of existing Medium Access Control (MAC) protocols in Section 12.1, the MDCF is introduced and evaluated in Section 12.2. It can serve as a potential approach to realize meshed networks of 802.11 APs. This chapter is summarized in Section 12.3 with a conclusion.

12.1 Facts on Medium Access Control

In wireless systems the MAC protocol plays a decisive role especially if QoS is under consideration. The resource “radio channel” has to be assigned to all competing stations in a proper way to meet their respective QoS requirements. Therefore the MAC protocol needs to acquire and supervise information about the channel status to be aware of the available resources and the requirements of the stations interested in using the resources. The way the MAC protocol controls the access to the medium significantly influences the level of QoS that the whole system is able to offer.

The radio resource allocation, in general, can be realized by central or decentral control and the relation of the stations to each other may be master–slave or some kind of group of equals.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

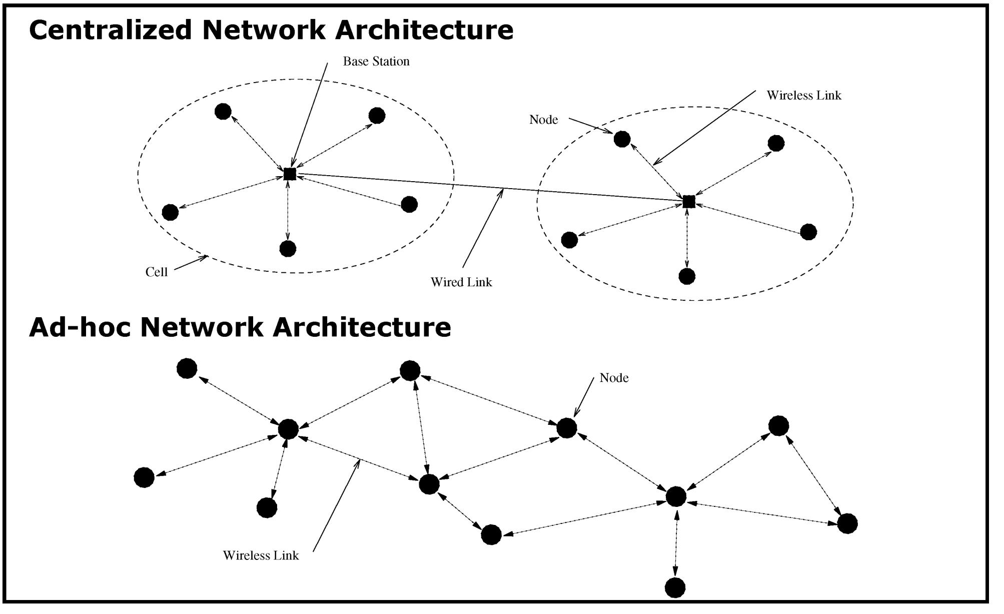


Figure 12.1 Centralized and ad-hoc network architecture.

A centrally controlled (master–slave) network architecture in the majority of cases features a specialized node, i.e., the base station, that coordinates and controls all resource allocations within its coverage area or cell. To increase network coverage, several base stations may be interconnected by land lines that eventually tie into an existing network. Thus, each base station also plays the role of an intermediary between the wired and wireless domains. Figure 12.1 shows a two-cell based network with central control.

Communication from a base station to a node takes place on a downlink channel, and the opposite direction is served on an uplink channel. The base station has full access control to the downlink channel, whereas the nodes share the uplink channels.

In most cases, at least one of these uplink channels is specifically assigned to collect control information from the nodes. The base station grants access to the uplink channels in response to service requests received on the control channel. Thus, the nodes simply follow the instructions of the base station. The concentration of intelligence at the base station leads to a greatly simplified node design that is both compact and energy efficient. The centralized control also simplifies QoS support and bandwidth management since the base station can collect the requirements and prioritize channel access accordingly. The centralized system also presents a single point of failure.

Even if the structure of the network architecture is centralized the coordination and control of resource allocations may be distributed between the nodes in a decentral manner, see Section 12.1.1.2.

An ad-hoc network architecture does not implicitly feature decentral coordination and control of resource allocation. A network built up in an ad-hoc manner does not necessarily lack a central controller solely responsible for resource allocation, see H/2 Home Environment Extension in Section 12.1.1.1.

The main characteristic of an ad-hoc network is the absence of any predefined structure. Service coverage and network connectivity are defined solely by node proximity and the prevailing RF propagation characteristics. Ad-hoc nodes usually communicate directly with one another in a peer-to-peer fashion. To facilitate communication between distant nodes, each ad-hoc node also may act as a router, storing and forwarding packets on behalf of other nodes. The result is a generalized wireless network that can be rapidly deployed and dynamically reconfigured to provide on-demand networking solutions. Figure 12.1 shows a simple completely decentrally organized ad-hoc network. The distributed coordination and control of resource allocation requires a close cooperation between the nodes.

Communication between nodes may be based on a reserved channel, called connection oriented. The transport service will guarantee that all data will be delivered to the other end in the same order as sent. Connection-oriented communication proceeds through three well-defined phases: connection establishment, data transfer, connection release (Walke, 2002).

Support of real-time oriented classes of service will be mandatory in future radio systems. To enable QoS support in packet-based systems the reservation of the medium in advance to the packet transmission is required, at least. Reservation is needed to be able to control interference during transmission as interference might degrade link quality in terms of bit error ratio and might even interrupt an ongoing service. Reservation per packet transmitted is performed in many widely used protocols. In what follows, the MAC protocols of known systems are reviewed with the aim to derive knowledge about their strengths and weaknesses to benefit the design of future protocols ending up in the MDCF described in Section 12.2.

12.1.1 State of the Art in Medium Access Control Protocols – A Taxonomy

The state of the art is represented by systems such as IEEE 802.11, 802.11e, 802.16 and the systems are briefly summarized in the following. For details see the respective Chapters of this book.

12.1.1.1 HiperLAN 2 (H/2)

The H/2 system was designed by ETSI-Broadband Radio Access Networks (BRAN) in 1999 to provide broadband wireless network access in areas of intermediate size, typically served by a WLAN, see Walke (2002). The MAC protocol later was used as a template for specifying the Wireless Metropolitan Area Network (WMAN) standard named ETSI-HiperMAN, which is identical to standard IEEE 802.16. Handover procedures are specified in H/2 and QoS is supported by this WLAN standard. The wireless link is operated connection oriented in that a time slot is reserved by the AP in advance to transmitting user data on both uplink and downlink. There are two ways to control user data exchange: centralized mode and direct mode.

In centralized mode one or several APs provide access to fixed-line networks. Mobile Terminals (MTs) are associated with an AP. If two MTs communicate with each other all their traffic goes through the AP.

Under direct mode (an option in H/2) two MTs that are associated to the same AP and within mutual radio range may exchange user data directly, under control of the AP. An H/2 network must be under central control at any time.

Radio resources to transmit data are scheduled by the AP on basis of a periodic MAC frame for the duration of that frame. To be able to support QoS of applications running on MTs, the AP needs to know the amount and priority of the packets queued for a certain connection at the MT that is communicated to the AP uplink either piggybacked to user data or via a scheduled data slot.

The MAC layer divides the physical channel into frames of constant length, namely 2 ms, which translates into 500 OFDM symbols each frame. The physical layer is more or less the same as used with 802.11a. A TDM/TDMA/TDD transmission scheme is used, very similar to that specified in IEEE 802.16. The specification of H/2 can be found in ETSI (1999).

12.1.1.2 DECT

The ETSI-DECT system, is very similar to the PHS system of Japan, and was designed to support digital cordless telephony, see Walke (2002). Its radio interface is based on the FDM/FDMA/TDM/TDMA/TDD methods. Handover procedures enable the user to roam in the area covered by more than one base station. The MAC protocol is connection oriented, based on TDM channels exclusively allocated to communicating stations and QoS is supported by this. Once a connection is established the negotiated timeslots are kept for the duration of the connection. The most advanced feature is the Dynamic Channel Selection, where mobiles may choose the best channel for their communication after having measured the respective signal strengths. Thus, the whole frequency spectrum, represented by the PHS/DECT frequency channels and the related TDMA channels, is available in each cell. No frequency planning is needed.

The MAC protocol of DECT applies a multi-frame structure of 16 frames, each comprising 24 slots, where a slot represents a TDMA channel. The first half of a frame is used for the downlink, the second half for the uplink. Thereby, a base station can operate 12 full-duplex TDMA connections (32 kbit/s each) to serve 12 mobile stations, concurrently (ETSI, 2004; Walke, 2002). A packet mode has been specified, too, able to multiplex data packets of different mobile stations between base and mobile stations. The DECT standard also comprises a Wireless Base Station (WBS), which can be termed the first standardized fixed wireless router in the world.

12.1.1.3 GPRS

The General Packet Radio Service (GPRS) is designed to multiplex the packet transmission of different concurrent mobile stations over a TDMA channel of the connection-oriented GSM system. GSM is based on FDM/FDMA/TDM/TDMA/FDD transmission. One frequency channel offers eight TDMA traffic channels, simultaneously. A number of traffic channels may be combined in the multi-slot option to provide a higher capacity TDMA channel for packet multiplexing.

The multiplexing on the DL is done by the downlink scheduler in the base station that knows all active MSs and the pending DL packets. The uplink is shared between several MSs transmitting to the BS. Access to the uplink is realized with a slotted ALOHA based reservation protocol. A Packet Channel Request sent on the Packet Random Access Channel to the base station is answered by a Packet Uplink Assignment message to the mobile indicating the time position of the radio resources reserved for the MS (Walke, 2002).

12.1.2 Potentials and Limitations of the State-of-the-art MAC Protocols

The MAC protocols mentioned can be categorized into three types where the protocols grouped to one type have in common the way of how to reserve the medium.

A **type 1** MAC protocol performs radio resource reservation on a per packet basis. The legacy IEEE 802.11 protocol is an example of this in that it either “reserves” the medium by spontaneously transmitting, or applies a reservation cycle RTS/CTS to reserve the medium in advance to user packet data transmission.

A **type 2** protocol uses a periodic MAC-frame to organize access to the radio medium by reservation, applying “TDMA in the short”. The group of type 2 protocols comprises the following standard systems:

- ETSI HiperLAN 2
- IEEE 802.11e for QoS support in WLANs

- IEEE 802.15.3 for Ultra-Wide Band communication on OFDM basis
- IEEE 802.16a (HiperMAN) WMAN for Point-to-Multipoint communication to connect fixed mobile stations to the base station.

A **type 3** MAC protocol is represented by standards such as GSM/GPRS or DECT. These perform the reservation of the medium applying "TDMA in the long" in that a TDMA channel is established for multiplexing packets of different stations for transmission from/to the base station (in DECT this only applies to the WBS). This way of resource allocation is especially well suited to predict the future use of a channel in terms of interference produced in a system, since the reservation of the channel is for quite a long duration, thereby it is easier to take into account by other potential users than is possible with type 1 and type 2 MAC protocols.

The most crucial problem in wireless systems operating under decentral control is to provide QoS support, which is easy to achieve if prediction of use of radio resources can safely be made from the current occupancy of radio resources. Type 3 protocols applying TDMA in the long are much better suited for this than TDMA in the short, as explained in the next section when introducing the MDCF.

In the following, limitations and potentials of the different types of MAC protocols introduced are reviewed with reference to the respective state-of-the-art systems using these protocols.

12.1.2.1 Reservation per Packet

The most simple architecture for a MAC protocol is implemented in standard IEEE 802.11 (legacy mode). There, the reservation of the medium is performed on a per packet basis. Each station listens to the carrier and if the medium is detected idle for a defined time duration (IFS), the station starts its transmission. The major drawback of this approach is that, the more mobile stations want to transmit, the higher the probability of collisions of transmitted packets. This is illustrated in Figure 12.2 where the probabilities for collision (solid line) and of access (dashed line) of stations are shown vs. the number of stations present in a network. The access probability decreases with increased number of stations since more and more stations are in backoff with large contention-window size.

One drawback is that under high network load resulting from many competing stations, the data rate per station and the total network throughput is reduced dramatically for any channel data rate (equivalent to a PHY mode), as can be seen in Figure 12.2. Apparently, the throughput drops for each channel data rate and packet size to less than half when increasing the number of stations from 1 to 100.

Another drawback of a random access based medium reservation on packet basis is that QoS cannot be supported. In order to support QoS a MAC protocol must be able to guarantee the transmission of a certain packet within a maximum delay time. This is not possible, in general, if a station has to contend for each packet transmission anew, since there is no guarantee that a station will win a contention within a given time duration. Figure 12.3 shows the Complementary Cumulative Distribution Function (CDF) of the packet service time (delay plus transmission time) for a given packet size and a PHY mode of 24 Mbit/s. It can be seen that with increasing number of stations the probability substantially increases that a packet service time will be 0.2 seconds or more.

Type 1 protocols in addition suffer from hidden stations as explained in Section 4.4, a problem that is partly solved when applying the RTS/CTS cycle specified in standard 802.11.

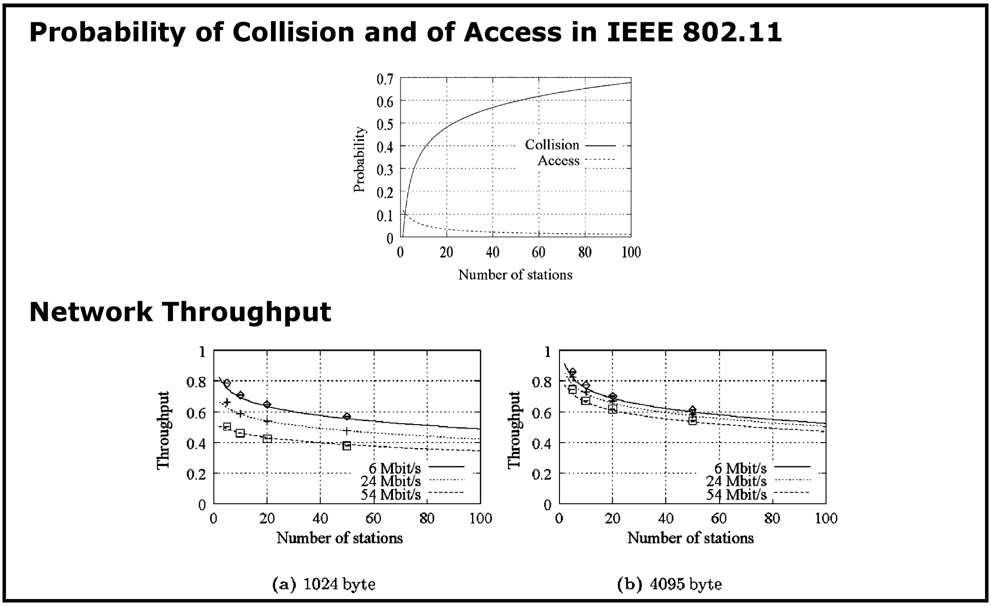


Figure 12.2 Probability of collision and of access versus number of stations in IEEE 802.11 (Kleinrock and Lam, 1975). Network throughput for packet sizes of 1024 and 4095 byte (Walke *et al.*, 2001).

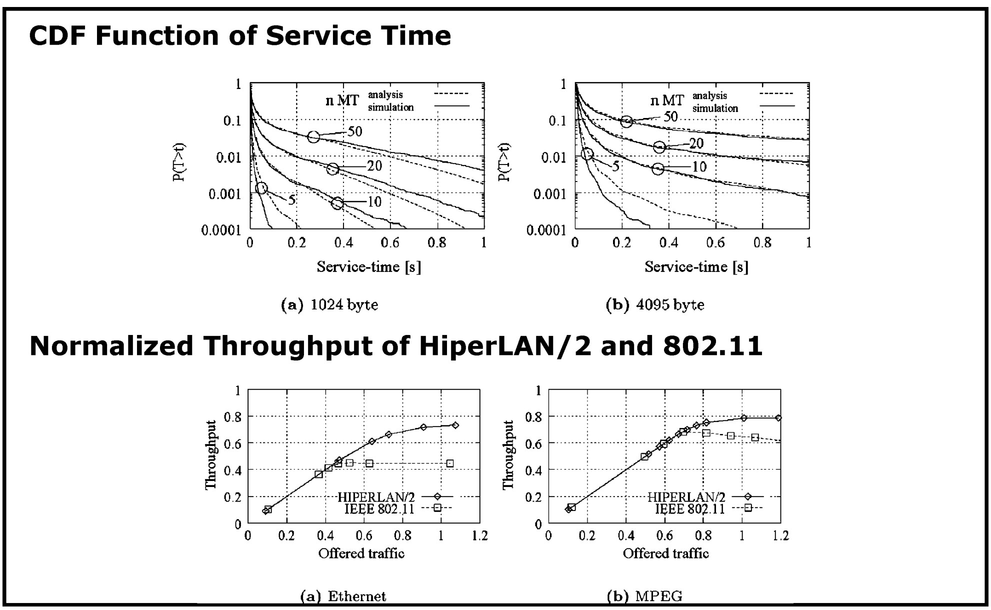


Figure 12.3 Complementary Cumulative Distribution Function of the service time for n stations. Packet size is 1024 or 4095 byte and data rate is 24 Mbit/s (Walke *et al.*, 2001). Normalized throughput of HiperLAN/2 and 802.11 WLAN for Ethernet and MPEG traffic.

12.1.2.2 TDMA in the Short

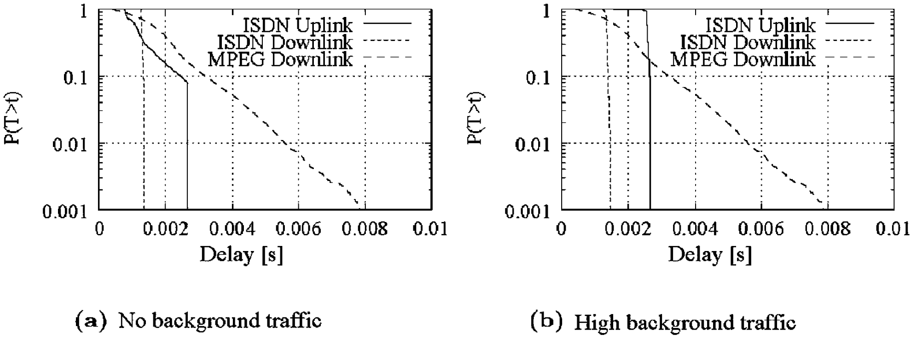
Type 2 protocols apply "TDMA in the short" whereby the medium is reserved on a per frame basis for a number of packet transmissions per frame to and from, typically, different mobile stations in the range of the AP. The amount of overhead spent per frame is reduced compared to a reservation per packet. Accordingly, compared to legacy 802.11 systems, type 2 protocols achieve a much higher throughput as shown in Figure 12.3. The normalized throughput is plotted vs. the offered traffic for both H/2 and IEEE 802.11 systems, which can be compared fairly, since both use the same physical layer. The results are gained by simulation in a homogeneous scenario with 10 stations active. The PHY data rate is 24 Mbit/s for the Ethernet-like load scenario and 27 Mbit/s for the MPEG load scenario (Walke *et al.*, 2001). The type 2 (H/2) protocol can carry roughly 60 % more traffic under Ethernet load than 802.11 and 10 % more traffic under MPEG traffic load.

Type 2 protocols owing to their ability to differentiate service classes by scheduling are able to guarantee a maximum delay for real-time services as shown in Figure 12.4. The CDF of packet delay is shown for a scenario where three traffic classes compete, namely one MT operating ISDN voice traffic, one MT is receiving MPEG traffic from the AP, and eight MTs exchanging Ethernet-like traffic with each other via Direct Link. It can be seen that ISDN and MPEG traffic are served with the same delay without and with background traffic.

To be able to support QoS, a MAC frame structure was added to IEEE 802.11 through the amendment 802.11e, making it a type 2 protocol 802.11e and greatly improving its performance, compared to the legacy version 802.11 (IEEE, 2003a, 2005b).

Different to 802.11 systems, which are capable of working either with decentralized or semi-centralized control, type 2 protocols operate under central control by the local AP. Ad-hoc

CDF of Packet Delay in Multimedia Scenario



- A type 2 MAC protocol can differentiate service classes according to its priority
- H/2, 802.16 and 802.11e (HCCA) apply central control and are comparable in performance

Figure 12.4 Complementary Cumulative Distribution Function of packet delay for a multimedia scenario. (a) No background traffic, (b) much Ethernet-like background traffic (Walke *et al.*, 2001).

capabilities as well as dynamic channel allocation are not easy to incorporate into type 2 protocols. It has been shown, however, that multi-hop operation and ad-hoc networking can be provided without losing the property to be able to guarantee QoS by type 2 protocols, e.g., by means of the H/2 Home Environment Extension (Habetha and Walke, 2002).

12.1.2.3 TDMA in the Long

Type 3 MAC protocols using "TDMA in the long" are represented by GSM/GPRS and DECT-like systems. Fixed TDMA channels are used to multiplex packets of different stations and thereby keep the duration of the "connection" much longer than possible for only one packet. With GPRS the establishment of a TDMA channel is under central control of the base, with DECT the mobile station may suggest a frequency channel and TDMA channel about where to establish a connection under non-central control.

A multi-hop capability is not specified in any of the state-of-the-art systems discussed, besides WBS in the DECT system. It will be shown in the example of the MDCF below in Section 12.2 that TDMA in the long is the key for superiority of non-centrally controlled multi-hop mesh networks compared to other designs, owing to its ability to extrapolate from the current occupation of a radio channel (derived from measurements) for predicting the interference expected in the future.

12.1.3 Key Methods for QoS Supporting Medium Access Control Protocols

12.1.3.1 Single-hop Links

This section relies on the findings from above and describes key methods to be taken into account when designing a MAC protocol for an advanced wireless multi-hop mesh network that is expected to support QoS sensitive applications.

When designing a MAC protocol for a packet based system, the following facts should be considered:

- The protocol should be able to reserve the radio medium (establish a connection) when needed for packet transmission.
- The time to set up a connection, *Tc*, should be small in relation to the duration, *Td*, the connection is being used; otherwise the overhead *Tc/Td* might be unacceptably high.
- If a station is multiplexing traffic of multiple applications to a connection used to realize a one-hop link, the duration *Td* of a connection will tend to increase compared to a connection used per application.
- If a station is not purely serving its own application but some of its traffic is relayed for other stations by multiplexing relay packets to a reserved one-hop connection, this is welcome to extend the duration *Td* that a connection is operated.
- Reservation of some resources of a radio medium can best be taken into account by all other stations competing for resources, if the occupancy duration *Td* is lasting for a long time.

Taking these facts into account it is possible to explain the limitations of the different types of MAC protocols introduced above:

- **Type 1** protocols reserving the medium on per packet basis are overhead prone.
- **Type 2** protocols applying "TDMA in the short" reservation for multiple transmissions at the same time (using a MAC frame) cause much lower overhead compared to type 1 systems. The reason is that the reservation of the medium is performed in the broadcast phase of a MAC frame for a number of packet transmissions of different radio links at the same time.
- **Type 3** protocols applying "TDMA in the long" by channel-based reservation of the medium for one hop might benefit from the resulting reduced *Tc/Td* ratio, if the number of packets multiplexed to a TDMA channel is large and the connection duration *Td* is long. Overhead applies only for connection establishment and not per packet, since no reservation overhead is needed per packet at all. This gives a guideline on how to design future MAC protocols for wireless multi-hop networks.

Regarding a technology evolution based increase of the transmission rate of wireless media and a respective reduction of packet lengths of a given application, it is important considering the impact on the different MAC protocol types (Walke, 2002):

- **Type 1** protocols with a reservation per packet will suffer more from a higher link bit rate than type 2 and type 3 protocols. Overhead percentage *Tc/Td* will be maximum for type 1 protocols: 802.11 is known to poorly perform in terms of capacity for short packets at current link speeds.
- **Type 2** protocols under higher link bit rate will benefit from sharing part of the reservation overhead for all the packets transmitted, since under a higher link data rate, more user packets will fit into a fixed-size MAC frame. Since there is also some individual overhead per packet, the total overhead will grow, but much less than with type 1 protocols.
- **Type 3** protocols under higher link bit rate will be much less affected than type 2 protocols by a non-favorable relation *Tc/Td*, since reservation overhead for connection establishment is per connection, not per packet and all the packets transmitted in multiplex are not causing any overhead at all. Further, to keep the reservation overhead for connection establishment low, under a low to medium network load a connection could be kept for the duration *hang-on* time, even if partially unused, until the radio resource occupied is requested by some other station.

12.1.3.2 Multi-hop Links

The coverage area of a station of future wireless broadband systems is likely to reduce compared to current systems since transmit power is limited and attenuation is higher at higher carrier frequency, see Figure 10.1. Mobile multi-hop relaying will be necessary to connect a station to non-direct neighbor stations and the design of the MAC protocol will affect the capability of a network supporting multi-hop.

Under multi-hop communication hidden and exposed stations must be taken into account to be able to efficiently allocate radio resources and neither waste capacity nor be subject to unpredictable interference or packet collisions:

- **Type 1** protocols are known to badly perform under multi-hop operation, see Walke (2002, Chapter 15).
- **Type 2** protocols have been proven to nicely perform under multi-hop operation in TDD mode, either performing relaying in the time or frequency domain, see Section 10.4.2.
- **Type 3** protocols can naturally be extended to support multi-hop operation, since they support one-hop links and can support multi-hop links by concatenating one-hop links.

A multi-hop supporting system based on type 3 protocols is represented in the next section, called MDCF, where any packet transmission is handled by the MAC protocol like a single-hop packet and packets on multi-hop routes are served by sequential single-hop links.

12.2 Mesh Networking for 802.11 WLAN[1]

In this section a novel MAC protocol is presented referred to as Mesh Distributed Coordination Function (MDCF) that may be used to interconnect APs to form a mesh network. MDCF has its origin in the Wireless Channel Oriented Ad-hoc Multi-hop Broadband (W-CHAMB) protocol (see Xu, 2002; Xu and Walke, 2002; Zhao and Walke, 2004; Zhao *et al.*, 2005).

The IEEE 802.11 infrastructure BSS formed by an AP and its associated stations is the most widely used WLAN system. There is a growing need to interconnect APs of BSSs by radio to form an Extended Service Set (ESS) mesh network, see Figure 12.5. APs thereby become mesh points (MPs) of an ESS mesh network and may deliver data packets by means of multi-hop relaying from a source to a destination MP. Some MPs may operate as a portal or gateway, providing access to other networks like the Internet. An MP represents a BSS in the ESS and meshes with the other MPs, while a station is associated to an AP in a BSS.

The MDCF is proposed to run on top of IEEE 802.11 a/g PHY with a minor driver modification. It is based on TDMA/TDD technology, operating under distributed control on a single frequency channel. More than one frequency channel may be used for increasing the bandwidth and reducing overhead to organize coexistence of both protocols. MDCF can properly handle hidden stations, exposed stations and problems resulting from signal capture as described in Sections 4.4.1 and 4.4.2, even at high load in a mesh network. MDCF is capable of distributing bandwidth fairly between end-to-end flows and performing multi-hop delivery of real-time services under specific QoS requirements.

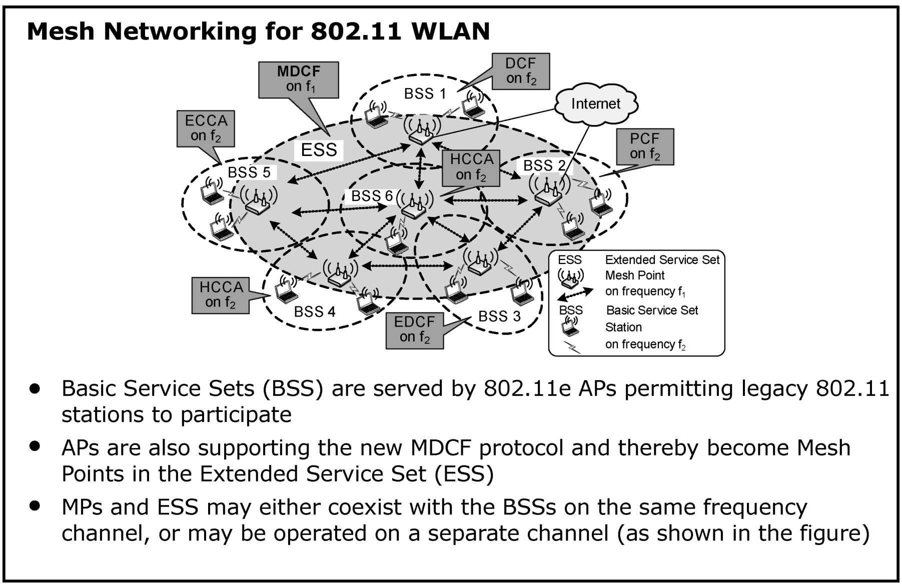


Figure 12.5 Proposed architecture for creating an ESS mesh network.

The IEEE 802.11b/g and IEEE 802.11a Physical layers (PHYs) provide 3 and 12 non-overlapped frequency channels, respectively, which may be used simultaneously. An ESS mesh network can be created with one of the 802.11 PHYs in the following way: APs form an MDCF mesh network under MDCF control on one frequency channel, while an 802.11 legacy or 802.11e station is associated with a nearby AP in a BSS and communicates with the AP on another frequency channel. This scheme provides a good backwards compatibility with the BSS networks operated with any of the current existing 802.11 protocols.

12.2.1 Mesh Distributed Coordination Function

The MDCF is a MAC protocol operating under distributed control on a single frequency channel. Transmissions take place in periodic time slots. The operation of the mesh network requires synchronization among the involved stations. The MDCF Timing Synchronization Function (MTSF) for implementing synchronization is described in IEEE (2005l). Unlike the Distributed Coordination Function (DCF) and Enhanced Distributed Channel Access (EDCA), where a station must contend for channel access to transmit one or several frames (Transmission Opportunities – TXOPs – in EDCA), in the MDCF a station contends for channel access to reserve a TDMA channel operated between adjacent stations. In terms of Section 12.1.2, MDCF applies a type 3 protocol. The channel, when allocated, enables QoS support controlled by the two stations it is connecting and, typically, is used to multiplex all packets routed single- or multi-hop across this link. Hidden stations are calmed down in MDCF by transmitting periodic energy signals at the receiver, known as receiver busy tone (Haas and Deng, 2002; Sobrinho and Krishnakumar, 1999). Energy signals serve for three purposes: (i) the implementation of an efficient, prioritized and fair channel access; (ii) calming down hidden stations nearby a receiver; and (iii) the implementation of a dynamic TDD scheme.

In the following, general operational mechanisms of the MDCF with a focus on challenging issues in mesh networks are presented. Unless otherwise stated, all the following time parameters are example values assuming a IEEE 802.11a PHY (IEEE, 2003a).

12.2.1.1 TDMA Frame and Energy Signals

An energy signal is a non-modulated single on-off pulse, occupying a short time slice, e.g. $6\,\mu$s. A receiver only needs to sense it to derive the meaning. Hence, the influence range of an energy signal transmitter is up to the carrier sense range of it.

As shown in Figure 12.6, a TDMA frame of the MDCF contains a number of time slots that are logically grouped into three types: (i) Access Channel (ACH), where energy signals and access control data are transmitted to implement a prioritized and fair channel access; (ii) Traffic Channel (TCH), where a slot can carry one MAC protocol data unit (MPDU) per TDMA frame; and (iii) Echo Channel (ECH), each is paired to a TCH slot, resulting in the same number of ECH slots in a TDMA frame as that of TCH slots. An ECH is used to transmit an energy signal per frame by an MP receiving in the corresponding TCH to notify its nearby MPs that the corresponding TCH is in use.

Energy signals transmitted in the ACH (Access-E-Signals, AES) have a single burst nature, while those transmitted in ECHs are periodic and called Busy-E-Signals (BES). BES might be Single Value BES (SVB) or Double Value BES (DVB) according to the signal duration, see Figure 12.6. An AES has the same waveform as a DVB. A SVB is transmitted on the ECH by a receiving MP for inhibiting hidden MPs. If the MP wants to turn around the transmit direction in TDD mode, a DVB is sent in the ECH instead of an SVB.

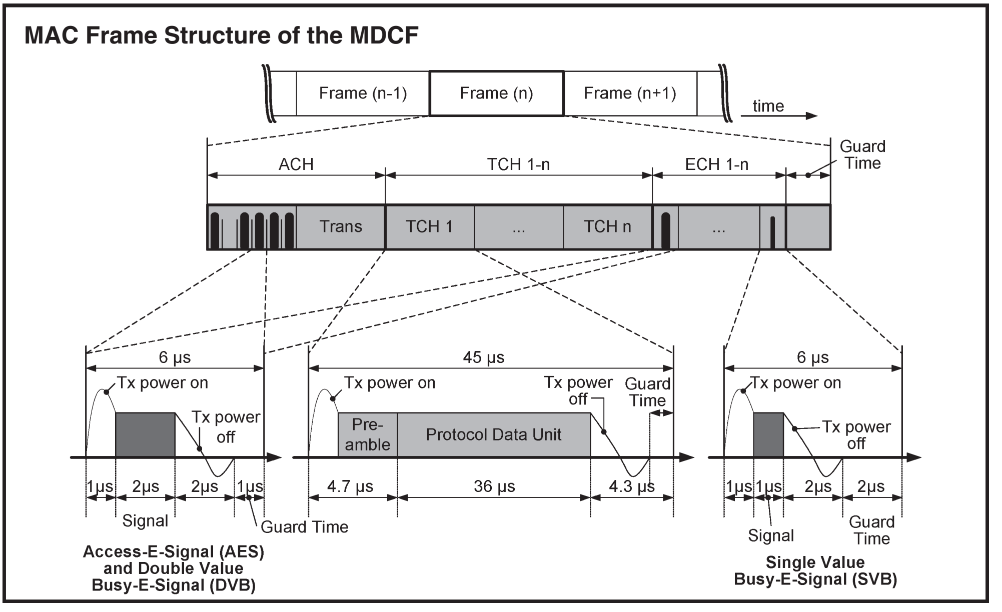


Figure 12.6 TDMA MAC frame structure of the MDCF and its usage of energy signals. Reproduced by permission of © 2006 IEEE[1].

An MPDU fits to a TCH slot, Figure 12.6. It is preceded by a PHY preamble to enable decoding at the receiver. The MPDU length depends on the PHY mode. Energy signals and MPDUs can be transmitted with the IEEE 802.11 a/b/g PHYs, requiring only minor driver modifications.

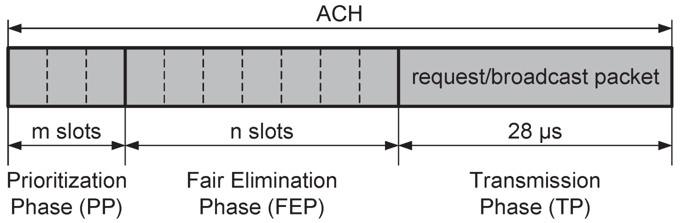
An ACH slot has three phases as shown in Figure 12.7(a): Prioritization Phase (PP), Fair Elimination Phase (FEP) and Transmission Phase (TP). The PP and FEP consist of *m* and *n* contention slots, respectively, each slot is one AES long. PP is used to differentiate high QoS level traffic flows from others by prioritization. FEP serves to guarantee with a high probability only one winner in each ACH contention and to ensure a fair channel access chance for each flow being maintained. A number of AESs are transmitted in the contention slots of the first two phases. Parameters like the number of TCHs in a TDMA frame, waveforms of energy signals and number of contention slots in the ACH may be different for different PHY layers and applications, but are never changed during operation.

12.2.1.2 Prioritized Channel Access

Contention process

HiperLAN/1 (Walke, 2001) and Black Burst Contention (Sobrinho and Krishnakumar, 1999) use successive bursts or pulses of energy to implement contention access. The amount of contention levels is equal to the number of contention slots. MDCF implements a contention level by transmitting a number of AESs as a binary number in contention slots. The contention levels amount to $2^{(\text{number of contention slots})}$. This is a significant improvement: HiperLAN/1 uses 31 contention slots to obtain 31 different levels, whereas MDCF uses 12 slots to obtain $2^{12} = 4096$ different levels. The contention levels in the PP and FEP are called PP contention levels (PPCLs)

(a) Access Channel



(b) Prioritization Phase Contention Levels (PPCL)

Access Category	Traffic Type	Delay	PPCL
AC_Beacon	Control	-	7
AC_VO_HIGH	Voice	> 30 ms	6
AC_VO	Voice	<= 30 ms	5
AC_VI_HIGH	Video	> 50 ms	4
AC_VI	Video	<= 50 ms	3
AC_BK_HIGH	Background	>200 ms	2
AC_BK	Background	<= 200 ms	1
AC_BE	Best Effort	-	1

Figure 12.7 (a) Access channel structure of the MDCF and (b) contention levels of prioritization phase. Reproduced by permission of © 2006 IEEE[1].

and FEP contention levels (FEPCLs), respectively. The amounts of PPCLs and FEPCLs are 2^m and 2^n, respectively.

An MP wanting to reserve a TCH, or transmit an MPDU in the TP must contend in the ACH for transmitting a request packet or an MPDU in the TP. The contention process is as follows: an MP selects a PPCL as specified in Figure 12.7(b). The higher a number, the higher is the access priority. Then it transmits the number bit by bit, when the bit is 1 it sends an AES, for 0 it listens. The most significant digit is transmitted first. When listening, if it receives an AES, it must cancel its pending bits and quit the contention in the current frame. If surviving the PP, the MP must contend again in the FEP with a number determined as Fair Elimination Channel Access below. If the MP wins both PP and FEP, it is allowed to transmit in the TP. If losing, it will contend again in the next TDMA frame.

Figure 12.8 illustrates a contention process. MP S_1, S_2 and S_3 contend for channel access at the same time. S_1 and S_2 want to reserve TCH(s) for transmitting Voice-over-IP (VoIP) MPDUs, while S_3 has pending video MPDUs. Assume that the PPCLs of the VoIP and video MPDUs are 5 (101) and 3 (011), respectively. Both S_1 and S_2 win in the PP by means of listening and sending AESs. After that, each of them generates a FEPCL and competes again in the FEP. The FEPCL of S_1 is 441 (110111001), whereas the FEPCL of S_2 is 283 (100011011). S_2 quits the contention when it senses an AES in the FEP. Finally, S_1 sends out a request packet in the TP.

PP Contention Levels (PPCLs)

Figure 12.7(b) defines the PPCLs used for different access categories in the PP. Beacons used for synchronization have the highest priority. According to the delay requirements and load

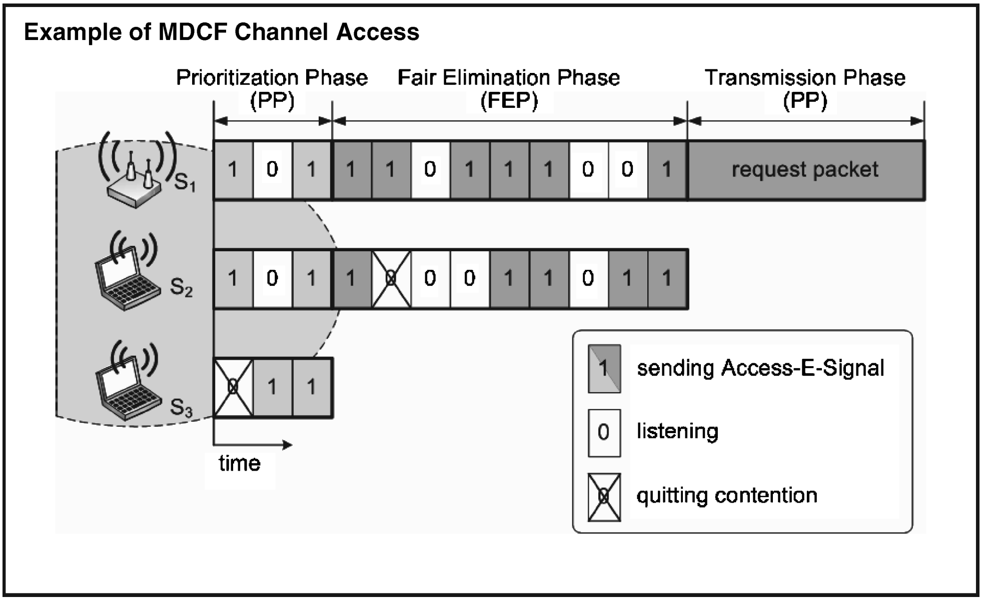


Figure 12.8 An example of channel access. Station S_1, S_2 and S_3 are in the transmission range of one another.

behaviors, traffic flows are given different PPCLs that can dynamically change according to the delay experienced by MPDUs waiting for transmission. The access categories are ordered as: AC_VO (Voice), AC_VI (Video), AC_BK (background) and AC_BE (best effort). When the highest delays of pending MPDUs in voice, video and background flows exceed 30, 50 and 200 ms, respectively, the related PPCLs of those traffic classes are upgraded to AC_VO_HIGH, AC_VI_HIGH and AC_BK_HIGH, respectively. If the highest delay of pending video MPDUs in a flow exceeds 75 ms (3/4 of the dropping threshold), the PPCL of the flow is upgraded to AC_VI_SUP, equal to AC_VO.

In principle, the PPCL is set in favor of time urgent and low traffic volume flows. Relaying MPs along a multi-hop route will be given higher PPCL since the packet delay on a multi-hop route tends to increase. Then, relaying MPs will win over others, reducing the end-to-end delay. Meanwhile, if a source MP loses contentions to the relaying MPs for other flows for a long time, it increases the PPCL. As a result, the starvation of the MP is avoided. When the residual life times are same, low traffic volume flows such as voice flows are served first for achieving shorter overall delays and jitters.

When an MP has multiple types of traffic flows going to a same next MP, it determines its PPCL from the flow with the highest PPCL to compete for TCH reservation.

Fair Elimination Channel Access

A wide range of contention levels implements fair elimination channel access during the FEP. Even when a large amount of MPs contending at the same time with same PPCL, with probability close to 1, only one MP survives and transmits in the TP. Others are eliminated for transmitting in the TP at the current TDMA frame. However, a fairness mechanism helps losing MPs to win contention in future. The more an MP loses contention, the higher chance it will win in the next TDMA frame.

2^n contention levels (FEPCLs) are grouped into K equal sized non-overlapping Contention Number Groups (CNG): $[0, (2^n/K) - 1], [2^n/K, (2 \times 2^n/K) - 1], \ldots, [(K-1) \times 2^n/K, 2^n - 1]$. At a time, there may be several MPs contending for channel access with same PPCL. Each contending MP maintains variables t^i counting the number lost for a given flow, where i is the flow ID. When an MP loses a contention in the FEP for competing for the flow i, it will increment t^i by 1. If it wins, it resets t^i to 0.

When several flows in an MP request TCH reservation at the same time, the MP will contend for the flow with the highest PPCL. If more than two flows with a same highest PPCL exist, the MP contends for the flow with the largest t^i. In case of more than two flows with a same largest t^i, the MP randomly selects one flow and competes for it. t^i of other flows will be incremented by 1.

After winning the contention in the PP, an MP determines a CNG for generating an FEPCL. Each CNG is associated with a group selection threshold $T^k, k \in [0, K-1]$. If $T^k \leq t^i < T^{k+1}$ when $k < K-1$, or $T^k \leq t^i$ when $k = K-1$, then the k^{th} CNG is selected, from which an FEPCL randomly is generated. Note that an FEPCL from a higher CNG is bigger than that from a lower one. The amount of contention levels in a CNG is $2^n/K$, which should be big enough to ensure only one winner even under heavy contention. The elimination phase is very important enabling MDCF to handle highly loaded situations.

12.2.1.3 Link Setup and Traffic Channel Reservation

When an MP wishes to transmit MPDUs, it needs to reserve TCHs in agreement with the receiving MP. The MP first checks the local TCH status. A TCH is considered free if:

- No carrier is sensed in the TCH; and
- No BES is sensed in the paired ECH of the TCH.

If the amount of available TCHs observed meets the traffic need, an MP will contend in the ACH. If it wins, it transmits a request packet for TCH reservation containing the receiver address, QoS-related Traffic Specification (QTS) and a list of proposed TCHs in the TP of the ACH. On reception of the request packet, the requested MP checks the free TCHs at its location, and then performs an Admission Control (AC) algorithm to check whether: (i) the common free TCHs at both sides are adequate for the QoS delivery; (ii) establishment of the link will not corrupt the QoS of the existing flows. It accepts the request if both conditions are met and then transmits SVB(s) in the ECH(s) paired to the accepted TCH(s). From the SVB(s), the requesting MP knows that the TCH(s) have been reserved, whereas nearby MPs derive that the TCH(s) are in use. Figure 12.9 shows the related message chart. Later on, transmission takes place in the TCH(s).

12.2.1.4 Transmission and On-demand-TCH Turnaround

Hidden MPs are in the carrier sense range of a receiving MP. In MDCF, the receiving MP transmits SVBs or DVBs in ECHs to prevent hidden MPs from using the related TCHs.

Allocation of fixed time slots to a transmission pair helps to support QoS. But when traffic is bidirectional and especially asymmetric, how to efficiently utilize time slots is a challenging

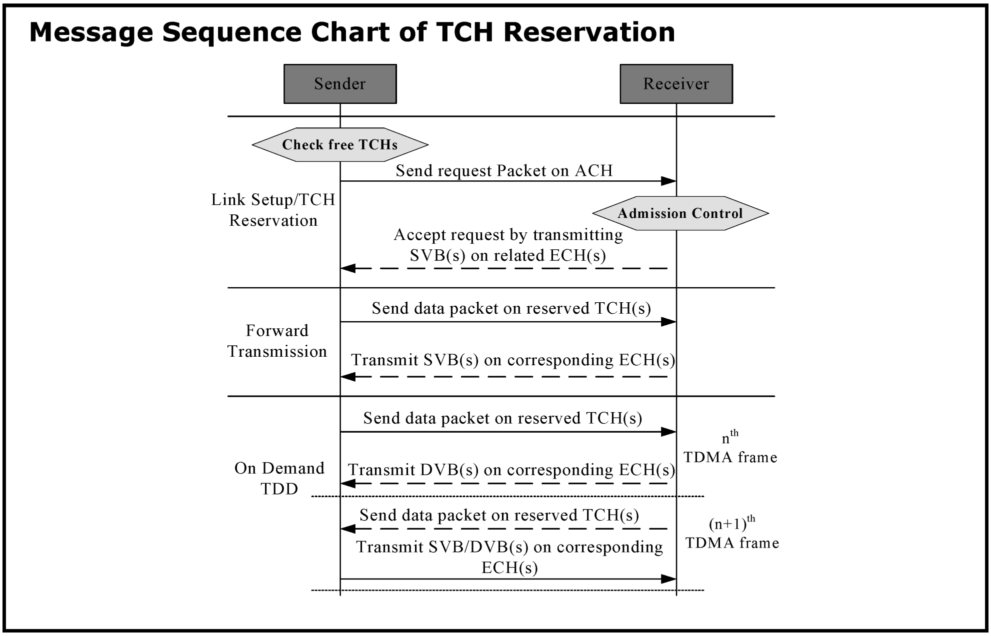


Figure 12.9 The processes of TCH reservation, transmission and on-demand-TDD. Reproduced by permission of © 2006 IEEE[1].

issue. We propose an on-demand TCH turnaround to address this issue. The transmission process is as follows:

Assume that two MPs reserve several TCHs, on each of which the sender transmit its MPDUs or dummy packets when it has no MPDU to transmit. As long as the receiver receives MPDUs or dummy packets in a reserved TCH (or a corrupted packet), it transmits an SVB in the ECH paired to the TCH, to signal the use of the TCH. If the receiver has MPDUs to send back, it transmits a DVB instead of an SVB in the ECH paired to a reserved TCH. If the sender senses the DVB, from the next MAC frame on, it stops transmission in the TCH and starts to transmit energy signals in the related ECH, and the receiver starts to send its MPDUs in the TCH. This mechanism is called On-Demand TCH turnaround, see Figure 12.10.

It is worth noting that there is no need to transmit the 802.11 addresses in each MPDU, which have been exchanged during reserving TCHs. On receiving an MPDU in a TCH, an MP knows where the MPDU comes from and is destined for. Overhead is greatly reduced compared to DCF/EDCA.

The On-demand TCH turnaround substantially increases the channel utilization. A receiving MP checks the number of pending MPDUs in both sides before using it. If necessary, one side will initiate *TCH reservation* to request more TCHs.

12.2.1.5 Packet Multiplexing and Multi-hop Operation

A multi-hop link consists of multiple one-hop links in tandem, each of which independently operates on one or more TCHs. A reserved TCH is used to multiplex any MPDUs transmitted on the route. The MPDU transmission sequence is according to their access categories (0–6) shown

On-demand-TCH Turnaround

If	$N_{pdu}^{T} = 0$ ***or*** $(N_{TCH}^{M} > 0$ ***and*** $N_{pdu}^{T}/N_{TCH}^{M} < Thr\_1)$
	Then exit
Else If	$N_{TCH}^{M} = 0$ **and** $N_{TCH}^{P} > 0$ **and** $N_{pdu}^{T} < Thr\_2$
	Then request On-demand TCH Turnaround
Else if	$N_{TCH}^{P} > 1$ **and** $N_{pdu}^{P}/(N_{TCH}^{P} - 1) < Thr\_3$
	Then request On-demand TCH Turnaround
Else if	$N_{TCH}^{F} > 0$ **and** $N_{pdu}^{r} > 0$ **and**
	$(N_{TCH}^{M} = 0$ ***or*** $N_{pdu}^{r}/N_{TCH}^{M} > Thr\_4)$
	Then initiate TCH Reservation
Else if	$N_{TCH}^{F} > Thr\_5$ **and**
	$(N_{TCH}^{M} = 0$ ***or*** $N_{pdu}^{T}/N_{TCH}^{M} > Thr\_6)$
	Then initiate TCH Reservation

Figure 12.10 Requesting On-demand-TCH turnaround or TCH reservation during transmission. Reproduced by permission of © 2006 IEEE[1].

in the table of Figure 12.7(b). Multiplexing of packets of any source–sink relationship to a link provided by a TCH increases its utilization. The transmitting sequence of packets is determined according to their QoS priorities. Under high mesh load the bottleneck MP will tend to have no more channel access overhead since it is then operating links on TCHs all the time to its neighbor MPs. As a result, the network throughput and delay performance are significantly improved. The MDCF supports routing protocols proposed for mobile ad-hoc networks.

12.2.1.6 Coexistence

IEEE 802.11 PHYs operate at unlicensed bands. Coexistence of MPs using MDCF with legacy stations using DCF/EDCA is an important issue. We have two solutions for this. Both rely on the ability of legacy stations to respect the superframe structure announced by a point coordinator using PCF: MDCF MPs periodically transmit beacons specifying a superframe containing Contention Period (CP) and Contention-Free Period (CFP). On reception of a beacon, a legacy station knows that it is not allowed to transmit during CFPs unless it is polled.

- *Simple solution:* If MPs announce and initiate a new CFP before the current CFP is expired, then legacy stations will be inhibited from transmitting. The approach requires low overhead. Beacons can be sent every 80 ms. This approach is adopted for the performance simulation studies described below.
- *Sharing a channel:* MDCF MPs announce the network wide uniform frame structure and operate in CFP, while legacy stations operate in CP, as described in Section 8.3.

12.2.2 Performance Evaluation Results

12.2.2.1 Simulation Tool

The MDCF is implemented in an event-driven simulator based on the Specification and Description Language (SDL) Performance Evaluation Tool Class Library (SPEETCL) (Steppler, 1998) in C++.

12.2.2.2 Simulation Results – QoS Performance in Mesh Networks

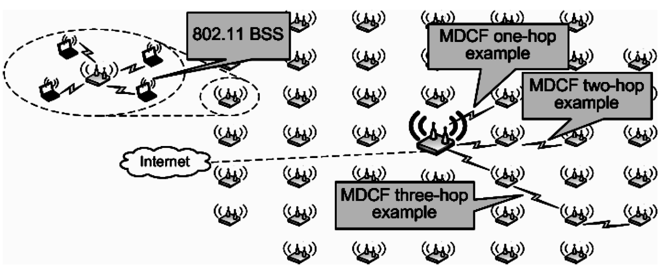
The section concentrates on evaluating the multi-hop and QoS performance of the MDCF. For comparison reasons, simulations are performed with exactly the same scenario by using the EDCA with parameter sets according to Section 5.5.

A grid topology shown in Figure 12.11 is used for evaluation. Every station is 100 m away from its direct neighbors. Each station is an AP of an 802.11 BSS network. The APs form an overlaying mesh network on one frequency channel while BSS networks operate on other frequencies. The central AP is the only station in this network connected with the Internet. This scenario is a typical ESS mesh network. Communication is done between a station in a BSS and a server in the Internet. Therefore communication in the mesh network is between the central AP and an AP forwarding the data from a station in its BSS. APs perform multi-hop forwarding to or from the central AP. Thus, the central AP is a bottleneck station. We evaluate the network capacity for delivering QoS traffic. The results achieved by using the MDCF and EDCA for forming a multi-hop network between APs are compared. We assume that each BSS network is lightly loaded. The packet delay from a station to its AP (less than 1 ms in a lightly loaded BSS) and packet delay from the central AP to a server in the Internet are ignored. The hops evaluated in Figure 12.12 and Figure 12.13 are counted from any AP to the central AP.

Important simulation parameters are summarized in the table of Figure 12.11. With the transmission power assumed, the transmission range of data packets is calculated to 100 m while the carrier sense ranges of data packets, BESs, and AESs are approximately 270 m, 270 m and 370 m, respectively. The PHY data rate with MCS 16-QAM 1/2 is 24 Mbit/s.

First Study: We investigate the number of real-time traffic flows that can be served in parallel for a given traffic type and given number of hops. The results obtained by using MDCF and EDCA, to form a multi-hop mesh network are compared. According to Figure 12.13(a), MDCF

Mesh Operation with MDCF – Simulation Scenario



Mesh Operation with MDCF – Simulation Parameters

Parameter	Setting
Transmission power for MPDUs and BESs	80 mW
Transmission power for AESs	160 mW
Carrier sense threshold	- 85 dBm
Minimum receiving sensitivity	- 74 dBm
PHY mode for data packets and rate	16QAM ½, 24 Mbit/s
Length of a MDCF MPDU (ARQ packet size)	108 bytes
ARQ Poll period	20 PDUs

Figure 12.11 Scenario and parameters used in simulating mesh operation with MDCF.

Traffic Loads and their QoS Requirements

Traffic	Load (kbit/s)	PDU Size (bytes)	Max. Delay (ms)	Max. Packet Loss Rate (PLR)
Video Conference	Mean:256 Max:1280	512	100	0.1%
VoIP	22.4	160	60	6%

Simulation Results – Complementary CDF of End-to-End Delay

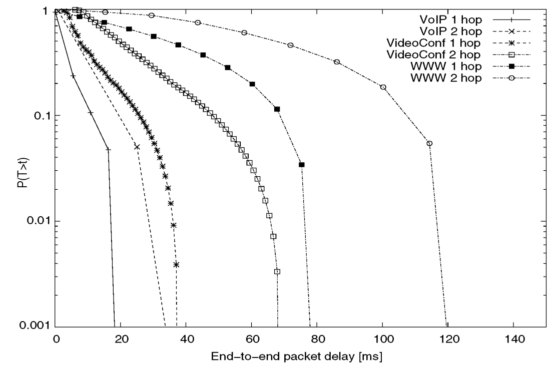


Figure 12.12 Comparison of simulation results of MDCF and EDCA for VoIP connections and video conferencing in mesh networks. Reproduced by permission of © 2006 IEEE[1].

is able to support 120 × 1-hop, 45 × 2-hop, and 18 × 3-hop concurrent VoIP flows with a Packet Loss Rate (PLR) ≤ 6%. The delay performance is shown in Figure 12.13(b).

As shown in Figure 12.13(b), under EDCA, end-to-end delay is much lower than with MDCF. Transmission of a VoIP packet in a one-hop EDCA network needs less than 0.4 ms under light load, where an MP can seize the channel and finish its transmission much faster than with MDCF. However, with the increase of traffic load, the performance of the EDCA-based ESS mesh network deteriorates sharply, see Figure 12.13(a), it is able to support 30 × 1-hop, 14 × 2-hop and 5 × 3-hop concurrent VoIP flows with PLR ≤ 6%. An EDCA MP drops a packet when it fails to retransmit the packet more than seven times. A relatively small Contention Window (CW) results in a considerable number of collisions. Increasing the retry time does not help to reduce the PLR but only increases the delay. EDCA cannot perform well in a highly loaded network. The only way to reduce the PLR is to increase the CW size for VoIP, which, however, leads to a longer delay and a low efficiency for delivering video and background traffic, because CW sizes for those traffic should be increased accordingly. In contrast, MDCF can support much more concurrent VoIP flows than EDCA with higher delay, but well below the delay requirement.

A video conference source generates highly bursty traffic. Given that the tolerable PLR is 0.1 %, the supported numbers of concurrent 1-, 2- and 3-hop video flows in the ESS mesh network when using MDCF are 24, 10 and 6, respectively, as plotted in Figure 12.13(c). In contrast, the ESS mesh network when using EDCA, can only support 8 × 1-hop, 2 × 2-hop and 1 × 3-hop concurrent video flows, respectively, 1/3, 1/5 and 1/6 of those under MDCF. A video flow offers a much higher load than a VoIP flow does, leading to a much higher collision rate. Though the packet delay is very small, the PLR of video packets is quite high. EDCA cannot make a tradeoff between the low delay and high PLR.

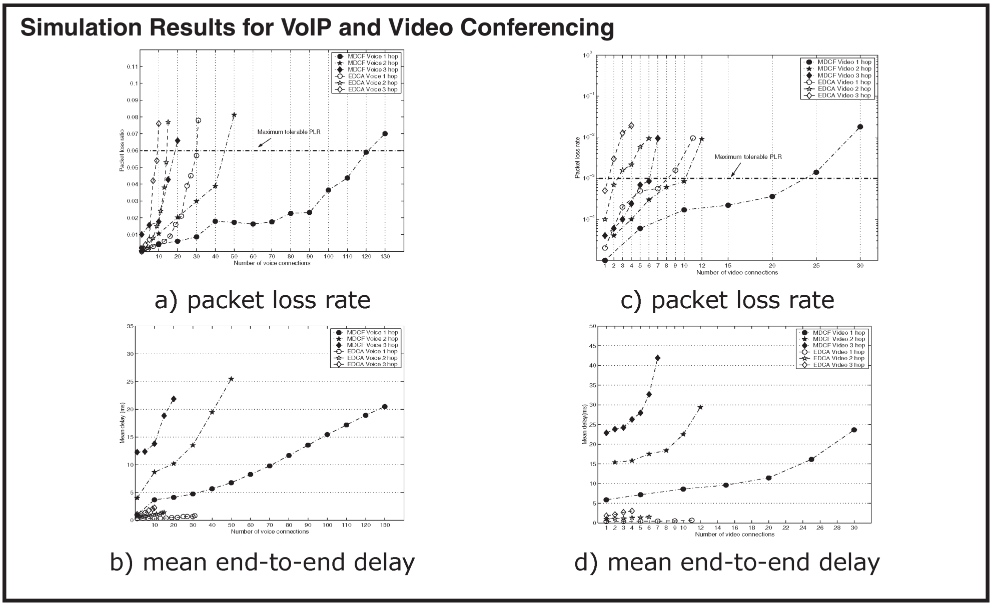


Figure 12.13 Comparison of simulation results of MDCF and EDCA for VoIP connections and video conferencing in mesh networks. Reproduced by permission of © 2006 IEEE[1].

Second Study: We reveal the QoS performance of the ESS mesh network when using MDCF to handle mixed traffic services in the grid topology shown in Figure 12.11. Since EDCA cannot support even three concurrent 2-hop video flows, Figure 12.13(c), no comparison is possible here.

The traffic flows are assumed to be divided into six groups. Flows generated from a given traffic type spanning a given number of hops constitute a group. The details of the group information are given in Figure 12.12 where the Complementary CDF of end-to-end delay for each group is shown. For a given number of hops, VoIP and video flows achieve the lowest and second lowest delay. For a given traffic type, the delay of 1-hop transmission is almost half than that of 2-hop transmission. It can be concluded that the ESS mesh network is able to serve 42 mixed end-to-end traffic flows in parallel, meeting the particular QoS requirements. The QoS support of MDCF works efficiently in an ESS mesh network.

Detailed results and a comprehensive description of the MDCF can be found in Zhao *et al.* (2006).

12.3 Conclusion

Originating in an analysis of state-of-the-art MAC protocols, we proposed a novel MAC protocol referred to as MDCF for creating an IEEE 802.11 ESS mesh network. MDCF as a type 3 protocol is based on TDMA and is able to operate with the IEEE 802.11a/g PHY. The MDCF can handle high load situations and is able to support in mesh operation a large number of QoS flows. It drastically outperforms the EDCA in the analyzed multi-hop environment. As a result, an ESS mesh network can be realized based on MDCF that is able to interconnect APs to form an ESS using one frequency channel while these APs serve a large number of 802.11(e) stations on other frequency channels. Coexistence of both MDCF and EDCA/DCF is also a possibility.

Note

1. Reproduced by permission of © 2006 IEEE. Source: R. Zhao, B. H. Walke, and G. R. Hiertz, "An Efficient IEEE 802.11 ESS Mesh Network Supporting Quality of Service," *IEEE Journal Selected Areas in Communications (JSAC), Special Issue on Multi-Hop Wireless Mesh Networks*, Vol. 24, No. 11, Nov. 2006 (Section 12.2 (in parts) Fig. 12.6, 12.7, 12.9, 12.10, 12.12, 12.13).

13

Cognitive Radio and Spectrum Sharing

Lars Berlemann, Stefan Mangold and Bernhard H. Walke

In recent years, various research organizations initiated programs and projects aiming at the improvement of flexibility in wireless communication. In the US, the SDRforum is targeting at a reconfigurable system architecture for wireless networks and user terminals on the basis of Software Defined Radio (SDR) (SDRforum, 2005). The DARPA NeXt Generation Communication (XG) Program, financed by the US government, aims at developing a de-facto standard for cognitive radio, and dynamic spectrum regulation (DARPA, 2003, 2004a, 2005). The IST projects WINNER (Mohr, 2005; WINNER, 2005) and E^2R (Bourse *et al.*, 2004, 2005; E^2R, 2005) of the Sixth Framework research funding Program (FP6) of the European Union, belonging to the Wireless World Research Initiative (WWI), both concentrate on different levels on flexible radio interfaces and network architectures. E^2R, just as DARPA XG, is also working on flexible and dynamic spectrum usage and related impacts on spectrum regulation.

13.1 From Software-Defined Radio to Cognitive Radio

13.1.1 Software-Defined Radio and Software Radio

The fundamental intent of software-based radios is to shift the hardware-oriented application-specific approach to communication devices to flexible software applications performing communication functions on a common computing platform. If the communication functions of a transceiver are completely realized as programs operating on a suitable processor, this transceiver is referred to as Software Radio (SR) (Mitola, 1995). The digitalization is done directly after radio wave reception at the antenna and all the signal processing is done by software. SDRs (Tuttlebee, 2002a, 2002b) are a prior step in the evolution to SR and thus more practicable: the analog signals are processed after a suitable band filter is selected. SDRs are a factual reality

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

when taking the recent convergence of software and digital radio into account. Although the SR concept is related to all layers of the protocol stack many research efforts concentrated in the past on the physical layer (PHY). In the last five years this scope was extended to the complete software of a device used for communication in introducing new concepts (and buzzwords) which are all based on SDRs as introduced in the next sections. A communication system of SDRs can have the following characteristics:

- Multi-band system operating in multiple frequency bands as for instance in the unlicensed bands at 2.4 and 5 GHz.
- Multi-standard system supporting more than one radio standard of the same protocol family (e.g. IEEE 802.11a/b/g) or of different radio access technologies (UMTS, GPRS and IEEE 802.11b).
- Multi-channel system enabling simultaneous transmission and/or reception on multiple channels.

The multi-mode systems, which are currently the focus of many research initiatives, combine these three characteristics. The SDRforum expanded its understanding of SDR to the same extent as described by the reconfigurable radio concept below (SDRforum, 2005).

13.1.2 Composite Radio and Reconfigurable Radio

The devices of a heterogeneous wireless infrastructure that use multiple modes (possibly simultaneously) can be referred to as *composite radios* (Demestichas *et al.*, 2003). The composite radio concept implies pre-installed modes complementing each other for optimized network utilization and QoS support. The dynamic selection, installation and adaptation of the devices' modes to the communication environment and changing user demands are introduced with the *reconfigurable radio* concept (Demestichas *et al.*, 2004). The term reconfigurable radio is mainly affected by European research. The technical realization of a reconfigurable terminal, such as mode monitoring, mode switching or software download, is discussed in Mehta *et al.* (2001). Reconfigurability provides the basis for the following key objectives (Dillinger *et al.*, 2003):

- Adaptation of the radio interface to the locally present communication environments and radio interface standards.
- The integration of new applications and services.
- Software updates and over-the-air download.
- Exploitation of flexible heterogeneous services provided by the radio network.

With these key objectives, reconfigurable radios raise the focus from the terminal centric view to the network perspective. In this context the term end-2-end reconfigurability is often used which considers all elements of the communication network including services and applications, as investigated in the Integrated Project E^2R. Contrary to this centralized view from the network level, the cognitive radio network, as introduced in Section 13.2, is based on a distributed network architecture.

13.1.3 Cognitive Radio

Cognitive radio will lead to a revolution in wireless communication with significant impacts on technology as well as regulation of spectrum usage to overcome existing barriers. Cognitive radio, including SDR as the enabling technology, was first suggested in Mitola and Maguire (1999) and Mitola (2000) for the flexible and efficient usage of spectrum. Cognitive radios are a further development of SRs, which again emerged from SDRs. Thus, cognitive radio is the necessary step from a flexible physical layer to a flexible system as a whole.

The term cognitive radio is derived from "cognition". According to Wikipedia (2005) cognition is referred to as

- *Mental processes* of an individual, with particular relation
- *Mental states* such as beliefs, desires and intentions
- *Information processing* involving learning and knowledge
- Description of the emergent *development of knowledge* and concepts within a group

Resulting from this definition, the cognitive radio is a self-aware communication system that efficiently uses spectrum in an intelligent way. It autonomously coordinates the usage of spectrum in identifying unused radio spectrum on the basis of observing spectrum usage. The consideration of spectrum as being unused and its usage involves regulation, as this spectrum can be originally assigned to a licensed communication system. This is introduced as vertical spectrum sharing in Section 13.3.3. Besides the cognition in radio resource management, cognition in services and applications is also provided by cognitive radios to enable transparency to the consumer. The mental processes of a cognitive radio based on the cognition circle from Mitola (2000) are depicted in Figure 13.1. Cognition is illustrated using the example of flexible radio spectrum usage and the consideration of user preferences. In observing the environment, the cognitive radio decides about its action. An initial switching on may lead to an immediate action, while usual operation implies a decision taking based on learning from observation history together with the actual state of the environment.

The FCC (2005a) has identified the following, less revolutionary features that cognitive radios can incorporate to allow a more efficient, flexible spectrum use:

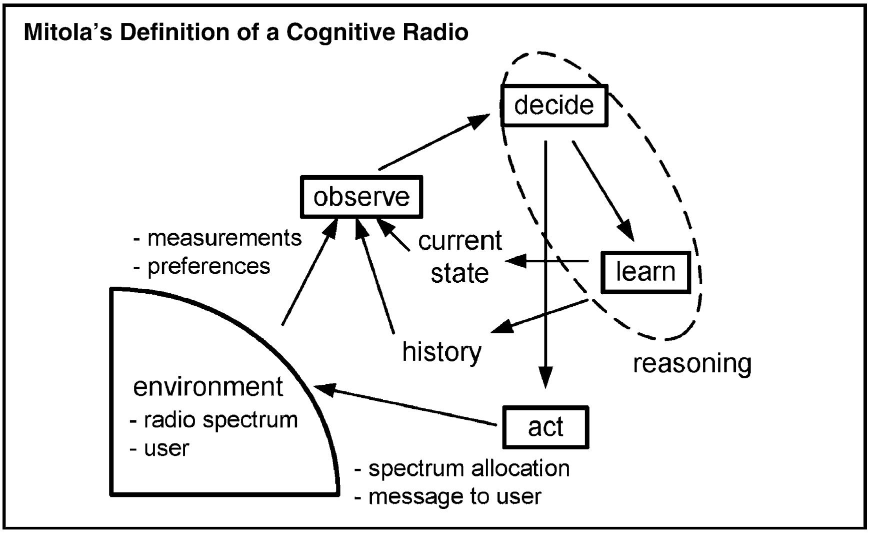


Figure 13.1 Mental processes of a cognitive radio based on the cognition cycle from Mitola (2000).

- **Frequency Agility** – The radio is able to change its operating frequency to optimize its use in adapting to the environment.
- **Dynamic Frequency Selection (DFS)** – The radio senses signals from nearby transmitters to choose an optimal operation environment, see Section 13.4.1.
- **Adaptive Modulation** – The transmission characteristics and waveforms can be reconfigured to exploit all opportunities for the usage of spectrum.
- **Transmit Power Control (TPC)** – The transmission power is adapted to full power limits when necessary on the one hand and to lower levels on the other hand to allow more efficient sharing of spectrum, see Section 13.4.2.
- **Location Awareness** – The radio is able to determine its location and the location of other devices operating in the same spectrum to optimize transmission parameters for increasing spectrum reuse.
- **Negotiated Use** – The cognitive radio may have algorithms enabling the sharing of spectrum in terms of prearranged agreements between a licensee and a third party or on an ad-hoc/real-time basis, see Section 13.3 and thereafter.

Strictly following this definition, modern wireless LANs could already be referred to as cognitive radios: IEEE 802.11 devices operate with a listen-before-talk spectrum access and with dynamically changing frequencies and transmission power.

In later research, cognitive radios are also referred to as "spectrum agile radios" (Mangold *et al.*, 2004b; 2005a) to indicate the authors' emphasis on dynamic spectrum usage. Mangold *et al.* (2004b) focus thereby on IEEE 802.11k for radio resource measurements as an approach to facilitate the development of spectrum agile radios, while Mangold *et al.* (2005a) introduce spectrum agile radios as a society of value-oriented machines. Basic concepts that are taken from social science to classify the social action of independent decision-makers are applied to define system strategy rules.

This understanding of cognitive radios is summarized in the definition for cognitive radio from Haykin (2005):

> Cognitive radio is an intelligent wireless communication system that is aware of its surrounding environment (i.e., outside world), and uses the methodology of understanding-by-building to learn from the environment and adapt its internal states to statistical variations in the incoming RF stimuli by making corresponding changes in certain operating parameters (e.g., transmit power; carrier-frequency, and modulation strategy) in real-time, with two primary objectives in mind: (i.) highly reliable communication whenever and wherever needed and (ii.) efficient utilization of the radio spectrum.

A highly reliable communication includes in this context the aspect of QoS. Cognitive radios have to share spectrum and require therefore means for mutually coordinating their spectrum utilization to enable the support of QoS in such distributed environments.

13.2 Cognitive Radio Networks[1]

Cognitive radio networks are a logical generalization of cognitive radios. The extension of the focus from the individual cognitive radio to a cognitive radio network aims especially at improving spectrum utilization through spectrum reuse. Additionally, the coverage area can be increased when a meshed wireless backbone network of infrastructure links is established based on Cognitive Access Points (CAPs) and fixed Cognitive Relay Nodes (CRNs) as illustrated in

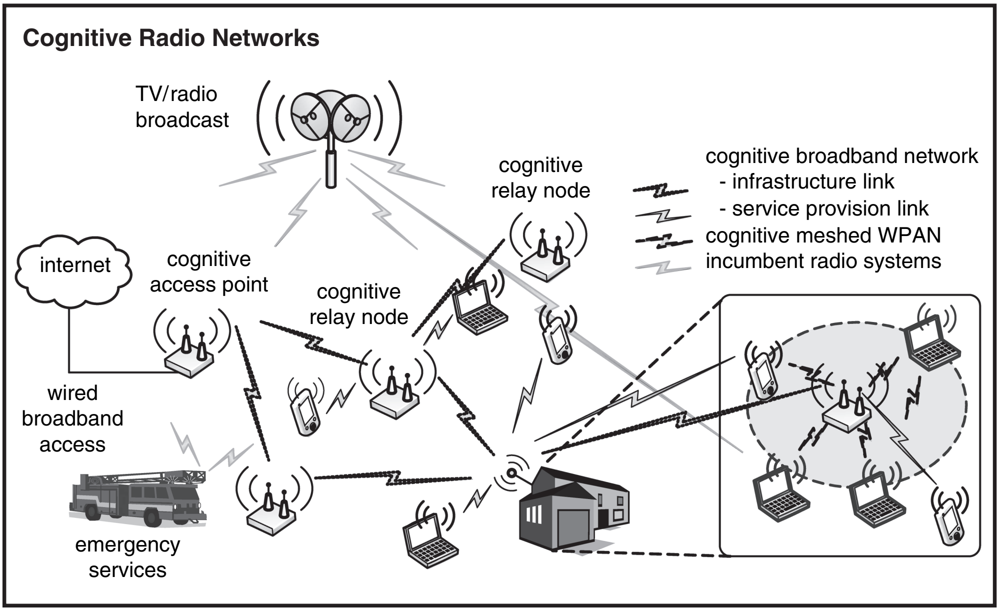


Figure 13.2 Scenario of different cognitive radio networks opportunistically operating in unused frequency bands.

Figure 13.2. The capacity of a CAP, connected via a wired broadband access to the Internet, is distributed into a large area with the help of a fixed CRN. The classical understanding of a planned relay-based deployment concept as introduced in Pabst *et al.* (2004) is extended here with the cognitive radio paradigm: The distributed operating cognitive radios form on their own a relay-based meshed network which flexibly uses radio spectrum. Such a cognitive radio network has for instance the ability to add temporarily or permanently bandwidth, i.e., spectrum, to the infrastructure links used for relaying in the case of high traffic load. The cognitive radio network operates on an opportunistic basis in unused frequency bands. Vertical spectrum sharing, as introduced in Section 13.3.3, thereby enables operation in frequency bands originally licensed to other radio systems such as TV/radio broadcasts or emergency services. The protocol stack of the cognitive radios supports multiple modes to optimally exploit the different characteristics of the wireless links of the cognitive radio network. The efficient realization of such a multi-mode protocol stack is introduced in Berlemann *et al.* (2005a), Berlemann *et al.* (2006e) and Schinnenburg *et al.* (2005).

13.2.1 Essential Characteristics

Several essential characteristics of a cognitive radio network can be identified which must be satisfied to increase the chances for economic success in the future:

- **Self-configuration**: CAPs/CRNs can be installed by the customer by simply switching on the device. This implies over-the-air software download for upgrades (e.g. of the communication protocol software or spectrum policies) and a self-healing in the case of malfunctions or partial network breakdowns. The self-configuration leads to a stable relaying infrastructure that enables a QoS guarantee involving power control, frequency selection, neighbor discovery and relaying configuration.

- **Low hardware costs**: All devices of the cognitive radio networks are mass-market products leading to consumer prices for CAPs/CRNs below US$100.
- **Small-sized hardware**: The cognitive CAPs/CRNs are more hamburger-sized than having the dimensions of a base station with an oversized transmission mast. CAPs/CRNs are inconspicuous, wall-mounted devices which lead to higher acceptance by the customers. This imposes deployments in revenue-rich urban areas as communal/neighborhood wireless broadband networks.
- **Transparency to the consumer**: The complexity of the cognitive radio network is hidden to the user, nevertheless a transparency of costs exists from the consumer perspective.
- **Energy consumption**: Spectrum Sensing and flexible modem/protocol stack should be designed for energy efficiency in order to realize mobile user devices.

These characteristics enable business models, where an operator offers the subvention of the hardware which builds the wireless network infrastructure. These CAPs/CRNs are installed by the consumer when buying terminal equipment of the cognitive radio network such as a WLAN/WPAN access point.

The policy-based reasoning as described in Section 13.7 enables operator-specific cognitive radio network deployments. Policies enable the distributed coordination necessary to realize self-configuration. They define the characteristics for the consideration of a link as relay link of the wireless infrastructure and take this into account in spectrum usage. Policies specify the restrictions to spectrum usage in terms of bandwidth and time of spectrum access that are required for the relay links to enable a QoS guarantee. From the operator perspective policies can be used to define behaviors of cognitive radios in the case of relaying requests from radios belonging to the same operator or others.

13.2.2 Spectrum Information Base

The Spectrum Information Base (SIB) is a location-specific database for spectrum use reflecting each cognitive radio's view on the "outside world". The SIBs are hierarchically administrated on the network level. Their comprehensiveness depends on the logical function fulfilled by the cognitive radio and its distance of potential interference. The potential interference to other radio systems depends mainly on its transmission power but also takes the propagation environment into account.

The SIB helps to increases accuracy of spectrum opportunity identification, eliminates the hidden station problem, reduces individual effort of measuring spectrum utilization, and is required for special reuse of spectrum.

The SIB is the basis for deciding, i.e. reasoning, about dedicated spectrum allocation. With its information about spectrum usage, the SIB is combined by the spectrum navigator with policies valid at this time and location, ending in the identification of spectrum opportunities. These spectrum opportunities are accessed under consideration of policies restricting radio transmission and enabling spectrum sharing. In the case of spectrum trading, as introduced in Section 13.3.1, the SIB helps to decide whether additional spectrum has to be bought from spectrum brokers.

13.2.3 Similar Approaches and Related Work

A less complex example for a cognitive radio network consists of multi-mode capable APs/RNs supporting IEEE 802.16a for the infrastructure links and IEEE 802.11e or IEEE 802.16e for service provision links to the mobile cognitive radios. The network operates in unlicensed frequencies and is able to support QoS in spite of coexistence, with the help of spectrum-sharing algorithms as discussed below.

A comparable approach is introduced in Buddhikot *et al.* (2005) with the Dynamic Intelligent Management of Spectrum for Ubiquitous Mobile-access Networks (DIMSUMnet) architecture that enables a coordinated real-time access to spectrum. The focus of the DIMSUMnet is limited to the improvement of the access to spectrum in time, frequency and space – different from an opportunistic spectrum usage as introduced below. The authors also envisage the advantages of relay-based systems enabling coordinated spectrum reuse and therefore introduce a DIMSUM-RelayCluster architecture.

In the field of Internet research, cognitive networks are discussed as a promising new architectural principle in Clark *et al.* (2003). There, the idea of a knowledge plane is suggested that is separated from the control and data plane. This knowledge plane enables the cognitive network to autonomously identify a problem and fix it with self-configuration. Based on this, Mähönen (2004) and Mähönen *et al.* (2004) suggest taking wireless networks and telephony into account from the beginning, when designing a knowledge plane for cognitive radio networks. In the context of end-2-end reconfigurability the knowledge plane is part of the distributed decision making of a reconfiguration management plane (Alonistioti *et al.*, 2004).

13.3 Spectrum Sharing and Flexible Spectrum Access

The DARPA XG Program and the Integrated Project E^2R are working on flexible and dynamic spectrum usage and related impacts on spectrum regulation. Regardless of the regulatory model, flexibility and efficiency need to be reflected in spectrum access. Techniques that sense and adapt to the radio environment are for instance essentially required in unlicensed bands or when improving spectrum access through methods such as secondary markets. Spectrum sharing therefore plays, especially in the open spectrum context, an important role in increasing spectrum utilization. In the following we differ between primary and secondary users of spectrum, while secondary users defer to primary users in utilizing spectrum.

13.3.1 Spectrum Trading

Currently, spectrum is a scarce and valuable source. The auctions of 3G spectrum licenses in Germany and the UK have shown that operators pay high values for the rights to use this spectrum. However, these high prices reflected the existing spirit in wireless communication industry and the stock markets. Economic development in wireless communication in the recent years indicates that such high prices will never be reached again, as the return of investment in 3G spectrum will take a long time. This questions the auctioning of spectrum especially when taking into account that the way the spectrum was sold artificially increased the prices (Valletti, 2001). The prohibition of using other spectrum with 3G technologies on the one hand and the limited quantity of spectrum on the other hand while government-owned bidders participate at the auction lead to prices not reflecting the de-facto market value of spectrum. Therefore new ways of handling spectrum are required.

Spectrum can be subdivided into quantities in many domains (e.g. frequency, time, code). Trading of spectrum refers to the transfer of spectrum usage rights for a certain quantity of spectrum, usually specified frequency bands. Comparable to trading with securities, spectrum trading is done at two markets. The initial issuing of spectrum through regulation bodies is done at the primary market, while the transfer of spectrum usage rights between different parties is done at the secondary market. Contrary to licensed spectrum, spectrum trading implies that the owner of spectrum usage rights may differ from the user of spectrum. Additionally as a third party, spectrum traders may deal with spectrum usage rights. Spectrum can be subdivided in many dimensions to be resold by the license holder to different parties.

The trading with spectrum increases efficiency of spectrum usage and is therefore attractive for regulation: The spectrum owner, such as a spectrum broker, has a financial interest to promote spectrum efficiency and innovation (Peha, 2000). Spectrum trading at the secondary market has impacts on regulation: the owner and the user of spectrum are different parties. In 2004, the FCC therefore made secondary markets for trading spectrum legal, allowing the licensee for the duration of its license to lease the rights to use spectrum (FCC, 2004b). Depending on the selected option for transferring usage rights the regulator might intervene in the pricing mechanisms to care for the "public interest".

Valletti (2001) recommends spectrum trading based on fully flexible transferable rights, which may be traded. The regulator intervenes in the case of market failures under the presumption that usually the market will solve problems on its own. The enforcement of antitrust laws and arbitration in the case of disputes are the task of the regulator. Additionally, Valletti (2001) suggests that the regulator licenses low power equipment for open access spectrum. Bidding credits are designated for social concerns and public safety services.

Spectrum usage rights can be transferred in different way analogous to other assets (economic goods):

- Lease – The right to use spectrum is temporarily transferred while the ownership remains with the license owner. The license holder can define its own rules in addition to regulation for using its spectrum. This option includes for instance spectrum brokers that lease spectrum in determining prices on the basis of auctions.
- Sale – The ownership of usage rights is permanently transferred.
- Options and futures – The right to access spectrum to a certain future point of time for a predefined duration is transferred for a pre-agreed price. Comparable to the financial markets, the trading with options and features can lead to complex constructs. Financial contracts will emerge to provide security against movement of spectrum prices, i.e. financial payments for reallocating risks.

A recent study by Analysys Consulting (2004) recommends a common overall approach to spectrum trading and liberalization of spectrum usage. A detailed implementation of spectrum trading framework is demanded. This refers to the creation of tradable rights and the establishment of an adequate forum for trading/transfer and managing rights. Additionally an interference management as well as the possibility of defining temporal usage rights and reclaiming them is suggested. Allowing flexible trading mechanisms lets the market decide about the success of the approach to transferring usage rights as introduced above. At the same time liberalization of spectrum is required to abolish restrictions on technologies and services associated with spectrum usage rights. Analysys Consulting (2004) also refers to this on existing spectrum licenses and suggests the reconfiguration as aggregation or partitioning of existing usage rights. A partitioning of usage rights is for instance possible depending on time, frequency or geographical location.

13.3.2 Underlay and Overlay Spectrum Sharing

Open access to most of the radio spectrum, even spectrum licensed for a dedicated technology, is today only permitted by radio regulation authorities for radio systems with minimal transmission powers in a so-called underlay sharing approach as illustrated in Figure 13.3. The simultaneous uncoordinated usage of spectrum in the time and frequency domains uses techniques to spread the emitted signal over a large band of spectrum so that the undesired signal power seen by the incumbent licensed radio devices is below a designated threshold. Spread spectrum, Multi-Band OFDM or Ultra-Wide Band as introduced below are examples for such techniques. To reduce potential interference, the transmission power is strictly limited in underlay spectrum sharing.

The Spectrum Policy Task Group of the FCC (2003b) suggested the Interference Temperature Concept for underlay spectrum sharing to allow low power transmissions in licensed (used) bands. The FCC suggests allowing secondary usage of shared spectrum if the interference cased by a device is below a sufficient threshold. The FCC sees a well-defined space between the original noise floor and the licensed signal of the incumbent radios, identified as “new opportunities for spectrum use” (2002, 2003b) and illustrated in Figure 13.4. This space refers to the power level of the signals at the receiver in a specific band at a geographic location.

Only a very small fraction of the radio spectrum is openly available as frequency band for unlicensed operation. Nevertheless these bands have enabled immense economic success of wireless technologies such as the popular WLAN IEEE 802.11. On the other hand, the actual availability of new spectrum is a seemingly intractable problem. Cognitive radios use flexible spectrum access techniques for identifying underutilized spectrum and to avoid harmful interference to other radios using the same spectrum. Such an opportunistic spectrum access to underutilized spectrum, whether or not the frequency is assigned to licensed, primary services, is referred as overlay spectrum sharing.

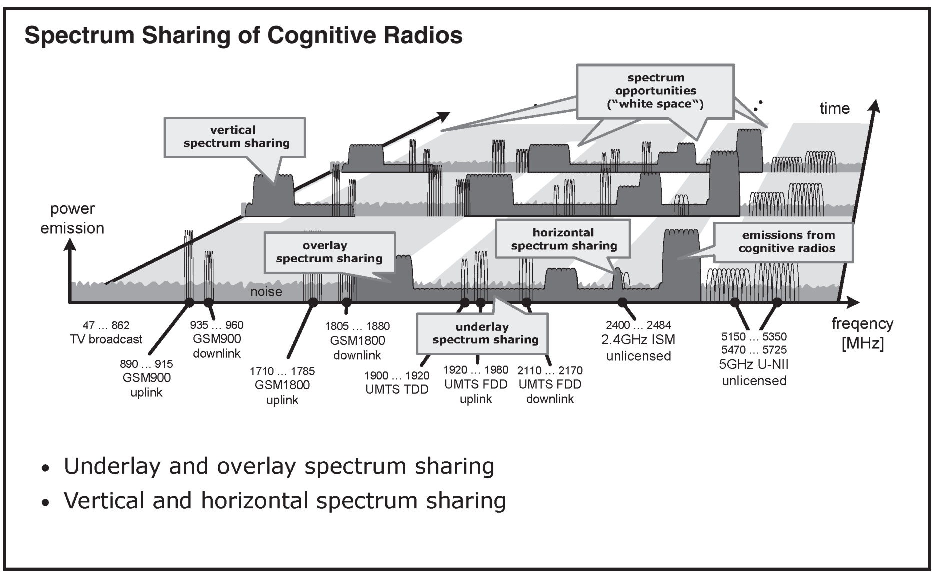


Figure 13.3 Underlay and overlay spectrum sharing of a frequency agile cognitive radio using spectrum on an opportunistic basis (Mangold *et al.*, 2005b).

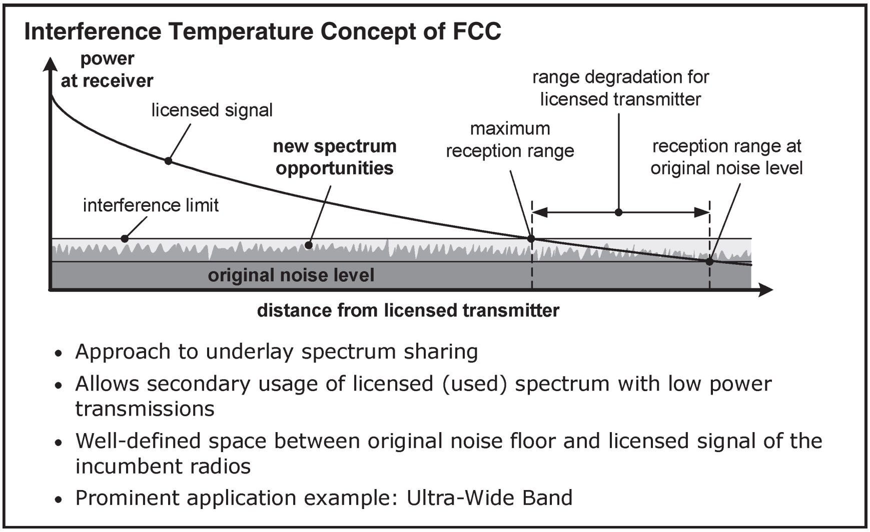


Figure 13.4 Underlay spectrum sharing corresponding to the interference temperature concept of the FCC (2003b).

Overlay sharing obviously requires new protocols and algorithms for spectrum sharing. Additionally, spectrum regulation is impacted, especially in the case of vertical spectrum sharing as introduced below. The operation of licensed radios systems must not be interfered when identifying spectrum opportunities and during operation in licensed spectrum. Dynamic Frequency Selection is a simple example for how unlicensed spectrum users (IEEE 802.11a) share spectrum with incumbent licensed users (radar stations) using overlay sharing.

13.3.2.1 Opportunistic Spectrum Usage

Underutilized spectrum is in the following referred to as spectrum opportunity. The terms “white spectrum” and “spectrum hole” can be used equivalently. To use spectrum opportunities with overlay sharing, cognitive radios adopt their transmission schemes such that they fit into the identified spectrum usage patterns, as illustrated in Figure 13.5. Thus spectrum opportunities have to be identified in a reliable way and their usage requires a distributed coordination. A spectrum opportunity is defined by location, time, frequency and transmission power. It is a radio resource that is either not used by licensed radio devices, or used with predictable patterns, such that idle intervals can be detected and reliably predicted. The accurate identification of spectrum opportunities is a challenge, as it depends on the predictability and the dynamic nature of spectrum usage. The less frequent and more predictable the spectrum usage by primary radio devices occurs, the higher the success of identification, and efficiency of opportunistic usage by cognitive radios. Therefore less frequent and more predictable spectrum usage can be regarded as a contribution to cooperation as introduced in Section 13.3.4.

Characteristic patterns in spectrum usage and characteristic signal features of the radio signals transmitted by primary radio systems facilitate an improvement of spectrum opportunity identification. Different spectrum usage patterns and their classification as spectrum opportunity are shown in Figure 13.5. In this figure, time is progressing from bottom to top, frequency increases

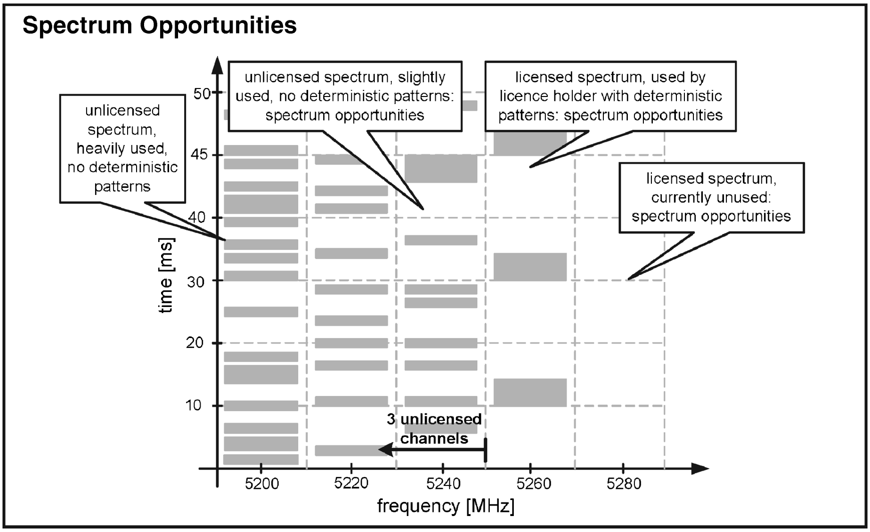


Figure 13.5 Spectrum usage example at 5 GHz. Three 802.11a channels and frequencies are depicted. A cognitive radio identifies spectrum opportunities.

from right to left. Here we see patterns of random IEEE 802.11a spectrum usage (in three channels of the unlicensed 5 GHz band and the frequencies above), in parallel to a predictable, deterministic spectrum usage.

13.3.2.2 IEEE 802.11k

A new type of measurement for improving spectrum opportunity identification is developed in the standardization group of IEEE 802.11k (IEEE, 2005k), which provides means for measurement, reporting, estimation and identification of characteristics of spectrum usage. Spectrum awareness for distributed resource sharing in IEEE802.11e/k is described in Mangold *et al.* (2004a) while radio resource measurements for opportunistic spectrum usage on the basis of 802.11k are analyzed in Mangold *et al.* (2004b). The improvement of confidence in radio resource measurements as an approach to the reliability of spectrum opportunity identification is considered in Mangold and Berlemann (2005).

13.3.3 Vertical and Horizontal Spectrum Sharing

The overlay spectrum sharing with licensed radio systems requires not only fundamental changes in spectrum regulation. Additionally, new algorithms for sharing spectrum are necessary, which reflect the different priorities for spectrum usage of the licensed, i.e., incumbent, and unlicensed radio systems. To reflect this priority, the terms primary and secondary radio systems are sometimes used for the licensed and unlicensed radio systems, respectively.

Cognitive radios will have to share spectrum: (i) either with unlicensed radio systems with limited coexistence capabilities enabling them to operate in spite of some interference from dissimilar radio systems or (ii) with licensed radio systems designed for exclusively using spectrum.

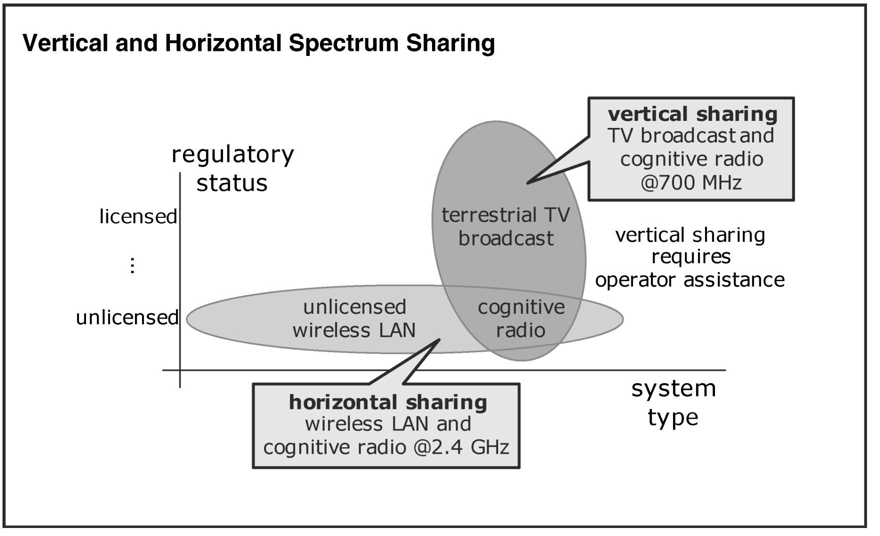


Figure 13.6 Cognitive radios share spectrum with different radio systems. Depending on regulatory status, vertical or horizontal spectrum sharing is performed.

The sharing of licensed spectrum with primary radio systems is referred to as vertical sharing, as indicated in Figure 13.6, and the sharing between equals as for instance in unlicensed bands can be referred to as horizontal sharing. These terms of horizontal and vertical spectrum sharing are first mentioned in Kruys (2003). Another example of horizontal spectrum sharing is the usage of the same spectrum by dissimilar cognitive radios that are not designed to communicate with each other. These dissimilar cognitive radio systems have the same regulatory status, i.e. similar rights to access the spectrum, comparable to the coexistence of devices operating in unlicensed spectrum. Vertical spectrum sharing promises to have the advantage that no lengthy or expensive licensing process and reallocation of spectrum is required.

Vertical and horizontal sharing requires the capability to identify spectrum opportunities as introduced above. Cognitive radios are able to operate without harmful interference in sporadically used licensed spectrum requiring no modifications in the primary radio system. Nevertheless, in order to protect their transmissions, licensed radio systems may assist cognitive radios to identify spectrum opportunities in vertical sharing scenarios. This help is referred to as “operator assistance” in the following.

In horizontal sharing, the cognitive radios autonomously identify opportunities and coordinate their usage with other cognitive radios in a distributed way. To avoid chaotic and unpredictable spectrum usage as in today’s unlicensed bands, advanced approaches such as “spectrum etiquette” are helpful and will be discussed in Section 13.7.

Example: Unlicensed Usage of TV Bands

The technology for terrestrial TV broadcasts is currently digitized. This process will be finalized in the near future (for instance in the US in 2009; in Germany by 2010 at the latest). This digitalization improves the utilization of spectrum, resulting in a reduction of the required spectrum when the number and quality of the TV channels remain unchanged. The usage of the corresponding frequency band is reorganized at the same time in many regulatory domains worldwide. As every broadcast site has to serve a large coverage area, radio transmission is done at high power to guarantee reliable reception throughout the complete coverage area. This implies for many receivers a robustness to interference in the case of proximity to the broadcast site, as the signal is received at a higher power than required. Thus reliable operation is possible even if cognitive radios emit some level of interference. Additionally, TV broadcast sites infrequently change their location and the frequencies they are using, which simplifies identification of spectrum opportunities.

It is therefore envisioned to allow such unlicensed reuse of the entire TV broadcast band for cognitive radios that scan all TV channels throughout the band and operate only upon identification of spectrum opportunities (FCC, 2004c). The working group 802.22 of the IEEE takes this idea and is working towards the standardization of the unlicensed secondary access to TV bands as outlined in Section 13.6.2. Figure 13.7 illustrates this scenario: Two adjacent TV broadcast sites and two independent pairs of communicating cognitive radio devices are shown. The cognitive radios identify locally underutilized spectrum, here unused TV channels, as spectrum opportunities. After some knowledge dissemination and negotiation, the pairs of cognitive radios communicate using these opportunities, while frequently scanning the spectrum for signals from primary radio systems.

Vertical spectrum sharing can be realized in different ways: A beacon signal or busy tone at a foreseen dedicated frequency for signaling permission and/or prohibition of secondary operation in licensed spectrum is one approach to vertical spectrum sharing. A licensee may sell temporarily underutilized spectrum for secondary usage to increase its revenue. Mechanisms therefore are introduced in Section 13.3.1 about spectrum trading. More complex approaches to vertical sharing such as a common control channel or the policy-based secondary spectrum usage on the basis of spectrum observation are introduced in Section 13.6.

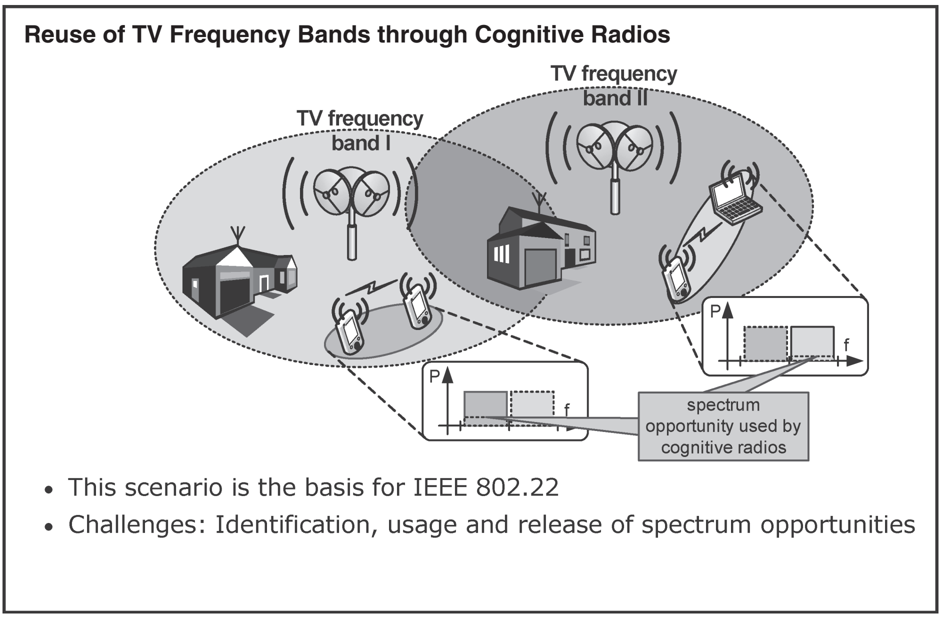


Figure 13.7 Cognitive radios operating in frequency bands of TV and radio broadcasts. At different locations the cognitive radios identify different frequencies as unused and regard them as spectrum opportunities.

13.3.4 Coexistence, Coordination and Cooperation

In the literature about spectrum sharing, the terms coexistence, coordination and cooperation are often used in different ways. As these terms are especially important in the context of QoS support, definitions are given in the following.

Means for coexistence target at interference avoidance in a distributed communication environment. Therefore no communication among coexisting devices is required and possible. In the case of a less utilized shared spectrum, coexistence capabilities suffice to enable a reliable communication. In recent years, coexistence is very successful but a victim of its own success. Under severe competition for accessing the shared spectrum, no QoS support is possible due to the missing coordination among the coexisting, often dissimilar radio systems like Wi-Fi and Bluetooth. Today's implemented approaches to coexistence have limited interference prevention and their spectrum utilization is very ineffective as coexistence implies no incentive to conserve spectrum.

Mutual coordination, either centralized or decentralized, is required in spectrum sharing to enable the support of QoS. QoS support refers in this context to the exclusive usage of spectrum to a predictable point of time for a certain duration.

Under cooperation, altruistic devices delimitate their spectrum usage and carry each other's traffic in the hope of gaining a potential cooperation when all radios participate. Cooperation comes along with the danger of being exploited by selfish, myopic radios resulting in a disadvantage for the cooperating radios. Cooperation is required in building one self-configuring network of mutually coordinated radios in a distributed communication environment. The usage of deterministic patterns when allocating spectrum can be regarded as cooperation. These deterministic patterns help to increase accuracy for other radios for identification of spectrum opportunities and enable a distributed coordination on the basis of observation.

In distributed environments, cooperation can be created and enforced through protocols, either as part of a standard or realized as spectrum sharing etiquette. The enforcement of cooperation is difficult for regulation authorities but may be easier for license holders.

13.4 Coexistence-based Spectrum Sharing

The amount of flexibility allowed in designing access protocols imposed by regulating authorities for the possibility of radios to coexist, share spectrum or even interoperate for coordination (Peha, 2000) defines therefore "Rules of Coexistence" considering the grade of flexibility from the least to the most restrictive:

- Spectrum usage is not constrained. This approach has the advantage of stimulating innovations but the big disadvantage of being only applicable in scenarios where less spectrum utilization can be expected. No QoS guarantee can be given and efficient spectrum usage is not encouraged.
- Constraining spectrum usage is limited to the transmission power. A maximum power level for radio emissions is defined. Finding the adequate level may be difficult and depends on the application scenarios of the envisaged radios operating in these frequencies. Limiting transmission power increases the reuse of spectrum but more radios are required for covering a certain area.
- Constraints on access protocols are imposed, which require no communication among coexisting devices. DFS in the U-NII band, as introduced below, is an example of this. Suitably

defined etiquettes for spectrum usage can facilitate the support of QoS, increase spectrum efficiency and add fairness to spectrum sharing. These etiquettes enable distributed coordination and require cooperation as introduced above. To design such etiquettes is a complex challenge and is introduced in Sections 13.5 and 13.6.

- A minimal standard is required for operation. A common signaling channel used by all radios operating in a shared frequency band is an example of this. The common signaling channel enables mutual coordination and thus increases the level of QoS.
- The interoperation of dissimilar radio networks is part of their standard. The Central Controller Hybrid Coordinator (CCHC) concept for interoperation of IEEE 802.11 and H/2 (Mangold *et al.*, 2001b) or the Base Station Hybrid Coordinator (BSHC) concept for integrating IEEE 802.11(e) in the frame structure of IEEE 802.16 (WiMAX) as outlined in Section 8.3.2 are examples for this.

In the following section several approaches to coexistence are introduced that do not require communication among spectrum sharing radios.

13.4.1 Dynamic Frequency Selection

The FCC (2003a) demands that devices operating in the U-NII band use Dynamic Frequency Selection (DFS) to protect radar systems from interference. DFS lets the transmitter dynamically switch to another channel whenever a certain condition is met. This condition is for instance a threshold level such as $-62\,\text{dBm}$ for devices with a maximum EIRP less than 200 mW. Once a radio signal has been detected, the channel must not be utilized for a certain period of time. Further, before initiating transmission a DFS using device senses the available spectrum for unused spectrum and accesses only these channels. The FCC requires that the DFS aims at a uniform spreading of load over the available channels. Additionally, a continuous monitoring of spectrum during operation is demanded. The possibilities and limitation of DFS for cognitive radios and policies are discussed in Horne (2003). For fundamental details on DFS and radio networks using DFS see Walke (2002).

13.4.2 Transmit Power Control

Transmit Power Control (TPC) is a mechanism that adapts the transmission power of a radio to certain conditions, e.g. a command signal from a communication target when the received signal strength falls below a predefined threshold. Corresponding to FCC (2003a) in the U-NII band a radio reduces its transmission power by 6 dB when TPC is triggered. Only radios operating at power levels higher than 500 mW require TPC in the U-NII band.

13.4.3 Ultra-Wide Band

Ultra-Wide Band (UWB) enables underlay spectrum sharing. In short, UWB is a transmission technique of pulses with a very short time duration using a very large frequency band of spectrum. UWB is discussed in detail in the context of WPANs in Chapter 6. Contrary to many other radio frequency communication techniques, UWB does not use RF carriers.

Instead UWB uses modulated high frequency pulses of low power with a duration less than one nanosecond. From the perspective of other communication systems the UWB transmissions are part of the low-power background noise. Therefore UWB promises to enable the usage of licensed spectrum without harmful interference to primary communication systems. UWB is a prominent example of the Interference Temperature Concept of the FCC introduced above.

The UWB coexistence and UWB based cognitive radios are discussed in Lansford (2004). General methods of UWB are explored to use time, frequency, power, space and coding for eliminating interference to other devices or to achieve graceful deterministic degradation. The application of UWB is limited to short range communication, due to the strictly restricted transmission power. The interference of UWB radios to incumbent radio systems is currently under research. Based on the current FCC mask for UWB emission, Pfletschinger *et al.* (2005) indicate major problems due to co-channel interference between UWB and B3G devices in indoor scenarios.

13.4.4 IEEE 802.16.2

Coexistence in Fixed Broadband Wireless Access (FBWA) systems of IEEE 802.16 operating in licensed bands is standardized in the Task Group IEEE 802.16.2. The activities of this Task Group are introduced in Section 7.3.4.

13.4.5 IEEE 802.16h

The IEEE 802.16h License-Exempt Task Group is developing improved mechanisms for enabling coexistence among license-exempt systems based on IEEE 802.16 as introduced in Section 7.3.7.

13.4.6 IEEE 802.19

The IEEE 802.19 working group (or the Coexistence Technical Advisory Group, 802.19) is aiming at the development and maintenance of policies defining the responsibilities of IEEE standardization efforts to consider coexistence with existing standards and other standards under development. If demanded 802.19 evaluates the conformance of the developed standard to the coexistence policies and offers a documentation of the coexistence capabilities to the public.

13.5 Coordination-based Horizontal Spectrum Sharing

13.5.1 Common Spectrum Coordination Channel

One widely used applied approach in research to spectrum sharing is the usage of a Common Spectrum Coordination Channel (CSCC). The basic idea of CSCC is to standardize a simple common protocol for periodically signaling radio and service parameters, in Raychaudhuri and Jing (2003) this is regarded as a spectrum etiquette mechanism. The CSCC enables coordination through mutual observability between different neighboring radio devices via a simple common protocol. As shown in Raychaudhuri and Jing (2003) using the contention example of 802.11b and Bluetooth devices, the CSCC approach impacts the complete protocol stack (e.g., physical layer, MAC layer, packet formats).

In general, a common control channel as part of the shared spectrum is highly vulnerable to interference so that the complete network might be disrupted though deliberate jamming or through coexisting radios operating in the same spectrum.

A permanently available signaling channel shared by several networks is also suggested in Hunold *et al.* (2000). A Network Access and Connectivity Channel (NACCH) is introduced to enable communication between different networks. In this way they form a larger network where users have universal access and roaming support. This approach extends the exchange of control information between networks for coordinating spectrum usage with the aspect of transferring user data among networks over a common radio channel.

13.5.2 Dynamic Spectrum Allocation

In Dynamic Spectrum Allocation (DSA), spectrum allocations are changed over time depending on network loads in assigning continuous spectrum quantities to different Radio Access Networks (RANs) (Leaves *et al.*, 2004). DSA aims at the exploitation of spatial and temporal variations to traffic loads in RANs. In the frequency domain for instance, a flexible guard band separating two adjacent frequency bands used by different RANs can be adapted to shift bandwidth between these RANs. In DSA, an exclusive usage of spectrum by one operator using one RAN is assumed.

13.5.2.1 Brokerage-based Spectrum Sharing

Spectrum can be regarded as an economic good, which is traded by a broker as introduced above. Auctions are an efficient way to determine the value of spectrum among many parties. An auction can be regarded as a partial information game in which the real valuation a bidder gives to spectrum is hidden to the broker and the other bidders. Auction theory is a very well-developed field of research in economics as well as in wireless communication. Courcoubetis and Webber (2003) give a technology-independent introduction and some theoretical results of spectrum auctioning.

The different treatments of spectrum in the context of a brokerage-based spectrum sharing lead to dissimilar approaches. The auctioning mechanisms introduced in Maheswaran and Basar (2003) consider spectrum as a public resource. Price and demand functions characterize the optimal response functions of the bidders leading to a unique Nash equilibrium for an arbitrary number of agents with heterogeneous quasilinear utilities. Contrary to Grandblaise *et al.* (2005) where another business model is applied, spectrum sharing between different operators providing their spectrum to a common spectrum pool is discussed. An operator temporarily leases spectrum from a broker who controls this pool of spectrum and performs the auctions. An analytical investigation of the problem and solution approach is described in Rodriguez *et al.* (2005): An operator of a CDMA cell populated by delay-tolerant terminals operating at various data rates, tries to optimize its revenue given an amount of spectrum depending on its own pricing policies. Brokerage-based spectrum sharing implies substantial periodical signaling depending on the number of participating parties. The automatic bidding through agents located in the MAC layer is also discussed in Kloeck *et al.* (2005) for an OFDMA/TDD system such as IEEE 802.16. Piggybacked multi-unit, sealed-bid auctions are suggested, which take heavy time and signaling constraints into account.

It can be questioned in general, if a QoS guarantee can be given when using spectrum based on auctions. Therefore, a customer-interested operator might not be willing to provide its spectrum to the pool even if in the short term the selling of its capacity leads to higher revenues.

13.5.2.2 Inter-operator Spectrum Sharing

Another approach to DSA is introduced in Pereirasamy *et al.* (2005) as Dynamic Inter-Operator Spectrum Sharing. Here, based on spectrum shared in UTRA FDD, each operator deploys its own independent radio access network. This work initiated in Pereirasamy *et al.* (2004) where a shared UMTS FDD RAN is assumed. From the regulatory perspective, Inter-Operator Spectrum Sharing is limited to one RAT and a frequency band licensed for use by this single RAT is dynamically divided under several operators. The consideration of a synchronous inter-operator system requires heavy cooperation, which is very unlikely for competing operators. Further it has been shown in Pereirasamy *et al.* (2005) that only partial spectrum sharing leads to favorable capacity gains, which depend heavily on network parameters (e.g., cell radios, TX power).

13.5.3 IEEE 802.11y

The IEEE 802.11y is the youngest Task Group of 802.11 and is working towards a standard for contention-based protocols operating in the 50 MHz frequency band at 3.650–3.700 GHz. The frequency band was opened to unlicensed services in FCC (2004a). A contention-based listen-before-talk protocol is demanded to operate in this frequency band. Therefore, the FCC defines (2005b) a contention-based protocol as:

> A protocol that allows multiple users to share the same spectrum by defining the events that must occur when two or more transmitters attempt to simultaneously access the same channel and establishing rules by which a transmitter provides reasonable opportunities for other transmitters to operate. Such a protocol may consist of procedures for initiating new transmissions, procedures for determining the state of the channel (available or unavailable), and procedures for managing retransmissions in the event of a busy channel.

Some of these characteristics are satisfied by currently available 802.11a systems operating in the U-NII band at 5 GHz: They are frequency agile, have the ability to sense signals from neighboring transmitters, offer adaptive modulation and use transmit power control. 802.11y aims at the definition of one extension to 802.11 OFDM, such that fixed and wireless devices can operate in the 3650–3700 MHz band.

13.5.4 Spectrum Sharing Games

In research, the application of solution concepts derived from game theory for the analysis and realization of spectrum sharing has many facets. The corresponding game models and their characteristics differ essentially, depending on the particular application scenario. The horizontal spectrum sharing of equally righted cognitive radios is suitable for modeling as a game of interacting players. Contrary the vertical spectrum sharing: Primary radio systems neither have nor need interaction capabilities and are not allowed to be interfered.

Potential game models in which networks of cognitive radios may alter transmitted energy and signature waveform are analyzed in Neel *et al.* (2002a) and Neel *et al.* (2002b). Specific conditions are delineated for which the models apply. These models are used to identify steady-state conditions of these networks in the presence of cognitive radios making myopic decisions as an initial step in network planning. Different convergence dynamics in cognitive radio networks are examined in Neel *et al.* (2004). Several distributed power control algorithms are examined in game models of different complexity levels. The convergence process and the steady-state behavior of cognitive radio networks are analyzed in modifying the objective functions to prevent the application of adaptation algorithms that result into suboptimal game outcomes.

The selfish behavior of nodes in CSMA/CA networks is investigated in Cagalj *et al.* (2005). In CSMA/CA the protocols rely on random deferment of packet transmission. By applying a model of dynamic games, the conditions for a stable and optimal operation in the presence of selfish nodes are analyzed.

The channel assignment in a Wi-Fi network is modeled as a game in Halldorsson *et al.* (2004), in which the players are service providers or APs. Stable points of operation are identified with the solutions to a maximal coloring problem in an appropriate graph. A “price of anarchy” is introduced as a value for the confession to the distributed channel assignment in comparison to a centralized one. In the case of easily realizable bargaining procedures, this “price of anarchy” is bounded to a constant.

A game theoretic formulation of the adaptive channel allocation problem for cognitive radios resulting into rules for spectrum etiquette is discussed in Nie and Comaniciu (2005). Cognitive radios measure the local interference temperature on different frequencies and may adapt their transmission rate to channel quality (in using adaptive channel coding) or switch to a different frequency channel. It is shown that the corresponding game formulation converges to a deterministic channel allocation scheme. Additionally, adaptive protocols and a learning algorithm are designed and their convergence behavior and tradeoffs are discussed.

The competition between independent radio systems for allocating a common shared radio channel can be modeled as a stage-based game: Players, representing radio systems, interact repeatedly in radio resource sharing games, without direct coordination or information exchange. Solution concepts derived from game theory allow the analysis of such models under the microeconomic aspects of welfare as described in detail in Section 9.1 in the context of spectrum etiquette.

13.6 Coordination-based Vertical Spectrum Sharing

All approaches to horizontal spectrum sharing can be used for vertical spectrum sharing when being combined with an accurate identification of spectrum opportunities. When multiple radios regard simultaneously the same spectrum as unused by the incumbent radio system, the access of the secondary radios needs to be coordinated (centralized or distributed) to enable a support of QoS on the one hand and to increase efficiency of spectrum usage on the other hand.

13.6.1 Common Control Channel

In the context of vertical spectrum sharing, Cabric *et al.* (2005) introduce opportunistic spectrum usage creating a “virtual unlicensed band”, also referred to as a spectrum pool, with the help of a common control channel. A hierarchical control channel structuring is suggested as differing between a universal control channel used by all groups for coordination and separate group control channels used by members of a group.

A dedicated control channel located in licensed spectrum is suggested by the DARPA XG Program (DARPA, 2003, 2004a) to enable coordination in shared spectrum. This simple approach has several disadvantages, as a fixed licensed channel is required. The licensing of such a channel is a long-winded and expensive process. The licensed channel has a fixed bandwidth and is therefore limited in scalability.

13.6.2 IEEE 802.22

The IEEE 802.22 working group is targeting the standardization of a cognitive air interface for fixed, point-to-multipoint, Wireless Regional Area Networks (WRANs) of 40 km or more operating on unused channels in the VHF/UHF TV bands between 54 and 862 MHz (IEEE, 2005m). A wireless broadband replacement of DSL and cable modem services in less populated areas with unused VHF/UHF TV bands is envisaged. In not causing harmful interference to the incumbent TV broadcast system, 802.22 realizes vertical spectrum sharing as introduced above and illustrated in Figure 13.7. Areas where wired networks are too expensive to deploy due to sparse population are the focus of IEEE 802.22.

13.6.3 Spectrum Pooling

In Weiss and Jondral (2004) an OFDM-based approach to secondary usage in overlay spectrum sharing is developed referred to as spectrum pooling. In an 802.11-like scenario the spectrum measurements of mobile terminals are gathered centrally by an access point. Unused spectrum of different owners is merged into a common pool optimized for a given application. A licensed system public rental hosts this common spectrum pool and users can temporarily rent spectrum during idle periods of the licensed users. Although Weiss and Jondral (2004) introduce a speeding-up protocol to bypass the MAC layer and just use the physical layer for signaling, an essential weakness of spectrum pooling is not sufficiently mitigated: The central gathering of measurement information takes considerable time and management effort.

13.6.4 Value Orientation

Spectrum sharing among different radio systems can be understood as a scenario forming a society of independent decision-makers (Mangold *et al.*, 2005a; Mangold *et al.*, 2006). Therefore, basic concepts to classify social action that are taken from social science can be applied to define system strategy rules. The rules represent algorithms for decision-making entities (referred to as actors) that reside in the radio systems. For a simple scenario of spectrum sharing, the need for regulation as opposed to voluntary rules is investigated. A spectrum-sharing scenario of a contention-based medium access is analyzed under the assumption that the radio systems do not communicate with each other, but operate using the same radio resources. Voluntary standards and social concepts to mitigate the two main problems of open spectrum access are addressed in the above mentioned publications: incumbent protection and fair coexistence.

13.6.5 Spectrum Load Smoothing

The application of waterfilling in the time domain enables a decentralized and coordinated, opportunistic usage of the spectrum. This is referred to as Spectrum Load Smoothing (SLS)

(Berlemann and Walke, 2005; Berlemann, 2006; Berlemann *et al.*, 2006d). With SLS, competing radio systems aim simultaneously at an equal utilization of the spectrum. In observing the past usage of the radio resource, the radio systems interact and redistribute their allocations of the spectrum under consideration of their individual QoS requirements. Due to the principle of SLS these allocations are redistributed to less utilized or unallocated spectrum. QoS requirements of the coexisting networks are considered. Further, SLS allows an optimized usage of the available spectrum: An operation in radio spectrum, which was originally licensed for other communication systems, is facilitated, as the SLS implicitly achieves usage of unused spectrum and its release if it is needed again.

13.7 Policies and Etiquette in Spectrum Usage[1]

Flexible and dynamic spectrum usage requires an intelligent medium access, especially in the face of QoS support. In this context, policies are required to restrict the dynamic spectrum usage of cognitive radios. A policy is a selection of facts specifying spectrum usage. These facts are interpreted through a reasoning instance, in this chapter referred to as a spectrum navigator. The spectrum navigator is able to consider a flexible amount of different policies realizing a policy-adaptive cognitive radio. Policies, as regulatory rules for spectrum usage, form a framework of behavior for using spectrum. They are mandatory for operation and are enforced by regulation authorities.

Etiquette on the other hand adds fairness and efficiency to spectrum allocation. Etiquette is a multitude of rules that may be voluntarily applied and can be either part of standards or imposed by regulation. Spectrum etiquette is already discussed for existing unlicensed bands in various regulatory bodies and standardization groups.

13.7.1 Policy Framework

Policy-enabled spectrum usage is one of the key features of cognitive radios. The decision taking and learning of a cognitive radio is not limited to policies but has to take many additional factors into account, such as radio capabilities and the environment (outside world, customers). This imposes the need for a formal description framework. Initial steps towards a description language for cognitive radios have been introduced in Mitola and Maguire (1999, 2000) as an ontology of radio knowledge defined in the Radio Knowledge Representation Language (RKRL). An initial step towards a policy framework is the DARPA XG policy language (DARPA, 2004b) which includes an Extendable Markup Language (XML) based policy description language. The SLS from Section 13.6.5 and strategies derived from game theory based on the approach from Section 9.1 are described in the DARPA XG policy language in Berlemann *et al.* (2005d) and Berlemann *et al.* (2006c). The spectrum access of cognitive radios can also be specified in the DARPA XG policy language (Berlemann *et al.*, 2006b).

Policies have their origin in spectrum usage restrictions imposed by a regulating authority. Further policies may come from other policy makers to reflect for instance preferences of the user or operators. The specification of algorithms for enabling spectrum sharing is another important aspect for using policies. The policies might have a limited validity, which depends on multiple factors as for instance the local time, the geographic location of the radio or the country where it is operating. A license holder may also impose policies for using its spectrum by a secondary radio system and might influence the access privileges to spectrum as well. Cognitive radios repeatedly seek for updates of policies (example: once a day) that are relevant for their regulatory domain. The radios that are located in the regulatory domain for which new policies have been published, download the machine-understandable policies and update their local information bases. Alternatively, policies are made available through memory devices such as flash cards, to allow cognitive radios that do not have access to servers to update their information bases. Thus, cognitive radios have to use policies in an adaptive way.

A well-defined policy framework is required to enable such a cognitive radio capable of updating policies. This framework implies language constructs for specifying a policy, a machine-understandable representation of these policies and a reasoning instance, here called a spectrum navigator, which decides about spectrum usage as further outlined below. The policy conformance validation is responsible for downloading, updating and validating policies. The syntactical correctness of a policy that has been downloaded to the cognitive radio is verified. After conformance validation, the cognitive radio translates the policies to a machine-understandable language to enable computation through the spectrum navigator.

13.7.2 Spectrum Navigation

Cognitive radios have a flexible protocol stack and modem part which can be both dynamically adapted to the local communication environment. Additionally, a reconfiguration management is required to fulfill all reconfiguration-related functions. All functions concerning the opportunistic usage of frequency spectrum, i.e., realizing a cognitive medium access, are done by a spectrum navigator as introduced in Figure 13.8. This spectrum navigator is part of the reconfiguration plane (in the case of a completely reconfigurable protocol stack and modem part as for instance considered in E^2R) or located in an “open spectrum mode” (in the case of a multi-mode capable radio of configurable modes, as for instance under discussion in WINNER). The decision about how to allocate which spectrum is taken by the spectrum navigator on the basis of policies. The spectrum navigator identifies spectrum opportunities with the help of frequent measurements of the spectrum usage provided by the protocol stack as for instance under standardization in IEEE 802.11k as introduced above. Means are developed there for measurement, reporting, estimation and identification of the current spectrum usage in the ISM bands. Additionally, the QoS requirements of the supported applications are taken into account together with preferences of the user such as transmission costs. The capabilities of a radio, as for example the frequency range that can be used for transmission, the available PHY modes, coding schemes, the number of transmission units etc. determine which spectrum the navigator selects. The reasoning of the spectrum navigator results in specification of the current spectrum usage and a corresponding configuration of the protocol stack as depicted in Figure 13.8.

13.7.3 Reasoning-based Spectrum Navigation

This section is based on Mangold *et al.* (2005b). The variety of diverse understandings of what “cognitive radio” refers to, often leads to confusion. The many promises of what can be achieved if cognitive radios are employed moreover lead to high expectations about the cognitive radio approach.

We therefore outline in this section the concept of reasoning as one of the core concepts for cognitive radio. This important aspect of cognitive radio is built on the DARPA XG vision (DARPA, 2004a). A cognitive radio is aware of its environment. “Cognition” refers to an act of knowing, being aware, recognizing, judgment, and reasoning. Recent developments in the area

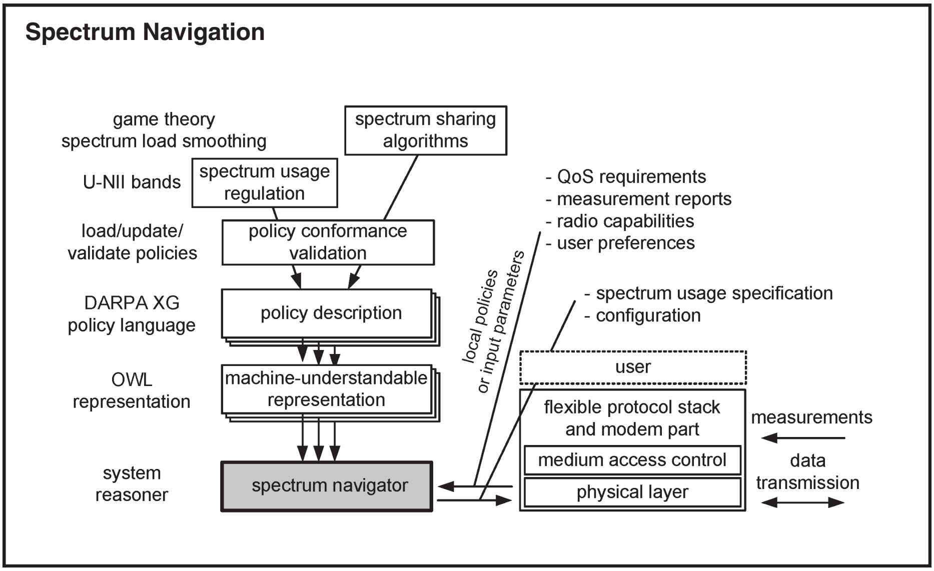


Figure 13.8 Flexible spectrum usage by a cognitive radio. A spectrum navigator takes multiple policies, spectrum usage measurements and additional restrictions into account, when deciding about spectrum allocation. Reproduced by permission of © 2005 IEEE[1].

of machine-learning, the semantic web, and machine-understandable knowledge representation, allow the efficient implementation of a cognitive radio. It is realized in the form of a so-called reasoner, which is introduced above as a spectrum navigator.

13.7.3.1 Reasoning

A reasoner makes the actual decisions on how to share spectrum. A reasoner is a software process that uses a logical system to infer formal conclusions from logical assertions. It is able to formally prove or falsify a hypothesis, and is capable of inferring additional knowledge. The so-called first order predicate logic is the simplest form of a logical system considered to be useful for such a reasoner. As a simple example, a reasoner may be fed with the knowledge ("all cognitive radio devices are capable of operating at frequencies below 3.5 GHz"). A statement ("white space at 2.0 GHz") would enable this reasoner to infer ("spectrum usage permitted at 2.0 GHz").

13.7.3.2 Knowledge Representation

However, inferring statements from other statements, as illustrated in the example, requires a structured and machine-understandable knowledge base for representing knowledge about radio communication. Such a knowledge base has to be constructed by human domain experts, before the machines will be able to interpret, consume, reuse and eventually extend the knowledge. For this, semantics are needed to define truth and valuations: so-called radio semantics. To construct radio semantics is one of the key research problems to be solved. Knowledge should be represented in a machine-understandable way, using languages such as the Web Ontology Language (OWL) (McGuinness and Harmelen, 2003; Smith *et al.*, 2003). OWL is a rich language based on XML, that allows not only first-order logics, but also higher-order, class-based reasoning (Baader *et al.*, 2003).

13.7.3.3 Traceability of Decision Making

Regulation targets at fair and efficient spectrum usage. Therefore the way a cognitive radio makes decisions must be transparent, contrary to today's algorithms for spectrum management. Current radio systems have vendor-specific solutions for spectrum management such as power control and channel selection and are thus not traceable for the public and the regulation bodies. As a result, today's standards and regulation have extreme restrictive parameters such as power levels and frequency ranges for operation, to achieve a minimum level of coexistence, spectrum efficiency and fairness in spectrum access. Due to the scarcity of free accessible spectrum on the one hand and frequently underutilized licensed spectrum on the other hand, spectrum regulation needs a fundamental rethinking towards less restricted spectrum usage. Cognitive radios realize that such weakly constrained radio resource management algorithms impose the requirement of visibility. The entire algorithms for decision making have to be visible to the outside world, and control mechanisms for regulators have to be developed.

13.7.4 Policy-defined Medium Access Control

A software-defined medium access control specified in the DARPA XG Policy Language is proposed in Berlemann *et al.* (2006b). Initial steps are discussed towards the realization of a policy-based MAC at the example of the Enhanced Distributed Channel Access (EDCA) of IEEE 802.11e. This channel access protocol is specified in a machine-understandable policy language, instead of lengthy textual description known from the standard. Such a machine-understandable description of the protocol enables cognitive radios to operate in distributed environments according to the 802.11(e) standard.

13.8 Summary and Conclusion

The different approaches to regulation of spectrum usage which have been introduced in this chapter are summarized in Figure 13.9, taking the aspect of QoS into account.

We refine Jon M. Peha's appropriate "Taxonomy for Spectrum Sharing" (Peha, 2005) in extending it with the aspect of vertical and horizontal spectrum sharing. His understanding of a cooperative meshed network matches the cognitive radio network vision as introduced above. The tables separate regulation options into primary and secondary spectrum usage. A QoS guarantee always requires some degree of exclusiveness. If a guarantee is not required, primary systems may share spectrum. Coexistence is less adequate to support QoS while cooperation increases the level of possible QoS support. Regulation authorities can delegate the control of spectrum access to one or multiple private entities to enable spectrum trading at the secondary market. A so-called spectrum manager inherits the role of the regulator in this context. Secondary usage might be allowed for underlay or overlay spectrum sharing, provided that secondary radio systems defer from spectrum utilization whenever the license-holding primary radios access their spectrum. Secondary radios can try to coexist with primary radios without interfering them. Cooperation

Regulation Options for Primary Spectrum Usage

Regulator controls access	Licensee controls access	Application requirements
Traditional licensing	Spectrum manager makes guarantees	Guaranteed QoS
Unlicensed band, regulator sets etiquette	Spectrum manager sets etiquette, no QoS guarantee	No QoS support, coexistence, horizontal spectrum sharing
Cognitive radio network, regulator sets protocol	Cognitive radio network, licensee sets protocol	QoS support, cooperation, horizontal spectrum sharing

Regulation Options for Secondary Spectrum Usage

Regulator controls access	Primary licensee controls access (secondary market)	Application requirements
Not possible	Licensee guarantees QoS	Guaranteed QoS
Unlicensed underlay with opportunistic access	Secondary market with overlay opportunistic access	No QoS support, coexistence, vertical spectrum sharing
Interruptible secondary operation, regulator sets cooperation protocol	Interruptible secondary operation, regulator sets cooperation protocol	Interruptible QoS support, cooperation, vertical spectrum sharing

Figure 13.9 Regulation options for primary and secondary spectrum usage as refinement of Peha (2005).

between secondary and primary radios enables the secondary radios to support QoS with known interruptions. Secondary radio systems are only able to guarantee QoS if the primary radio systems commit themselves not to interfere. This commitment of the licensee introduces trading of spectrum.

In the short term, commercial broadband and cellular networks require exclusive spectrum access to enable QoS guarantees for the customers. Restricted secondary spectrum usage and spectrum trading are grades of flexibility to increase the overall efficiency of spectrum utilization. The licensing process itself needs to be accelerated and requires more flexibility to reflect the rapid developments in the wireless communication market.

In the long term, spectrum used by future wireless broadband systems covering wide areas will most likely be a combination of exclusively accessed spectrum and unlicensed/open spectrum. The exclusively used spectrum allows to guarantee the customer a minimum level of QoS. The shared spectrum enables an extension of network capacity to provide more services and to increase the number of served customers. Intelligent spectrum sharing algorithms for coordination, as introduced in Sections 13.5 and 13.6 improve the efficiency of spectrum usage and increase the radio networks' capabilities to support QoS while using the shared spectrum. The sharing of a common network infrastructure, as for instance introduced in Beckman and Smith (2005), will further facilitate spectrum sharing.

For indoor, short-range communication at high data rates, such as wireless USB, the advantages of liberalizing spectrum access outweigh its dangers. The cognitive radio approach for flexible spectrum access is ideal for realizing such communication systems: It lies between the two extremes of open and unlicensed spectrum on the one hand, and the "command-and-control" licensing on the other hand. Cognitive radios can be modified to any level of freedom between these two extremes. Due to the locally limited operation of this application scenario, no tragedy of commons exists and freeing access to spectrum will stimulate innovations and economic success. In distributed environments, policy adaptive cognitive radios provide the necessarily required flexibility and intelligence in spectrum access: Local usage constraints are taken into account while etiquettes enable distributed coordination through cooperation.

The self-organization of cognitive radios to a cognitive radio network will further enhance coverage, capacity and QoS in wireless communication. Therefore, a flexible regulatory framework is required enabling less restricted spectrum usage. Operator assistance plays an important role, especially in the field of secondary usage of spectrum and vertical spectrum sharing. Operators may wish to assist in identification of spectrum opportunities and to protect incumbent radio systems. Thus operators might help the cognitive radio network to find an optimal configuration. Cognitive radio networks will not compete but complement the existing cellular networks operating in licensed frequency bands. The development from time-based pricing to service-based revenue models will be further intensified through a further fall in prices for wireless communication.

With cognitive radio, spectrum assignment and licensing will become more dynamic. Greater flexibility in responding to emerging demands of the information society, as well as to market requirements, will be the result.

Note

1. Reproduced by permission of © 2005 IEEE[1]. Source: L. Berlemann, S. Mangold, and B. H. Walke, "Policy-based Reasoning for Spectrum Sharing in Cognitive Radio Networks," in Proc. of *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN2005*, Baltimore MD, USA, 8-11 November 2005. (Section 13.2 (in parts), Fig. 13.8, Section 13.7).

14

Conclusions

Bernhard H. Walke, Lars Berlemann and Stefan Mangold

The short history of wireless systems standardization presented in the Preface and the description in the Introduction of current actors in standardization activities towards future mobile and wireless systems have made it clear: We already have a plethora of different radio system standards in use, Figure 1.4, and the future trend is towards ever higher data rates and capacities of wireless and mobile systems.

In this book, we have presented the most important wireless IEEE 802 standards created during the last five to seven years, and have demonstrated their delay and throughput performance in typical use cases. We have also discussed and proposed improvements to these systems and have addressed spectrum regulatory constraints, currently targeted by various standardization groups, with a focus on the next-generation systems. It is felt to have been a rare opportunity to have the full range of 802.11 mainstream standards covered by experts at ComNets who have worked towards its development and contributed their experience to this book. This made it possible to apply the same way of presentation and analysis, applied to the different systems, within the limits of the differing wordings used in the 802 Working Groups (WGs).

After reviewing fundamentals of wireless communications (Chapter 2), a detailed introduction to radio transmission and of random versus controlled medium access is given to establish the basis for an understanding of wireless systems' functioning and related problems described in later chapters. Since random access cannot be avoided in any wireless/mobile systems, Medium Access Control (MAC) protocols are compared to underline the importance of well-designed MAC protocols.

Spectrum Regulation

The availability of a sufficient amount of radio spectrum is crucial for the economic success of current and future radio communication systems (Chapter 3). The triumph of wireless technologies operating in unlicensed frequency bands, especially of the 802.11 standards family, indicates the economic advantage of unlicensed spectrum. The scarcity of free available spectrum on

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

the one hand and underutilized licensed spectrum on the other hand leads to a rethinking of spectrum regulation in general. Although national regulation authorities have signaled readiness to liberalize regulation of radio spectrum, the concrete realization of spectrum liberalization has been observed to substantially differ when comparing the approaches and public statements of national regulation authorities.

The reliable operation of wireless communication systems in unlicensed frequency bands is a main focus of this book and has been discussed in the context of coexistence, interworking and spectrum sharing of 802-based systems (Chapters 5 to 8). The potential and current trends in the field of intelligent radio communication systems capable of efficiently using spectrum (either unlicensed or underutilized licensed spectrum) are discussed under the umbrella of cognitive radio (Chapter 13).

Basics of Mesh Networks

Wireless mesh networks (Chapter 4) have been found to be a main driver in wireless systems development. Mesh networks are based on multi-hop relaying of data frames and therefore need to cope with problems such as hidden and exposed stations, capture effects and congestion at bottleneck nodes. Mesh networks need highly efficient routing protocols, to cope with suddenly appearing obstacles, too short transmission ranges and link breakage owing to mobile devices. Fast rerouting and route reestablishment appear to be challenges. State-of-the-art routing protocols providing multiple routes between each source–destination device pair to reduce impact of link breakage through fast and seamless switching to an alternate route have been found to be beneficial combined with the use of cross-layer information exchange between MAC and IP layers.

802.11 Wireless Local Area Networks

The impressively large family of WLAN related 802.11 standards (Chapter 5), and the more mature options are found to be very complex systems. As said in the Preface, the standard of the first days (IEEE, 1997) specified a very robust and low complexity system, starting from 1 Mbit/s capacity operating at 2.4 GHz, stepwise was improved to 2 Mbit/s and in 2002/3 to between 6 and 54 Mbit/s peak data rate (802.11g). Besides that, functionality towards a secure, radio interference aware and QoS supporting system was developed, thereby adopting the initial properties of HiperLAN Type 2 (H/2). The 802.11e option can be viewed as the closest possible clone of H/2 within the limitations of a CSMA/CA based MAC protocol, without sacrificing downwards compatibility to "legacy" 802.11 stations following the rules of the early standard: A contention-free period is part of the MAC protocol to allow access to the wireless medium under central control, based on polling, to become the key element for supporting QoS of real-time applications. Downwards compatibility to legacy stations made the advanced 802.11 protocol options much more complex than the H/2 MAC protocol. In the early days of WLANs, 802.11 conquered H/2 mainly because of much lower complexity, today it is much more complex but still inferior in performance (Walke, 2002, Chapter 14.2.9). As a consequence, 802.16 WMAN, from the start, followed the path outlined by H/2 to become what it is today.

The 802.11 protocol, which reserves the channel per data frame (MSDU) in a preceding random access phase, was found to be inefficient under short frame length with a technology driven increasing air interface data rate, most data frames become short, questioning the future-proofness of 802.11 systems for high-speed WLANs – most recently, a "quarter Gigabit per second" air interface 802.11n has been specified and the "one Gbit/s" air interface is expected to come soon.

Accordingly, medium reservations after successful random access for a longer time duration, namely, for a number of frames and related block acknowledgment were introduced in 802.11. In addition, the Direct Link option of H/2 is resurrected in 802.11 to save overhead, if neighbor stations communicate with each other under central control. Further, use of subchannels out of a wideband OFDM channel under OFDMA access control and use of Multi-Carrier CDMA are under discussion for 802.11 systems (Orfanos *et al.*, 2005) to artificially extend the length of an MSDU of given data content through transmitting it on a more narrow bandwidth channel, to enlarge transmit duration of data frames and the overhead-to-useful-data-rate relationship.

Measures for coexistence support of adjacent 802.11 systems' operation have been introduced in 802.11k, allowing stations to report measurement results, channel load, noise histogram and beacon reports, frame and hidden station reports, etc. These functions contribute to enable 802.11 systems to not only closely work adjacent or overlapping to each other but also to establish service areas much larger than a hotspot, with many APs cooperating in a harmonized way.

The different types of 802 wireless systems, namely WLAN, WPAN and WMAN, are expected to benefit greatly from extending the range of radio coverage of an AP: Relaying introduced as a function to a subset of stations and meshing to connect stations and APs with each other on a multi-hop route are seen as promising future technologies.

802.15 Wireless Personal Area Networks

Starting with IEEE Bluetooth standardization in 1999 (IEEE, 2002c), the WPAN WG 802.15 soon split its activities into two application areas of short-range device communication: Low data rate and high-speed WPAN. Short-range communication allows for extremely high data rates (480 to 1.300 Mbit/s over a few meters is one aim) and this is the main differentiator to WLANs. It is fascinating to observe the manifold technical solutions for PHY and MAC that compete in the standardization process, temporarily blocking any progress. Apparently, the expected market for applications such as wireless USB is seen to be huge, creating much competition, and thus quite different from the early days of WLAN standardization. It is clear how UWB and broadband OFDM based systems differ and what their pros and cons are. It appears that high-speed WPAN standards will need time to emerge, instead, industry consortia have formed to gather forces for bringing through their preferential system solution. Multi-hop and Mesh, again, are big hopes and challenges for these forthcoming systems.

802.16 Wireless Metropolitan Area Networks

Fixed (Broadband) Wireless Access (F(B)WA) Point-to-Multipoint (PMT) technology until recently was based on proprietary solutions and, although known to be competitive to wired access in the local loop of telecommunications systems such as Digital Subscriber Line (DSL), cable modems and even fiber optic cables, appeared to be applied to telecommunications applications in rural countries. Since 2001/2, ETSI BRAN and IEEE started to develop their standards HiperMAN and 802.16, respectively, aiming at mass production of low-cost systems for F(B)WA. Soon after, it became clear that connecting 802.11 APs to the wired backbone network could be a second enabler of 802.16 systems and meanwhile, in 2005, use of the system for cellular broadband to compete to 3rd and 3.5th generation mobile systems appeared attractive and resulted in an option 802.16e. Currently (2006), a Mobile Multi-hop Relay (MMR) Task Group (TG) was formed (802.16j) to specify a cellular system supported by fixed and mobile Subscriber Stations acting as layer-2 relays. The dynamics of these developments was a surprise to cellular infrastructure manufacturers and there is still a debate whether the regulator should permit operation of 802.16 mobile cellular in frequency bands licensed for cellular operators, especially, in 3rd generation systems extension bands, according to ITU-R/WRC regulations.

Although there are three PHYs standardized, the MAC and its extensions were developed and accepted without great debate, partly harmonized with ETSI/BRAN HiperMAN. As a result, some basic system standards are already mature at the end of 2005, and ready for products in the fixed wireless domain. According to the authors' view, 802.16 as a system able to support QoS of applications, will enter not only the licensed bands for the different purposes mentioned, but will also enter the unlicensed bands to stepwise replace 802.11 systems. For this purpose, as described in detail in Chapter 8, it needs to be able to also support 802.11 functions, especially PCF, to silence 802.11 stations whenever necessary. A migration from 802.11 operation in licensed bands to higher quality 802.16 operation would then be possible.

The capacity of the IEEE 802.16 system useful for the user in typical scenarios is found to be 90% of the overall capacity, underlining the low (10%) overhead typical for that system. The systems appears applicable also under NLOS radio propagation conditions to meet the expectation, especially if smart antennas and beamforming are introduced as foreseen in the standard.

Mesh for 802.11, 802.15 and 802.16

The approaches to mesh discussed in the book are not yet part of the 802 standards. What makes these approaches attractive is that they are standard compliant and can be realized on top of the existing standards. Especially, the proposals based on central control only require some MAC modifications in the BS/AP, so that company-specific, proprietary implementations seem to be promising for the implementation of mesh networks.

Current 802 standardization tends to not extend MAC protocols for multi-hop operation. The solutions introduced and evaluated demonstrate the feasibility of multi-hop operation in homogeneous as well as heterogeneous 802.11 and 802.16 wireless networks. The centrally controlled MAC frame-based protocol of 802.16 and the HCCA of 802.11e are clearly solid bases for the introduction of efficient multi-hop networks. The delay and throughput performance of alternate proposals show that a limited number of hops is acceptable without degradation of QoS, so that even VoIP communication can be realized in highly loaded heterogeneous (802.11 and 802.16 combined) mesh networks.

Coexistence Control of 802 Radio Systems

Transfer of solutions from game theory and social science to the control of wireless systems competing for the same channel appears useful for realizing QoS support in interfering wireless systems.

Results of analysis and simulation indicate that cooperation is an achievable equilibrium, suited to improve the overall spectrum efficiency. Learning in games to substitute lack of information about competing systems and game models of multiple players (central controllers of networks) appear promising research areas for introducing cooperation in cognitive radio-based systems and to mitigate mutual interference of wireless networks sharing a channel.

Additionally, mechanisms for enabling operation of IEEE 802.16 in spectrum shared with 802.11(a) have been proposed. Coexistence is enabled in partly blocking 802.11(a) out of the medium. This enables a guarantee of QoS in 802.16. A drawback of operating 802.16 in unlicensed frequency bands shared with 802.11 is identified: in the worst case, approximately 20 % of the available transmission time is wasted leading to a reduced efficiency of channel/spectrum usage.

Fixed Relays for Cellular Systems

The trend is to use IEEE 802 systems, especially 802.11 and 802.16, for outdoor and cellular-like deployment to provide Internet access. This is a recent development for WLAN systems. Since its early days, wireless broadband systems suffer from a capacity, substantially decreasing with the distance from the BS/AP. When assuming a constant user density in the area, the number of users increases with distance from the BS/AP inversely to the capacity offered in a cell. The technology trend towards ever higher air interface data rate makes it even worse. Shifting wireless systems upwards in the spectrum where more spectrum is available for high-speed communication results in substantially reduced radio range of BS/AP, leading to higher pathloss, with cells shrinking and cabling cost exploding.

Relays introduced to 802 cells appear well suited to bring broadband to wider areas than possible with single-hop current systems. And, as shown in Chapter 10, are suited to provide more fairness in terms of throughput capacity available throughout the cell area.

Relay enhanced cells differ from mesh networks, in principle, since routes follow tree topologies, with the BS/AP as source and sink of the data flows. Meshing of adjacent relay stations, belonging to different tree topologies, appear to be of advantage to increase the reliability of relay-based networks. Apparently, most of the work done in standardizing mesh networks in IEEE 802 WGs is beneficial for cellular systems, too.

Mutual Integration and Cooperation in Wireless Systems

As visible from Figure 1.4 and explained throughout the book, the number of standardized systems is large and increasing, making interworking of wireless networks an attractive option, even a must. The User Terminal (UT) supported by interworking wireless networks will always be able to connect best in terms of cost, performance and other parameters. Interworking covers loose and tight coupling of networks, and even cooperation for mutual assistance. And the comfort of usage seen by the UT might vary substantially, dependent on the interworking support provided at the current location. The systems involved when addressing interworking range from mobile cellular for high mobile terminals to wireless serving movable terminals that are quasi-fixed when operating. 3GPP and IEEE have active Task Groups working towards close cooperation of systems under their responsibility. WLAN/3G interworking and 802.21 Media Independent Handoff WGs are the main actors, besides IETF concerning IP mobility.

User (personal) mobility enables a user to maintain its identity irrespective of the terminal used and its network point of attachment. Terminals may be of different types. *Service mobility* enables the customer to use a particular (subscribed) service irrespective of the location of the user and the terminal. *Session mobility* enables the user to move an active session between terminals. *Terminal mobility* enables a terminal to change its location, i.e., network point of attachment, without interrupting the communication service.

Mobility Management can be a very complex set of functions, differentiating a fixed from a mobile network. Accordingly, many levels of comfort are available for mobility support and the interworking of networks. The current discussion is on trigger functions to ease handover between systems. Depending on the origin and destination of triggers, horizontal and vertical triggers are differentiated. Both serve for information provision and context transfer, either within the same layer or across layers.

It might appear strange to have included in an IEEE 802 text book considerations on horizontal handover (between systems of the same kind) and vertical handover (between systems following different standards). But, this is what is required in the future in order to be supported by wireless 802 based systems.

Future Mesh Technologies

The currently available 802 systems protocols are not well suited to support multi-hop or even mesh networking. The IETF MANET initiative after many years of research and development still does not have a suitable solution to wireless ad-hoc networking. It therefore makes sense to spend some space in the book on understanding the strengths and weaknesses of existing MAC protocols and typifying them, leading to better designs for future wireless mesh networks.

One conclusion is that current 802 MAC protocols are not a good basis for enhancement towards mesh networking. Consequently, the results from typifying protocols are translated into the design of a new MAC protocol using decentral control that outperforms 802.11 when used in a mesh network. It is based on the 802.11a PHY and can coexist with 802.11 on the same channel.

Cognitive Radio and Spectrum Sharing

The available options for regulation of radio spectrum, as discussed in Chapter 13, introduce different spectrum-sharing scenarios. The introduction of a QoS guarantee always requires some degree of exclusiveness when sharing spectrum. If a guarantee is not required, primary systems may share spectrum. Coexistence is less adequate to support QoS while cooperation increases the level of possible QoS support. Regulation authorities can delegate the control of spectrum access to one or multiple private entities to enable spectrum trading at the secondary market. Secondary usage might be allowed for underlay or overlay spectrum sharing, provided that secondary radio systems defer from spectrum utilization whenever the license-holding primary radios access their spectrum. Secondary radios can try to coexist with primary radios without interfering them. Cooperation between secondary and primary radios enables the secondary radios to support QoS with known interruptions. Secondary radio systems are only able to guarantee QoS if the primary radio systems commit themselves not to interfere.

In the short term, commercial broadband and cellular networks require exclusive spectrum access to enable QoS guarantees for the customers. Restricted secondary spectrum usage and spectrum trading are grades of flexibility to increase the overall efficiency of spectrum utilization. The licensing process itself needs to be accelerated and requires more flexibility to reflect the rapid developments in the wireless communication market.

In the long term, spectrum used by future wireless broadband systems covering wide areas will most likely be a combination of exclusively accessed spectrum and unlicensed/open spectrum. The exclusively used spectrum allows the customer a minimum level of QoS. The shared spectrum enables an extension of network capacity to provide more services and to increase the number of served customers. Intelligent spectrum sharing algorithms for coordination, as introduced in Chapter 13, improve the efficiency of spectrum usage and increase the radio networks' capabilities to support QoS also using the shared spectrum.

For indoor, short-range communication at high data rates, such as wireless USB, the advantages of liberalizing spectrum access outweigh its risk. The cognitive radio approach for flexible spectrum access is ideal for realizing such communication systems: It lies between the two extremes of open and unlicensed spectrum on the one hand, and the "command-and-control" licensing on the other hand. Cognitive radios can be modified to any level of freedom between these two extremes. Due to locally limited operation of this application scenario, no tragedy of commons exists and freeing access to spectrum will stimulate innovations and economic success. In distributed environments, policy adaptive cognitive radios provide the necessarily required flexibility and intelligence in spectrum access: Local usage constraints are taken into account while etiquettes enable distributed coordination through cooperation.

With cognitive radio, spectrum assignment and licensing will become more dynamic. Greater flexibility in responding to emerging demands of the information society, as well as to market requirements, will be the result.

There are still so many problems, addressed in the book, unsolved. The authors are happy about this, since this opens opportunities for researchers and developers to contribute and also leaves space towards updating the book, or writing another one, in a few years.

Abbreviations

3G	Third Generation
3GPP	3rd Generation Partnership Project
4BOK	Quaternary Bi-Orthogonal Keying
4G	Fourth Generation
5C	5 Criteria
AAA	Authentication, Authorization, Accounting
AAS	Advanced Antenna System
AC	Access Category
ACH	Access Channel
ACK	Acknowledgment
AES	Access-E-Signal
AIEE	American Institute of Electrical Engineers
AIFS	Arbitration Interframe Space
AMC	Adaptive Modulation and Coding
AODV	Ad-Hoc On-Demand Distance Vector Routing
AP	Access Point
API	Application Programming Interface
ARF	Auto Rate Fallback
ARIB	Association of Radio Industries and Businesses
ARP	Address Resolution Protocol
ARQ	Automatic Repeat Request
AS	Authentication Server
ATM	Asynchronous Transfer Mode
ATS	Abstract Test Suite
B3G	Beyond Third Generation
BE	Best Effort
BER	Bit Error Ratio
BES	Busy-E-Signal

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

BIFS	Backoff Interframe Space
BnetzA	Bundesnetzagentur
BP	Beacon Period
BPSK	Binary Phase Shift Keying
BS	Base Station
BSC	Base Station Controller
BSHC	Base Station Hybrid Coordinator
BSID	Base Station Identifier
BSN	Block Sequence Number
BSS	Basic Service Set
BSSID	Basic Service Set Identification
BTC	Block Turbo Coding
BTS	Base Transceiver Station
BW	Bandwidth
BWA	Broadband Wireless Access
C	cooperative behavior
C/I	Carrier-to-Interference
CA	Collision Avoidance
CACR	Channel Reservation
CACR + SR	Channel Reservation with Spatial Reuse
CAP	Cognitive Access Point
CAP	Contention Access Period
CAP	Controlled Access Phase
CAPEX	Capital Expenditure
CBR	Constant Bit Rate
CC	Central Controller
CC	Convolutional inner Code
CCA	Clear Channel Assessment
CCF	Common Channel Framework
CCHC	Central Controller Hybrid Coordinator
CCK	Complementary Code Keying
CCPCH	Common Control Physical Channel
CCW	Channel Coordination Window
CDF	Complementary Cumulative Distribution Function
CDM	Code Division Multiplex
CDMA	Code Division Multiple Access
CE	Consumer Electronics
CEPT	European Conference of Post and Telecommunications Administration
CF	Coordination Function
CFA	Call for Applications
CFP	Contention Free Period
CFP	Call for Proposals
CID	Connection Identifier
COOP	cooperation strategy
CP	Contention Period
CP	Cyclic Prefix
CPG	Conference Preparatory Group
CPS	Common Part Sublayer

CRC	Cyclic Redundancy Check
CRN	Cognitive Relay Node
CS	Convergence Sublayer
CS	Circuit Switched
CSCC	Common Spectrum Coordination Channel
CSMA	Carrier Sense Multiple Access
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Access with Collision Detection
CTA	Channel Time Allocations
CTAP	Channel Time Allocation Period
CTC	Convolutional Turbo Coding
CTS	Clear To Send
CTX	Clear to Switch
CW	Contention Window
D	defective behavior
DA	Destination Address
DARPA	Defense Advanced Research Project Agency
DBPC-REQ	Downlink Burst Profile Change Request
DBPC-RSP	Downlink Burst Profile Change Response
DC	Direct Current
DCD	Downlink Channel Descriptor
DCF	Distributed Coordination Function
DCM	Dual-Carrier Modulation
DCS	Dynamic Channel Selection
DECT	Digital Enhanced Cordless Telecommunication
DEF	defection strategy
DEV	Independent Data Devices
DEV	Device
DFS	Dynamic Frequency Selection
DHCP	Dynamic Host Configuration Protocol
DIFS	Distributed Coordination Function Interframe Space
DiL	Direct Link
DIMSUM	Dynamic Intelligent Management of Spectrum for Ubiquitous Mobile-access Networks
DIUC	Downlink Interval Usage Code
DL	Direct Link
DL	Downlink
DLC	Data Link Control
DLFP	Downlink Frame Prefix
DLP	Direct Link Protocol
DNA	Detecting Network Attachment
DPCH	Dedicated Physical Channel
DQPSK	Differential QPSK
DRCA	Distributed Reservation Channel Access
DRPIE	DRP Information Element
DRRM	Distributed Radio Resource Management
DS	Distribution System
DSA	Dynamic Spectrum Allocation

DSA	Dynamic Service Addition
DSA++	Dynamic Slot Assignment
DSC	Dynamic Service Change
DSD	Dynamic Service Deletion
DSL	Digital Subscriber Line
DSM	Distribution System Medium
DSS	Distribution System Services
DSSS	Direct Sequence Spread Spectrum
DTP	Data Transfer Period
DVB	Double Value Busy-E-Signal
E2E	End to End
ECA	European Common Allocation Table
ECC	Electronic Communications Committee
ECH	Echo Channel
ECMA	European Computer Manufacturers Association
ECSG	Executive Committee Study Group
EDCA	Enhanced Distributed Channel Access
EIFS	Extended Interframe Space
EIRP	Equivalent Isotropic Radiated Power
EKS	Encryption Key Sequence
ERL	Error Ratio Limit
ERRA	Early Route Rearrangement
ERU	Early Route Update
ESS	Extended Service Set
ETSI	European Telecommunications Standards Institute
ETSI BRAN	European Telecommunications Standards Institute Broadband Radio Access Networks
FAC	Flow Admission Control
FBWA	Fixed Broadband Wireless Access
FC	Fragment Control
FCC	Federal Communications Commission
FCH	Frame Control Header
FCS	Frame Check Sequence
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplex
FEC	Forward Error Correction
FEP	Fair Elimination Phase
FEPCL	FEP Contention Level
FFT	Fast Fourier Transformation
FHSS	Frequency-Hopping Spread Spectrum
FM	Frequency Management Working Group
FM	Forwarding Mode
FP6	6th Framework research funding Program
FRN	Fixed Relay Node
FRS	Fixed Relay Station
FSH	Fragmentation Subheader
FSN	Fragment Sequence Number
FUT	Fixed User Terminal

FuTURE	Future Technologies for Universal Radio Environment
GERAN	GSM/EDGE Radio Access Network
GSM	Global System for Mobile Communications
H/2	High Performance Local Area Network Type 2
HBFSA	Hierarchical Beacon with Fixed Slot Allocation
HC	Hybrid Coordinator
HCCA	HCF Controlled Channel Access
HCF	Hybrid Coordination Function
HCS	Header Check Sequence
HFDD	Half-duplex Frequency Division Duplex
HHO	Horizontal HandOver
HIPERACCESS	High Performance Radio Access Network
Hiperlan 1	High Performance Local Area Network Type 1
HiperMAN	High Performance Metropolitan Area Network
HO	HandOver
HR/DSSS	High Rate Direct Sequence Spread Spectrum
IAPP	Inter AP Protocol
IBSS	Independent Basic Service Set
ICI	Inter-Channel Interference
IE	Information Element
IEEE	Institute of Electrical and Electronics Engineers
IEEE-SA	IEEE Standards Association
IETF	Internet Engineering Task Force
IFFT	Inverse Fast Fourier Transform
IFS	Interframe Space
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
IR	Infrared
IRE	Institute of Radio Engineers
IRTF	Internet Research Task Force
ISI	Inter-Symbol Interferences
ISM	Industrial, Scientific and Medical
IST	Information Society Technology
ITG	Informationstechnische Gesellschaft
ITU-R	Radiocommunication Sector of the ITU
ITU-T	International Telecommunication Union
L2	Link Layer
L3	Network Layer
LA	Link Adaptation
LAN	Local Area Network
LBT	Listen-Before-Talk
LCI	Location Configuration Information
LLC	Logical Link Control
LMSC	LAN/MAN Standards Committee
LOS	Line-Of-Sight
LRC	Long Retry Counter
LRE	Limited Relative Error

LWMP	Light-Weight Mesh Point
MA	Multiple Access
MAC	Medium Access Control
MAC-SAP	MAC Service Access Point
MANET	Mobile Ad-hoc Networks
MAP	Mesh Access Point
MAS	Medium Access Slot
MB-OFDM	Multiband-OFDM
MBWA	Mobile Broadband Wireless Access
MCF	Mesh Coordination Function
MCHO	Mobile-Controlled HandOver
MCM	Multi-carrier Modulation
MCS	Modulation and Coding Scheme
MCTA	Management Channel Time Allocations
MDA	Mesh Deterministic Access
MDAOP	Mesh Deterministic Access Opportunity
MDCF	Mesh Distributed Coordination Function
MIB	Management Information Base
MIC	Ministry of Internal Affairs and Communications
MICH	Mobile-Initiated HandOver
MIFS	Minimum Interframe Space
MII	Ministry of Information Industry
MIMO	Multiple Input Multiple Output
MIP	Mobile IP
mip4	Mobility for IPv4
mip6	Mobility for IPv6
mITF	mobile IT Forum
ML	Mesh Link
MLME	MAC Layer Management Entity
MM	Mobility Management
MMR	Mobile Multi-hop Relaying
MMS	Multimedia Messaging Service
mmW	millimeter Wave
MobOpts	Mobility Optimizations
MP	Mesh Point
MPDU	MAC Protocol Data Unit
MQAP	Mesh Quality-of-Service Access Point
MR	Measurement Request/Report
MRN	Mobile Relay Node
MRS	Mobile Relay Station
MS	Mobile Station
MSDU	MAC Service Data Unit
MSG	Multi Stage Game
MSTH	Request Medium Sensing Time Histogram Request
MTSF	MDCF Timing Synchronisation Function
MUT	Mobile User Terminal
MWF	Multi Wall and Floor
N	Node

NACCH	Network Access and Connectivity Channel
NAV	Network Allocation Vector
NCHO	Network-Controlled HandOver
NE	Nash Equilibrium
NesCom	New Standards Committee
NG	Next Generation
NGN	Next Generation Network
NICH	Network-Initiated HandOver
NiCT	National Institute of Communication Technology
NLOS	Non-Line-Of-Sight
nQSTA	Non-Quality-of-Service Station
NRS	Nomadic Relay Station
nrtPS	non-real-time Polling Service
NTIA	National Telecommunications and Information Administration
ODMA	Opportunity Driven Multiple Access
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OPEX	Operational Expenditure
OSI	Open System Interconnection
OWL	Web Ontology Language
PAC	Piconet Acquisition Codeword
PAPR	Peak to Average Power Ratio
PAR	Project Authorization Request
PC	Point Coordinator
PC	Personal Computer
PCA	Prioritized Contention Access
PCF	Point Coordination Function
PCS	Personal Communications Service
PDA	Personal Digital Assistant
PDU	Protocol Data Unit
PER	Packet Error Ratio
PHS	Payload Header Suppression
PHY	Physical layer
PICS	Protocol Implementation Conformance Statement
PID	Piconet ID
PLCP	Physical Layer Convergence Protocol
PLME	PHY Layer Management Entity
PLR	Packet Loss Rate
PMD	Physical Medium Dependent
PMP	Point-to-Multipoint
PNC	Piconet Coordinator
PoA	Point of Access
PP	Prioritization Phase
PPCL	PP Contention Level
PPDU	PLCP Protocol Data Unit
PPP	Point-to-Point Protocol
PSDU	PHY Service Data Unit
PSFD	Power Spectral Flux Density

PSH	Packing Subheader
QAM	Quadrature Amplitude Modulation
QAP	Quality-of-Service Access Point
QBSS	QoS supporting BSS
QoS	Quality-of-Service
QoS CF-Poll	Quality-of-Service Contention Free Poll
QPSK	Quaternary Phase Shift Keying
QSTA	QoS supporting Station
QTS	QoS-related Traffic Specification
R&D	Research and Development
RA	Receiving station Address
RAP	Radio Access Point
RCT	Radio Conformance Tests
REC	Relay Enhanced Cell
RegTP	Regulatory Authority for Telecommunications and Posts
RERR	Route Error
RF	Radio Frequency
RIFS	Retransmission Interframe Space
RISE	Radio Interference Simulation Engine
RKRL	Radio Knowledge Representation Language
RN	Relay Node
RNC	Radio Network Controller
RNG-REQ	Ranging Request
RNG-RSP	Ranging Response
RPI	Receive Power Indication
RR	Radio Regulatory
RRC	Radio Resource Control
RREP	Route Reply
RREQ	Route Request
RRM	Radio Resource Management
RS	Reed–Solomon
RTG	Receive/transmit Transition Gap
rtPS	real-time Polling Service
RTS	Request To Send
RTX	Ready to Switch
RWM	Random Waypoint Model
SA	Source Address
SAP	Service Access Point
SC	Single-Carrier
SC	Standing Committee
SCI	Strategy Comparison Index
SDL	Specification and Description Language
SDM	Space Division Multiplex
SDMA	Space Division Multiple Access
SDR	Software Defined Radio
SDT	SDL Design Tool
SDU	Service Data Unit
SE	Spectrum Engineering

SF	SubFrame
SFID	Service Flow Identifier
SFIR	Spatial Filtering for Interference Reduction
SG	Study Group
SIB	Spectrum Information Base
SIFS	Short Interframe Space
SIM	Subscriber Identity Module
SINR	Signal to Interference plus Noise Ratio
SIP	Session Initiation Protocol
SIR	Signal to Interference Ratio
SLS	Spectrum Load Smoothing
SME	Station Management Entity
SMS	Short Message Service
SNMP	Simple Network Management Protocol
SNR	Signal-to-Noise Ratio
SOHO	Small Offices and Home Offices
SPEETCL	SDL Performance Evaluation
SR	Software Radio
SRC	Short Retry Counter
SS	Station Services
SS	Subscriber Station
SSG	Single Stage Game
STA	802.11 Station
STC	Space-Time Coding
SVB	Single Value Busy-E-Signal
TA	Transmitting station Address
TAG	Technical Advisory Group
TBTT	Target Beacon Transmission Time
TCH	Traffic Channel
TCP	Transmission Control Protocol
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TFC	Time-Frequency Code
TFT	TitForTat strategy
TFTP	Trivial File Transfer Protocol
TG	Task Group
TP	Transmission Phase
TPC	Transmit Power Control
TSF	Time Synchronization Function
TSS & TP	Test Suite Structure and Test Purposes
TSWR	Time Sharing Wireless Router
TSWR + SR	Time Sharing Wireless Router with Spatial Reuse
TTG	Transmit/receive Transition Gap
TTL	Time-To-Live
TU	Time Unit
TXOP	Transmission Opportunity
TXOPlimit	Transmission Opportunity Limit
UCA	Uniform Circular Array

UCD	Uplink Channel Descriptor
UDA	Unused DRP Announcement
UDP	User Datagram Protocol
UDR	Unused DRP Response
UGS	Unsolicited Grant Service
UIUC	Uplink Interval Usage Code
UL	Uplink
UML	Universal Modeling Language
UMTS	Universal Mobile Telecommunication System
U-NII	Unlicensed National Information Infrastructure
UPCS	Unlicensed PCS
USB	Universal Serial Bus
UT	User Terminal
UTRA-FDD	UMTS Terrestrial Radio Access Frequency Division Duplex
UTRAN	UMTS Terrestrial Radio Access Network
UTRA-TDD	UMTS Terrestrial Radio Access Time Division Duplex
UWB	Ultra-Wide Band
VCI	Virtual Channel Identifier
VDE	Verband der Elektrotechnik, Elektronik und Informationstechnik e.V.
VHO	Vertical HandOver
VoIP	Voice over IP
VPI	Virtual Path Identifier
WAR	Wireless Application Protocol
WARP2	Wireless Access Radio Protocol 2
WCDMA	Wideband Code Division Multiple Access
W-CHAMB	Wireless Channel Oriented Ad-hoc Multihop Broadband
WEP	Wired Equivalent Privacy
WG	Working Group
WiBro	Wireless Broadband
WIEN	Wireless Interworking with External Networks
Wi-Fi	Wireless Fidelity
WiMA	Wi-Mesh Alliance
WiMAX	Worldwide interoperability for Microwave Access
WINNER	Wireless World Innitiative New Radio
WirelessHUMAN	Wireless High-speed Unlicensed Metropolitan Area Network
WLAN	Wireless Local Area Network
WM	Wireless Medium
WMAN	Wireless Metropolitan Area Network
WNG	Wireless Next Generation
WPAN	Wireless Personal Area Network
WR	Wireless Router
WRAN	Wireless Regional Area Network
WRC	World Radio Conference
WWI	Wireless World Research Initiative
WWRF	Wireless World Research Forum
XG	DARPA NeXt Generation Communication
XML	Extendable Markup Language

References

3GPP (1998). *Selection Procedures for the Choice of Radio Transmission Technologies of the Universal Mobile Telecommunication System UMTS*. UMTS 30.03, 3G TR 101 112. 3rd Generation Partnership Project.

3GPP (2002). *Technical Specification Group Services and System Aspects; 3GPP System to Wireless Local Area Network (WLAN) Interworking; Functional and Architectural Definition (Release 6)*. TR 23.934 V1.0.0. 3rd Generation Partnership Project.

3GPP (2003). *Technical Specification Group Services and System Aspects; Feasibility Study on 3GPP System to Wireless Local Area Network (WLAN) Interworking (Release 6)*. TR 22.934 V6.2.0. 3rd Generation Partnership Project.

3GPP (2004). *Technical Specification Group Radio Access Network "Physical Layer Items Not For Inclusion In Release '99"*. 3G TR 25.833 V1.1.0. 3rd Generation Partnership Project.

3GPP (2005a). *Technical Specification Group GSM/EDGE Radio Access Network; Radio Transmission and Reception (Release 1999)*. TS 05.05 V8.19.0. 3rd Generation Partnership Project.

3GPP (2005b). *Technical Specification Group Services and System Aspects; 3GPP System to Wireless Local Area Network (WLAN) interworking; System Description (Release 6)*. TS 23.234 V6.5.0. 3rd Generation Partnership Project.

3GPP (2005c). *Technical Specification Group Services and System Aspects; Telecommunication Management; Charging Management; Wireless Local Area Network (WLAN) Charging (Release 6)*. TS 32.252 V.1.4.0. 3rd Generation Partnership Project.

3GPP (2005d). *Technical Specification Group Services and System Aspects; Requirements on 3GPP System to Wireless Local Area Network (WLAN) Interworking (Release 7)*. TS 22.234 V7.2.0. 3rd Generation Partnership Project.

3GPP (2005e). *Technical Specification Group RAN; User Equipment (UE) Radio Transmission and Reception (FDD) (Release 7)*. TS 25.101 V7.1.0. 3rd Generation Partnership Project.

3GPP (2005f). *Technical Specification Group Services and System Aspects; 3G Security; Wireless Local Area Network (WLAN) Interworking Security (Release 6)*. TS 33.234 V6.4.0. 3rd Generation Partnership Project.

3GPP (2005g). *Technical Specification Group Services and System Aspects; Vocabulary for 3GPP Specifications (Release 6)*. TR 21.905 V6.9.0. 3rd Generation Partnership Project.

AGGELOU, G. N. and TAFAZOLLI, R. (2001). On the Relaying Capability of Next-Generation GSM Cellular Networks. *IEEE Personal Communications*, **8**(1), 40–7.

IEEE 802 Wireless Systems B. Walke, S. Mangold and L. Berlemann

ALONISTIOTI, N., GLENTIS, A., FOUKALAS, F. and KALOXYLOS, A. (2004). Reconfiguration Management Plane for the Support of Policy-based Network Reconfiguration. In: *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, September 5–8.

ALTMAN, E., BASAR, T., JIMENEZ, T. and SHIMKIN, N. (2002). Competitive Routing in Networks with Polynomial Costs. *IEEE Transactions on Automatic Control*, **47**(1), 92–6.

ANALYSYS (2004). *Study on Conditions and Options in Introducing Secondary Trading of Radio Spectrum in the European Community*. Final Report for the European Commission. Analysys Consulting Ltd.

ANDERSEN, J. B., RAPPAPORT, T. S. and YOSHIDA, S. (1995). Propagation Measurements and Models for Wireless Communications Channels. *IEEE Communications Magazine*, **33**(1), 42–9.

BAADER, F., CALVANESE, D., MCGUINNESS, D., NARDI, D. and PATEL-SCHNEIDER, P. (2003). *The Description Logic Handbook*. Cambridge: Cambridge University Press.

BANA, S. V. and VARAIYA, P. (2000). Space Division Multiple Access (SDMA) for Robust Ad Hoc Vehicle Communication Networks. In: *IEEE Intelligent Transportation Systems 2001*, Oakland, CA, August 25–9, 962–7.

BAYER, N., SIVCHENKO, D., XU, B., RAKOCEVIC, V. and HABERMANN, J. (2006). Transmission Timing of Signalling Messages in IEEE 802.16 based Mesh Networks. In: *European Wireless 2006, EW'06*, Athens, Greece, April 2–5.

BECKMAN, C. and SMITH, G. (2005). Shared Networks: Making Wireless Communication Affordable. *IEEE Wireless Communications*, **12**(2), 78–85.

BELLMAN, R. (1958). On a Routing Problem. *Quarterly of Applied Mathematics*, **16**. Available from: http://wisl.ece.cornell.edu/ECE794/Jan29/bellman1958.pdf [March 2006].

BERG, J. E. (1995). A Recursive Method For Street Micro-cell Path Loss Calculations. In: *6th International IEEE Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC'95*, Toronto, Canada, September 27–9, 140–3.

BERGER, R. J. (2003). Open Spectrum: A Path to Ubiquitous Connectivity. *ACM Queue*, **1**(3). Available from: http://www.acmqueue.org/modules.php?name=Content&pa=showpage&pid=37&page=1.

BERLEMANN, L. (2006). *Distributed Quality-of-Service Support in Cognitive Radio Networks*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Aachen, Germany: Wissenschaftsverlag Mainz.

BERLEMANN, L., HIERTZ, G. R., WALKE, B. and MANGOLD, S. (2005c). Strategies for Distributed QoS Support in Radio Spectrum Sharing. In: *IEEE International Conference on Communications, ICC 2005*, Seoul, Korea, May 16–20, 3271–7.

BERLEMANN, L., HOYMANN, C., HIERTZ, G. R. and MANGOLD, S. (2006a). Coexistence and Interworking of IEEE 802.16 and IEEE 802.11(e). In: *IEEE 63rd Vehicular Technology Conference, VTC2006-Spring*, Melbourne, Australia, May 7–10.

BERLEMANN, L., MANGOLD, S., HIERTZ, G. R. and WALKE, B. (2005b). Radio Resource Sharing Games: Enabling QoS Support in Unlicensed Bands. *IEEE Network, Special Issue on "Wireless Local Area Networking: QoS Provision and Resource Management"*, **19**(4), 59–65. Available from: http://www.comsoc.org/ni/Public/2005/jul/ [August 2005].

BERLEMANN, L., MANGOLD, S., HIERTZ, G. R. and WALKE, B. (2006b). Policy Description of Distributed Medium Access in IEEE 802.11(e) – Invited Paper. In: *European Wireless 2006, EW'06*, Athens, Greece, April 2–5.

BERLEMANN, L., MANGOLD, S., HIERTZ, G. R. and WALKE, B. H. (2006c). Policy-Defined Spectrum Sharing and Medium Access for Cognitive Radios. *Journal of Communications (JCM)*, **1**(1), 1–12.

BERLEMANN, L., MANGOLD, S., HIERTZ, G. R. and WALKE, B. H. (2006d). Spectrum Load Smoothing: Distributed Quality-of-Service Support for Cognitive Radios in Open Spectrum. *European Transactions on Telecommunications*, **17**(3), 395–406.

BERLEMANN, L., MANGOLD, S. and WALKE, B. H. (2005d). Policy-Based Reasoning for Spectrum Sharing in Cognitive Radio Networks. In: *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN2005*, Baltimore, MD, November 8–11, 1–10.

BERLEMANN, L., PABST, R. and WALKE, B. (2005a). Multi-Mode Communication Protocols Enabling Reconfigurable Radios. *EURASIP Journal in Wireless Communications and Networking, Special Issue on "Reconfigurable Radio for Future Generation Wireless Systems"*, **3**, 390–400. Available from: http://www.hindawi.com/journals/wcn/volume-2005/ issue-3.html.

BERLEMANN, L., PABST, R. and WALKE, B. H. (2006e). Efficient Multi-Mode Protocol Architecture for Complementary Air-Interfaces in Relay-Based 4G Networks. *IEEE Wireless Communications Magazine*, **13**(3), 15–23.

BERLEMANN, L. and WALKE, B. (2005). Spectrum Load Smoothing for Optimized Spectrum Utilization – Rationale and Algorithm. In: *IEEE Wireless Communication and Networking Conference, WCNC 2005*, New Orleans, LA, March 13–17.

BERLEMANN, L., WALKE, B. and MANGOLD, S. (2004). Behavior Based Strategies in Radio Resource Sharing Games. In: *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, September 5–8, 840–6.

BERLEMANN, L., HOYMANN, C., HIERTZ, G. R. and WALKE, B. H. (2006f). Unlicensed Operation of IEEE 802.16: Coexistence with 802.11(a) in Shared Frequency Band. In: *17th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2006*, Helsinki, Finland, September 11–14 2006.

BETTSTETTER, C., HARTENSTEIN, H. and PÉREZ-COSTA, X. (2003). Stochastic Properties of the Random Waypoint Mobility Model. *ACM/Kluwer Wireless Network, Special Issue on Modelling and Analysis of Mobile Networks*.

BEYER, J. and JAKOBY, R. (1997). Two Semi-Empirical and Fast Prediction Models for Urban Microcells Compared with Measurements at 919 and 1873 MHz. In: *EPMCC'97*, Bonn, Germany, 291–8.

BIANCHI, G. (1998a). IEEE 802.11 – Saturation Throughput Analysis. *IEEE Communication Letters*, **2**(12), 318–20.

BIANCHI, G. (1998b). Throughput Evaluation of the IEEE 802.11 Distributed Coordination Function. In: *5th Workshop on Mobile Multimedia Communications*, Berlin, Germany, October, 12–14.

BIANCHI, G. (2000). Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, **18**(3), 535–47.

BIANCHI, G. and TINNIRELLO, I. (2005). Remarks on IEEE 802.11 DCF Performance Analysis. *IEEE Communication Letters*, **9**(8), 765–7.

BISLA, B., ELINE, R. and FRANCA-NETO, L. M. (2004). RF System and Circuit Challenges for WiMAX. *Intel Technology Journal: "WiMAX"*, **8**(3).

BLUETOOTH (2006). Available from: http://www.bluetooth.com/.

BOTTIGLELIENGO, M., CASETTI, C., CHIASSERINI, C.-F. and MEO, M. (2004). Short-Term Fairness for TCP Flows in 802.11b WLANs. In: *23rd Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE INFOCOM 2004*, March 7–11.

BOURSE, D., EL-KHAZEN, K., BERTHET, N., WIEBKE, T., DILLINGER, M., BRAKENSIEKs, J., MOESSNER, K., VIVIER, G. and ALONISTIOTI, N. (2005).

End-to-End Reconfigurability: Towards the Seamless Experience. In: *14th IST Mobile and Wireless Communications Summit 2005*, Dresden, Germany, June 19–23.

BOURSE, D., EL-KHAZEN, K., DELAUTRE, A., WIEBKE, T., DILLINGER, M., BRAKENSIEK, J., MOESSNER, K., VIVIER, G. and ALONISTIOTI, N. (2004). European Research in End-to-End Reconfigurability. In: *13th IST Mobile and Wireless Communications Summit 2004*, Lyon, France, June 27–30.

BUDDHIKOT, M. M., KOLODZY, P., MILLER, S., RYAN, K. and EVANS, J. (2005). DIMSUMNet: New Directions in Wireless Networking Using Coordinated Dynamic Spectrum Access (position paper). In: *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2005*, Taormina, Italy, June 13–16.

BUNDESNETZAGENTUR (2005). *Strategic Aspects of the Spectrum Regulation of the RegTP* [online]. Available from: http://www.bundesnetzagentur.de/media/archive/2155.pdf.

CABRIC, D., MISHRA, S. M., WILLKOMM, D., BRODERSEN, R. and WOLISZ, A. (2005). A Cognitive Radio Approach for Usage of Virtual Unlicensed Spectrum. In: *14th IST Mobile and Wireless Communications Summit*, Dresden, Germany, June 19–23.

CAGALJ, M., GANERIWAL, S., AAD, I. and HUBAUX, J.-P. (2005). On Selfish Behavior in CSMA/CA Networks. In: *IEEE INFOCOM 2005*, Miami, FL, March 13–17.

CHERRY, S. M. (2003). The Wireless Last Mile. *IEEE Spectrum*, **40**(9), 18–22.

CIMINI, L. J. (1985). Analysis and Simulation of a Digital Mobile Channel Using Orthogonal Frequency Division Multiplexing. *IEEE Transactions on Communications*, **33**(7), 665–75.

CLARK, D. D., PARTRIDGE, C., RAMMING, J. C. and WROCLAWSKI, J. T. (2003). A Knowledge Plane for the Internet. In: *ACM Special Interest Group on Data Communications Conference, ACM SIGCOMM 2003*, Karlsruhe, Germany, August 25–29.

COASE, R. H. (1959). The Federal Communications Commission. *Journal of Law and Economics*, **2**, 1–40.

COASE, R. H. (1960). The Problem of Social Cost. *Journal of Law and Economics*, **3**, 1–44.

CORLEIAL, A. B. and HELLMANN, M. E. (1975). Bistable Behavior of Aloha-Type Systems. *IEEE Transactions on Communications*, **23**, 404–9.

CORREIA, L. M. (2001). *Wireless Flexible Personalised Communications, COST259: European Co-operation in Mobile Radio Research*. New York: John Wiley & Sons, Inc.

COST (1999). *Action 231, Final Report, Digital Mobile Radio towards Future Generation Systems*. Technical Report. European Commission.

COURCOUBETIS, C. and WEBBER, R. (2003). *Pricing Communication Networks – Economics, Technology and Modelling*. Chichester, UK: John Wiley & Sons, Ltd.

CUNNINGHAM, G. A. (1990). Delay versus Throughput Comparisons for Stabilized Slotted ALOHA. *IEEE Transactions on Communications*, **38**(11), 1932–4.

DARPA (2003). *The XG Architectural Framework. Request for Comments*. Version 1.0. Prepared by BBN Technologies, Cambridge, MA: DARPA XG Working Group. Available from: http://www.ir.bbn.com/projects/xmac/vision.html [May 2005].

DARPA (2004a). *The XG Vision. Request for Comments*. Version 2.0. Prepared by BBN Technologies, Cambridge, MA: DARPA XG Working Group. Available from: http://www.ir.bbn.com/projects/xmac/vision.html [May 2005].

DARPA (2004b). *XG Policy Language Framework. Request for Comments*. Version 1.0. Prepared by BBN Technologies, Cambridge, MA: DARPA XG Working Group. Available from: http://www.ir.bbn.com/projects/xmac/pollang.html [May 2005].

DARPA (2005). Available from: http://www.darpa.mil/ato/programs/xg/ [June].

DAS, S. K., LIN, H. and CHATTERJEE, M. (2004). An Econometric Model for Resource Management in Competitive Wireless Data Networks. *IEEE Network*, **18**(6), 20–6.

DEERING, S. and HINDEN, R. (1998). *Internet Protocol, Version 6 (IPv6)*. IETF Network Working Group, RFC 2460.

DEMESTICHAS, P., KOUTSOURIS, N., KOUNDOURAKIS, G., TSAGKARIS, K., OIKONOMOU, A., STAVROULAKI, V., PAPADOPOULOU, L., THEOLOGOU, M. E., VIVIER, G. and EL KHAZEN, K. (2003). Management of Networks and Services in a Composite Radio Context. *IEEE Wireless Communications*, **10**(4), 44–51.

DEMESTICHAS, P., VIVIER, G., EL-KHAZEN, K. and THEOLOGOU, M. (2004). Evolution in Wireless Systems Management Concepts: From Composite Radio Environments to Reconfigurability. *IEEE Communications Magazine*, **42**(5), 90–8.

DIJKSTRA, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik 1*, 269–327. Available from: http://www.digizeitschriften.de/contentserver/ contentserver? command=docconvert&docid=314537 [March 2006].

DILLINGER, M., MADANI, K. and ALONISTIOTI, N. (2003). *Software Defined Radios: Architectures, Systems and Functions*. Chichester, UK: John Wiley & Sons, Ltd.

E2R (2005). Available from: http://www.e2r.motlabs.com/e2r [June].

ECC (2004). *ECC Decision of 12 November 2004 on the Harmonized Use of the 5 GHz Frequency Bands for the Implementation of Wireless Access Systems including Radio Local Area Networks (WAS/RLANs)*. ECC/DEC/(04)08. Electronic Communications Committee.

ECMA (2005a). *MAC-PHY Interface for ECMA-368*. Standard ECMA-369. Geneva, Switzerland: European Computer Manufacturers Association.

ECMA (2005b). *High Rate Ultra Wideband PHY and MAC Standard*. Standard ECMA-368. Geneva, Switzerland: European Computer Manufacturers Association.

EINHAUS, M., KLEIN, O. and LOTT, M. (2005). Interference Averaging and Avoidance in the Downlink of an OFDMA System. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, September 11–14.

ENGELS, M. (2002). *Wireless OFDM Systems: How to Make Them Work?* Kluwer Academic.

ERC (1996). *Harmonized Radio Frequency Channel Arrangements for Digital Terrestrial Fixed Systems Operating in the Band 10.0–10.68 GHz*. CEPT/ERC/REC 12-05 E. European Radiocommunications Committee.

ERC (1998). *Preferred Frequency Bands for Fixed Wireless Access in the Frequency Range between 3 and 29.5 GHz*. Technical Report CEPT/ERC/REC 13-04. European Radiocommunications Committee.

ERC (1999). *ERC Decision on the Designation of the Harmonised Frequency Band 40.5 to 43.5 GHz for the Introduction of Multimedia Wireless Systems (MWS), including Multipoint Video Distribution Systems (MVDS)*. ERC/DEC 99-15. European Radiocommunications Committee.

ERC (2000). *ERC Recommendation 00-05 Use of Parts of the Band 24.5–26.5 GHz for Fixed Wireless Access*. ERC/REC 00-05. European Radiocommunications Committee.

ERC (2001a). *ERC Recommendation 01-02 Preferred Channel Arrangement for Digital Fixed Service Systems Operating in the Frequency Band 31.8–33.4 GHz*. ERC/REC 01-02. European Radiocommunications Committee.

ERC (2001b). *ERC Decision of 12 March 2001 on Harmonized Frequencies, Technical Characteristics and Exemption from Individual Licensing of Short Range Devices used for Radio Local Area Networks (RLANs) Operating in the Frequency Band 2400–2483.5 MHz*. ERC/DEC 01-07. European Radiocommunications Committee.

ERC (2001c). *ERC Recommendation 01-03 Use of Parts of the Band 27.5–29.5 GHz for Fixed Wireless Access (FWA)*. ERC/REC 01-03. European Radiocommunications Committee.

ERC (2005). *ERC Recommendation 70-03 Relating to the Use of Short Range Devices (SRD)*. ERC/REC 70-03. European Radiocommunications Committee.

ESSELING, N. (2005). *Ein Relaiskonzept für das hochbitratige drahtlose lokale Netz HIPERLAN/2*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Aachen, Germany: Wissenschaftsverlag Mainz.

ESSELING, N., VANDRA, H. S. and WALKE, B. (2000). A Forwarding Concept for HiperLAN/2. *In: European Wireless 2000*, Dresden, Germany, September 13–14.

ESSELING, N., VANDRA, H. S. and WALKE, B. (2001). A Forwarding Concept for HiperLAN/2. *Computer Networks*, **37**(1), 25–32.

ESSELING, N., WALKE, B. and PABST, R. (2004). Performance Evaluation of a Fixed Relay Concept for Next Generation Wireless Systems. In: *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, September 5–8.

ETSI (1993). *High Performance Radio Local Area Network (HIPERLAN); Services and Facilities – ETR 096. Technical Report, ETSI RES; Radio Equipment and Systems*. ETSI TC RES10.

ETSI (1995). *Equipment and Systems (RES); HIPERLAN*. Functional Specifications.

ETSI (1997). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Requirements and Architecture for Wireless Broadband Access*. TR 101 031. ETSI BRAN.

ETSI (1999). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions*. DTS 0020004-1. Sophia Antipolis, France: ETSI BRAN.

ETSI (2000). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Functional Specification; Data Link Control (DLC) Layer; Part 1: Basic Data Transport Functions*. TS 101 761-1. Sophia Antipolis, France: ETSI BRAN.

ETSI (2001a). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Requirements and Architectures for Interworking between HIPERLAN/2 and 3rd Generation Cellular Systems*. TR 101 957 V1.1.1. Sophia Antipolis, France: ETSI BRAN.

ETSI (2001b). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 4; Extension for Home Environment Standard*. TS 101 761-4. Sophia Antipolis, France: ETSI BRAN.

ETSI (2002). *ETSI Broadband Radio Access Networks (BRAN); HIPERLAN Type 2; Data Link Control (DLC) Layer; Part 2; Radio Link Control (RLC) Sub-layer Standard*. TS 101 761-2 V1.3.1. Sophia Antipolis, France: ETSI BRAN.

ETSI (2003a). *ETSI Broadband Radio Access Networks (BRAN); Functional Requirements for Fixed Wireless Access Systems below 11 GHz: HIPERMAN*. 101 856 V1.1.1. Sophia Antipolis, France: ETSI BRAN.

ETSI (2003b). *ETSI Broadband Radio Access Networks (BRAN); HIPERMAN; Physical (PHY) Layer*. TS 102 177. Sophia Antipolis, France: ETSI BRAN.

ETSI (2003c). *ETSI Broadband Radio Access Networks (BRAN); HIPERMAN; Data Link Control (DLC) Layer*. TS 102 178. Sophia Antipolis, France: ETSI BRAN.

ETSI (2004). *Digital Enhanced Cordless Telephony (DECT); Common Interface (CI); Part 3: Medium Access Control (MAC) Layer*. Technical Report. ETSI ETS 300 175-3 V1.8.0.

EU (2002a). *Directive 2002/21/EC on a Common Regulatory Framework for Electronic Communications Networks and Services (Framework Directive)*. The European Parliament and the Council of the European Union. Available from: http://www.euractiv.com/Article?tcmuri=tcm:29-117547-16&type=LinksDossier.

EU (2002b). *Decision No 676/2002/EC on a Regulatory Framework for Radio Spectrum Policy in the European Community (Radio Spectrum Decision)*. The European Parliament and the Council of the European Union. Available from: http://www.euractiv.com/Article?tcmuri=tcm:29-117547-16&type=LinksDossier.

FCC (2002). *Report of the Spectrum Rights and Responsibilities Working Group*. ET Docket No. 02-135. Federal Communications Commission Spectrum Policy Task Force.

FCC (2003a). *Report and Order (FCC 03-287): Revision of Parts 2 and 15 of the Commission's Rules to Permit Unlicensed National Information Infrastructure (U-NII) Devices in the 5 GHz Band*. ET Docket No. 03-122. Federal Communications Commission.

FCC (2003b). *Notice of Inquiry and Notice of Proposed Rulemaking (FCC 03-289): Establishment of an Interference Temperature Metric to Quantify and Manage Interference and to Expand Available Unlicensed Operation in Certain Fixed, Mobile and Satellite Frequency Bands*. ET Docket No. 03-237. Federal Communications Commission.

FCC (2003c). *Notice for Proposed Rulemaking (FCC 03-322): Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies*. ET Docket No. 03-108. Federal Communications Commission.

FCC (2004a). *Notice of Proposed Rulemaking (FCC 04-100): Unlicensed Operation in the Band 3650–3700 MHz*. ET Docket No. 04-151. Federal Communications Commission.

FCC (2004b). *Second Report and Order (FCC 04-167): Promoting Efficient Use of Spectrum Through Elimination of Barriers to the Development of Secondary Markets*. WT Docket No. 00-230. Federal Communications Commission.

FCC (2004c). *Notice of Proposed Rulemaking (FCC 04-113): Unlicensed Operation in the TV Broadcast Bands*. ET Docket No. 04-186. Federal Communications Commission.

FCC (2005a). *Report and Order (FCC 05-57): Facilitating Opportunities for Flexible, Efficient, and Reliable Spectrum Use Employing Cognitive Radio Technologies*. ET Docket No. 03-108. Federal Communications Commission.

FCC (2005b). *Report and Order and Memorandum Opinion and Order (FCC 05-56): Wireless Operations in the 3650–3700 MHz Band*. ET Docket No. 04-151. Federal Communications Commission.

FCC (2005c). *Part 15 – Radio Frequency Devices – of the Commissions Rules to Permit Unlicensed National Information Infrastructure (U NII) Devices in the 5 GHz Band*. Federal Communications Commission. Available from: http://www.fcc.gov/oet/info/rules/part15/part15_4_05_05.pdf.

FEUERSTEIN, M., BLACKARD, K. L., RAPPAPORT, T. S., SEIDEL, S. Y. and XIA, H. H. (1994). Path Loss, Delay Spread and Outage Models as Functions of Antenna Height for Microcellular System Design. *IEEE Transactions on Vehicular Technology*, **43**(3), 487–9.

FLOYD, R. W. (1962). Shortest Path. *Communications of the ACM*, **5**(6), 345.

FORKEL, I. and SALZMANN, M. (2001). Radio Propagation Modeling and its Application for 3G Mobile Network Simulation. In: *Aachen Symposium on Signal Theory (ASST)*, Aachen, Germany, September 20–21, 363–70.

FUCHS, M., DEL GARDO, G. and HAARDT, M. (2005). A Novel Tree-based Scheduling Algorithm for the Downlink of Multi-user MIMO Systems with ZF Beamforming. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, March 18–23.

FUDENBERG, D. and TIROLE, J. (1998). *Game Theory*. Cambridge, MA: MIT Press.

FURUSKÄR, A. (2003). *Radio Resource Sharing and Bearer Service Allocation for Multi-Bearer Service, Multi-Access Wireless Networks*. Thesis (PhD), Radio Communication Systems Laboratory, Royal Institute of Technology, Sweden.

GARG, V. K. (2000). *IS-95 and cdma2000*. Prentice-Hall.

GHOSH, A., WOLTER, D. R., ANDREWS, J. G. and CHEN, R. (2005). Broadband Wireless Access with WiMax/802.16: Current Performance Benchmarks and Future Potential. *IEEE*

Communications Magazine, **43**(2), 129–36.

GODARA, L. (1997). Application of Antenna Arrays to Mobile Communications, Part II: Beamforming and Direction-of-Arrival Considerations. *Proceedings of IEEE*, **85**(8), 1195–245.

GRANDBLAISE, D., KLOECK, C., MOESSNER, K., MOHYELDIN, E., PEREIRASAMY, M. K., LUO, J. and MARTOYO, I. (2005). Towards a Cognitive Radio-Based Distributed Spectrum Management. In: *14th IST Mobile and Wireless Communications Summit*, Dresden, Germany, June 19–23.

HAAS, Z. J. and DENG, J. (2002). Dual Busy Tone Multiple Access (DBTMA) – A Multiple Access Control Scheme for Ad Hoc Networks. *IEEE Transactions on Communications*, **50**(6), 975–85.

HABETHA, J., DUTAR, R. and WIEGERT, J. (2002). Performance Evaluation of HIPERLAN/2 Multihop Ad Hoc Networks. In: *European Wireless Conference 2002*, Florence, Italy, February 25–8, 25–31.

HABETHA, J. and WALKE, B. (2002). Fuzzy Rule-Based Mobility and Load Management for Self-organizing Wireless Networks. *Journal of Wireless Information Networks: Standards, Research, Applications*, **9**, 119–40.

HALLDORSSON, M. M., HALPERN, J. Y., LI, L. and MIRROKNI, V. (2004). On Spectrum Sharing Games. In: *23rd Annual ACM Symposium on Principles of Distributed Computing, PODC 2004*, St. John's, Newfoundland, Canada, July 25–8, 107–14.

HAN, S. H. and LEE, J. H. (2005). An Overview of Peak-to-Average Power Ratio Reduction Techniques for Multicarrier Transmission. *IEEE Wireless Communications*, **12**(2), 56–65.

HARROLD, T. J. and NIX, A. R. (2000a). Capacity Enhancement Using Intelligent Relaying for Future Personal Communication Systems. In: *IEEE 52nd Vehicular Technology Conference, VTC 2000-Fall*, Boston, MA, September 24–28, 2115–20.

HARROLD, T. J. and NIX, A. R. (2000b). Intelligent Relaying for Future Personal Communication Systems. In: *IEE Colloquium on Capacity and Range Enhancement Techniques for 3rd Generation Mobile Communications and Beyond*, February.

HATA, M. (1980). Empirical Formula for Propagation Path Loss in Land Mobile Radio Services. *IEEE Transactions on Vehicular Technology*, **29**(3).

HAYKIN, S. (2005). Cognitive Radio: Brain-Empowered Wireless Communications. *IEEE Journal on Selected Areas in Communications*, **23**(2), 201–20.

HAZLETT, T. W. (1998). Spectrum Flash Dance: Eli Noam's Proposal for "Open Access" to Radio Waves. *Journal of Law and Economics*, **44**, 805–21.

HAZLETT, T. W. (2005). Spectrum Tragedy. *Yale Journal on Regulation*, **22**.

HERRMANN, C. (1995). *Stochastische Modelle für ATM-Konzepte*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Aachen, Germany: Wissenschaftsverlag Mainz.

HETTICH, A., ALDINGER, M., DU, Y. and LITZENBURGER, M. (1997). Frequency Sharing Rules for HIPERLAN-/ U-NII-Systems. In: *ACTS Mobile Communications Summit*, Aalborg, Denmark, October 7–10.

HIERTZ, G. R. and HABETHA, J. (2003). A New MAC Protocol for a Wireless Multi-Hop Broadband System Beyond IEEE 802.11. In: *9th Meeting of the Wireless World Research Forum, WWRF9*, Zurich, Switzerland, July 1–2.

HIERTZ, G. R., HABETHA, J., MAY, P., WEISS, E., BAGUL, R. and MANGOLD, S. (2003). A Decentralized Reservation Scheme for IEEE 802.11 Ad Hoc Networks. In: *14th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2003*, Beijing, China, September 7–10.

HIERTZ, G. R., STIBOR, L., HABETHA, J., WEISS, E. and MANGOLD, S. (2005b). Throughput and Delay Performance of IEEE 802.11e Wireless LAN with Block Acknowledgments. In: *European Wireless Conference 2005, EW'05*, Nicosia, Cyprus, April 10–13.

HIERTZ, G. R., ZANG, Y., HABETHA, J. and SIRIN, H. (2005a). IEEE 802.15.3a Wireless Personal Area Networks – The MBOA Approach. In: *11th European Wireless Conference 2005, EW'05*, Nicosia, Cyprus, April 10–13, 204–10.

HORNE, W. D. (2003). Adaptive Spectrum Access: Using the Full Spectrum Space. In: *31st Research Conference on Communication, Information and Internet Policy, TPRC'03*, Arlington, VA, September 19–21.

HOYMANN, C. (2005). Analysis and Performance Evaluation of the OFDM-based Metropolitan Area Network IEEE 802.16. *Computer Networks, Selected Papers from the European Wireless 2004 Conference*, **49**(3), 341–63.

HOYMANN, C. (2006). MAC Layer Concepts to Support Space Division Multiple Access in OFDM based IEEE 802.16. *Wireless Personal Communications*, 37(3–4).

HUNOLD, D., BARRETO, A. N., FETTWEIS, G. P. and MECKING, M. (2000). Concept for Universal Access and Connectivity in Mobile Radionetworks. In: *11th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2000*, London, UK, September 18–21, 847–51.

HUTTER, A., HOYMANN, C., WÄCHTER, B., CIANCA, E. and CASSIAU, N. (2004). Advanced Link Level Techniques for Broadband MIMO Metropolitan Area Networks. In: *7th International Symposium on Wireless Personal Multimedia Communications, WPMC'04*, Abano Terme, Italy, September 12–15.

IEEE (1996). *Standard for High Performance Serial Bus*. IEEE Standard 1394-1995. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (1997). *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. Standard IEEE 802.11. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (1998). *IEEE Standard for Information Technology Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks Common Specifications Part 3: Media Access Control (MAC)*. IEEE 802.1D. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (1999a). *Supplement to IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: High-Speed Physical Layer in the 5 GHZ Band*. IEEE Std 802.11a-1999. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (1999b). *Supplement to IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band*. IEEE Std 802.11b-1999. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2002a). *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Std 802.16-2001. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2002b). *P802.15-03/030r0 TG3a Technical Requirements*. IEEE 802.15.3a. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2002c). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements, Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications* *for Wireless Personal Area Networks (WPANs)*. IEEE 802.15.1. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2002d). *IEEE Recommended Practice for Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements, Part 15.2: Coexistence of Wireless Personal Area Networks with Other Wireless Devices Operating in Unlicensed Frequency Band*. IEEE 802.15.2. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003a). *IEEE Wireless LAN Edition – A Compilation Based on IEEE Std 802.11TM-1999 (R2003) and its Amendments*. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003b). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 5: Spectrum and Transmit Power Management Extensions in the 5 GHz Band in Europe*. IEEE Std 802.11h-2003. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003c). *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2–11 GHz*. IEEE Std 802.16a-2003. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003d). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements, Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)*. IEEE 802.15.4. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003e). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band*. IEEE Std 802.11g-2003. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2003f). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements, Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)*. IEEE 802.15.3. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2004a). *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems*. IEEE Std 802.16-2004. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2004b). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – Local and Metropolitan Area Networks – Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 6: Medium Access Control (MAC) Security Enhancements*. IEEE Std 802.11i-2004. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2004c). *Recommended Practice for Local and Metropolitan Area Networks – Coexistence of Fixed Broadband Wireless Access Systems*. IEEE Std 802.16.2-2004. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2004d). *P802.15-04/0137r3 DS-UWB Physical Layer Submission to 802.15 Task Group 3a*. IEEE 802.15.3a. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2004e). *Project Authorization Request for Draft: Amendment to Standard for Telecommunications and Information Exchange Between Systems – LAN/MAN – Specific Requirements – Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) – Specifications: Higher Speed Physical Layer Extension for the High Rate Wireless Personal Area Networks (WPAN)*. IEEE P802.15.3a. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2005a). *Submission 802.11-05/0710r0* [online]. Available from: http://802wirelessworld.com [March 2006].

IEEE (2005b). *IEEE Standard for Information Technology – Telecommunications and Information Exchange between Systems – LAN/MAN Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment: Medium Access Control (MAC) Quality of Service Enhancements*. IEEE 802.11e/ D13.0. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2005c). Available from: http://www.ieee.org [December 2005].

IEEE (2005d). *IEEE 802.21 Media Independent Handoff Working Group, Project Authorization Request*. New York: Institute of Electrical and Electronics Engineers, Inc. Available from: http://www.ieee802.org/21/802_21_PAR.doc [October 2005].

IEEE (2005e). *Amendment to IEEE Standard for Local and Metropolitan Area Networks – Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Improved Coexistence Mechanisms for License-Exempt Operation*. Working document for IEEE 802.16h. New York: Institute of Electrical and Electronics Engineers, Inc. Available from: http://ieee802.org/16/le/docs/80216h-05_013.pdf.

IEEE (2005f). *802.11 Task Group s* [online]. Available from: http://grouper.ieee.org/groups/802/11/Reports/tgs_update.htm [December 2005].

IEEE (2005g). *802.15 Working Group for WPAN* [online]. Available from: http://www.ieee802.org/15 [December 2005].

IEEE (2005h). *802.15 WPAN Task Group 5 (TG5)* [online]. Available from: http://www.ieee802.org/15/pub/TG5.html [December 2005].

IEEE (2005i). *802.15 WPAN High Rate Alternative PHY Task Group 3a (TG3a)* [online]. Available from: http://www.ieee802. org/15/pub/TG3a.html [December 2005].

IEEE (2005j). *802.15 WPAN Task Group 3b (TG3b)* [online]. Available from: http://www.ieee802.org/15/pub/TG3b.html [December 2005].

IEEE (2005k). *Draft Amendment to Standard for Information Technology – Telecommunications and Information Exchange between Systems – LAN/MAN Specific Requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment 7: Radio Resource Measurement*. IEEE 802.11k/D2.0. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2005l). *A MAC Partial Proposal for IEEE 802.11s*. P802.11-05/595r1. New York: Institute of Electrical and Electronics Engineers, Inc., IEEE 802.11 Task Group"s" on ESS Mesh Networking. Available from: http://www.802wirelessworld.com [May 2006].

IEEE (2005m). *Information Technology – Telecommunications and Information Exchange Between Systems – Wireless Regional Area Networks (WRAN) – Specific Requirements – Part 22: Cognitive Wireless RAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Policies and Procedures for Operation in the TV Bands*. IEEE 802.22. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2006a). Available from: http://ieee802.org [2006 April].

IEEE (2006b). *IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Fixed Broadband Wireless Access Systems – Amendment for Physical* *and Medium Access Control Layers for Combined Fixed and Mobile Operation in Licensed Bands*. IEEE Std 802.16e-2005. New York: Institute of Electrical and Electronics Engineers, Inc.

IEEE (2006c). *Draft Amendment to Standard for Information Technology – Telecommunications and Information Exchange between Systems – LAN/MAN Specific Requirements – Part 11: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Amendment: ESS Mesh Networking*. IEEE 802.11s/D0.01. New York: Institute of Electrical and Electronics Engineers, Inc.

IETF (2004). *The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)*. Internet Draft. Available from: http://www.ietf.org/internet-drafts/draft-ietf-manet-dsr-10.txt [March 2006].

IETF (2005a). *Mobility for IPv4 (mip4)*. IETF Working Group. Available from: http://www.ietf.org/html.charters/mip4-charter.html [October 2005].

IETF (2005b). *Mobility for IPv6 (mip6)*. IETF Working Group. Available from: http://www.ietf.org/html.charters/mip6-charter.html [October 2005].

IETF (2005c). *Detecting Network Attachment (DNA)*. IETF Working Group. Available from: http://www.ietf.org/html.charters/dna-charter.html [October 2005].

IETF (2006a). *The Optimized Link-State Routing Protocol version 2*. Draft IETF MANET OLSR v2-01. Available from: http://www.ietf.org/internet-drafts/draft-ietf-manet-olsrv2-01.txt [March 2006].

IETF (2006b). *Internet Engineering Task Force, Mobile Ad-hoc Networks (MANET)* [online]. Available from: http://www.ietf.org/html.charters/manet-charter.html [March 2006].

IST-STRIKE (2006). Available from: www.ist-strike.org [March 2006].

ITU (1994). *Information Technology – Open System Interconnection – Basic Reference Model: The Basic Model*. ITU-T Recommendation X.200. Geneva, Switzerland: International Telecommunication Union.

ITU (2005a). *ITU-T Study Group 16: Multimedia Terminals, Systems, and Applications* [online]. Available from: http://www.itu.int/ITU-T/studygroups/com16/sg16-q29.html [October 2005].

ITU (2005b). *ITU-T Study Group 13: Next Generation Networks*. (Study Period 2005–8). Geneva, Switzerland: International Telecommunication Union. Available from: http://www.itu.int/ITU-T/studygroups/com13/ [October 2005].

JING, X. and RAYCHAUDHURI, D. (2005). Spectrum Co-existence of IEEE 802.11b and 802.16a Networks using the CSCC Etiquette Protocol. In: *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN2005*, Baltimore, MD, November 8–11, 243–50.

JOHNSON, D., PERKINS, C. and ARKKO, J. (2004). *Mobility Support in IPv6*. IETF Network Working Group, RFC 3775.

JOHNSTON, D. (2004). 802.21 – L2 Services for Handover Optimization. In: *59th IETF*, Seoul, Korea, February 29–March 4.

KAMERMAN, A. and MONEBAN, L. (1997). WaveLAN II: A High-Performance Wireless LAN for the Unlicensed Band. *Bell Labs Technical Journal*, 118–33.

KATZELA, I. and NAGHSHINEH, M. (1996). Channel Assignment Schemes for Cellular Mobile Telecommunication Systems: A Comprehensive Survey. *IEEE Personal Communications*, **3**(3), 10–31.

KEPLER, J. F., KRAUSS, T. P. and MUKTHAVARAM, S. (2002). Delay Spread Measurements on a Wideband MIMO Channel at 3.7 GHz. In: *IEEE Vehicular Technology Conference, VTC2002-Fall*, Vancouver, Canada, September 24–28.

KHERANI, A. A. and SHOREY, R. (2004). Throughput Analysis of TCP in Multi-hop Wireless Networks with IEEE 802.11 MAC. In: *IEEE Wireless Networking Conference*, *WCNC 2004*, Atlanta, GA, March 21–24.

KHUN-JUSH, J., SCHRAMM, P., WACHSMANN, U. and WENGER, F. (1999). Structure and Performance of the HIPERLAN/2 Physical Layer. In: *IEEE 50th Vehicular Technology Conference, VTC1999-Fall*, Amsterdam, Netherlands, September 19–22.

KLEINROCK, L. (1976). *Queuing Systems, Volume II: Computer Applications*. New York: John Wiley & Sons, Inc.

KLEINROCK, L. and LAM, S. (1975). Packet Switching in a Multiple Broadcast Channel: Performance Evaluation. *IEEE Transactions on Communications*, **23**(4), 410–23.

KLEINROCK, L. and SILVESTER, J. (1987). Spatial Reuse in Multi-hop Packet Radio Networks. *Proceedings of IEEE*, **75**(1), 156–67.

KLEINROCK, L. and TOBAGI, F. A. (1975). Packet Switching in Radio Channels: Part I – Carrier Sense Multiple-Access Modes and Their Throughput-Delay Characteristics. *IEEE Transactions on Communications*, **23**(12), 1400–16.

KLOECK, C., JAEKEL, H. and JONDRAL, F. K. (2005). Auctions Sequence as a New Spectrum Allocation Mechanism. In: *14th IST Mobile and Wireless Communications Summit*, Dresden, Germany, June 19–23.

KOFFMAN, I. and ROMAN, V. (2002). Broadband Wireless Access Solutions Based on OFDM Access in IEEE 802.16. *IEEE Communications Magazine*, **40**(4), 96–103.

KOODLI, R. (2003). *Fast Handovers for Mobile IPv6*. draft-ietf-mobileip-fast-mipv6-06.txt.

KOTZ, D., NEWPORT, C. and ELLIOTT, C. (2003). *The Mistaken Axioms of Wireless-Network Research*. Dartmouth College Computer Science Technical Report TR2003-647. Available from: http://www.pdos.lcs.mit.edu/decouto/papers/kotz03.pdf [March 2006].

KREAMLING, A., SCHEIBENBOGEN, M. and WALKE, B. (2000). Dynamic Channel Allocation in Wireless ATM Networks. In: *Wireless Networks*, 6, 381–389.

KRUYS, J. (2003). *Co-existence of Dissimilar Wireless Systems* [online]. Available from: http://www.wi-fi.org/opensection/pdf/co-existence_dissimilar_systems.pdf [July 2005].

KUSUDA, A., YAMOMOTO, K. and YOSHIDA, S. (2005). Performance Comparison of Multihop Relaying and Rate Adaptation in Cellular Communication Systems. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, September 11–14.

LAIHO, J., WACKER, A. and NOVOSAD, T. (2002). *Radio Network Planning and Optimization for UMTS*. Chichester, UK: John Wiley & Sons, Ltd.

LANSFORD, J. (2004). UWB Coexistence and Cognitive Radio. In: *International Workshop on Ultra Wideband Systems, 2004*, Kyoto, Japan, May 18–21, 35–9.

LEAVES, P., MOESSNER, K., TAFAZOLLI, R., GRandBLAISE, D., BOURSE, D., TONJES, R. and BREVEGLIERI, M. (2004). Dynamic Spectrum Allocation in Composite Reconfigurable Wireless Networks. *IEEE Communications Magazine*, **42**(5), 72–81.

LEE, M., KIM, G. and PARK, S. (2005). Seamless and Secure Mobility Management with Location Aware Service (LAS) Broker for Future Mobile Interworking Networks. *Journal of Communications and Networks (JCN)*, **7**(2).

LEE, W. C. Y. (1982). *Mobile Communication Engineering*. New York: McGraw-Hill.

LEFEVRE, F. and VIVIER, G. (2000). Understanding TCP's Behavior over Wireless Links. In: *Symposium on Communications and Vehicular Technology*, 123–30.

LEITH, D. J., CLIFFORD, P., MALONE, D. and NG, A. (2005). TCP Fairness in 802.11e WLANs. *IEEE Communications Letters*, **9**(11), 964–6.

LIN, S. and CASTELLO, D. J. (2004). *Error Control Coding*. Prentice-Hall.

LIN, Y. D. and HSU, Y. C. (2000). Multihop Cellular: A New Architecture for Wireless Communications. In: *IEEE Infocom 2000*, Tel Aviv, Israel, March 26–8, 1273–82.

LOTT, M. and FORKEL, I. (2001). A Multi-Wall-and-Floor Model for Indoor Radio Popagation. In: *IEEE Vehicular Technology Conference, VTC2001-Spring*, Rhodes, Greece, May 6–9.

LUCKY, R. W. (2001). The Precious Radio Spectrum. *IEEE Spectrum*, **38**(9), 90.

MAHESWARAN, R. T. and BASAR, T. (2003). Nash Equilibrium and Decentralized Negotiation in Auctioning Divisible Resources. *Group Decision and Negotiation*, **12**(5), 361–95.

MÄHÖNEN, P. (2004). Cognitive Trends in Making: Future of Networks. In: *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, 5–8 September, 1449–54.

MÄHÖNEN, P., RIIHIJARVI, J., PETROVA, M. and SHELBY, Z. (2004). Hop-by-Hop toward Future Mobile Broadband IP. *IEEE Communications Magazine*, **42**(3), 138–46.

MANGOLD, S. (2003). *Analysis of IEEE 802.11e and Application of Game Models for Support of Quality-of-Service in Coexisting Wireless Networks*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Aachen, Germany: Wissenschaftsverlag Mainz.

MANGOLD, S. and BERLEMANN, L. (2005). IEEE 802.11k: Improving Confidence in Radio Resource Measurements. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, September 11–14.

MANGOLD, S., BERLEMANN, L. and NANDAGOPALAN, S. (2006). Spectrum Sharing with Value-Orientation for Cognitive Radio. *European Transactions on Telecommunications*, **17**(3), 383–94.

MANGOLD, S., BERLEMANN, L. and WALKE, B. (2003). Equilibrium Analysis of Coexisting IEEE 802.11e Wireless LANs. In: *14th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2003*, Beijing, China, September 7–10, 321–5.

MANGOLD, S., CHOI, S. and ESSELING, N. (2001a). An Error Model for Radio Transmissions of Wireless LANs at 5 GHz. In: *Aachen Symposium on Signal Theory 2001*, Aachen, Germany, September 20–21.

MANGOLD, S., CHOI, S., BUDDE, W. and NGO, C. (2001b). 802.11a/e and H2 Interworking Using HCF. *IEEE Working Document 802.11-01/414r0*. Available from: http://www.ieee802.org/11.

MANGOLD, S., HABETHA, J., CHOI, S. and NGO, C. (2001c). Co-existence and Interworking of IEEE 802.11a and ETSI BRAN HiperLAN/2 in MultiHop Scenarios. In: *IEEE 3rd Workshop in Wireless Local Area Networks*, Boston, MA, September 27–28.

MANGOLD, S., JAROSCH, A. and MONNEY, C. (2005b). Cognitive Radio – Trends and Research Challenges. *Swisscom Comtec Magazine 03/2005*, 6–9.

MANGOLD, S., SHANKAR, S. and BERLEMANN, L. (2005a). Spectrum Agile Radio: A Society of Machines with Value-Orientation (invited paper). In: *European Wireless Conference 2005, EW'05*, Nicosia, Cyprus, April 10–13, 539–46.

MANGOLD, S., ZHONG, Z., CHALLAPALI, K. and CHOU, C. T. (2004b). Spectrum Agile Radio: Radio Resource Measurements for Opportunistic Spectrum Usage. In: *47th Annual IEEE Global Telecommunications Conference, Globecom 2004*, Dallas, TX, November 29–December 3.

MANGOLD, S., ZHONG, Z., HIERTZ, G. R. and WALKE, B. (2004a). IEEE 802.11e/802.11k Wireless LAN: Spectrum Awareness for Distributed Resource Sharing. *Wireless Communications and Mobile Computing*, **4**(8), 881–902.

MANNER, J. and KOJO, M. (2004). *Mobility Related Terminology*. IETF Network Working Group RFC 3753.

MAY, R. J., SPETA, J. B., DIXON, K. D., GATTUSO, J. L., GIFFORD, R. L., SHELANSKI, H. A., SICKER, D. C. and WEISMAN, D. (2005). *Digital Age Communications Act – A Proposal for a New Regulatory Framework – Release 1.0*. Washington, DC: Progress & Freedom Foundation.

MCGUINNESS, D. L. and HARMELEN, F. V. (2003). *OWL Web Ontology Language Overview, W3C Candidate Recommendation* [online]. World Wide Web Consortium. Available from: http://www.w3.org/TR/2003/CR-owl-features-20030818/ [June 2005].

MEHTA, M., DREW, N., VARDOULIAS, G., GRECO, N. and NIEDERMEIER, C. (2001). Reconfigurable Terminals: An Overview of Architectural Solutions. *IEEE Communications Magazine*, **39**(8), 82–9.

MIC (2003). *Radio Policy Vision: Report from Telecommunications Council*. Ministry of Internal Affairs and Communications.

MILLER, M. J. and LIN, S. (1981). The Analysis of Some Selective-Repeat ARQ Schemes with Finite Receiver Buffer. *IEEE Transactions on Communications*, **29**(9), 1307–15.

MIND (2005). *IST Project MIND (IST-2000-28584)* [online]. Available from: http://www.ist-mind.org [2005 October].

MITOLA, J., III (1995). The Software Radio Architecture. *IEEE Communications Magazine*, **33**(5), 26–38.

MITOLA, J., III (2000). *Cognitive Radio: An Integrated Agent Architecture for Software Defined Radio*. Thesis (PhD), Dept. of Teleinformatics, Royal Institute of Technology (KTH), Stockholm, Sweden.

MITOLA, J., III and MAGUIRE, G. Q., JR. (1999). Cognitive Radio: Making Software Radios More Personal. *IEEE Personal Communications Magazine*, **6**(4), 13–18.

MOHR, W. (2005). The WINNER (Wireless World Initiative New Radio) Project – Development of a Radio Interface for Systems beyond 3G. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, September 11–14.

MOHR, W., LÜDER, R. and MÖHRMANN, K.-H. (2002). Data Rate Estimates, Range Calculations and Spectrum Demand for New Elements of Systems Beyond IMT-2000. In: *5th International Symposium Wireless Personal Multimedia Communication*, Honolulu, HI, October.

MONTAVONT, N. and NOËL, T. (2002). Handover Management for Mobile Nodes in IPv6 Networks. *IEEE Communications Magazine*, **30**(8), 38–43.

MONTAVONT, N. and NOËL, T. (2003). Analysis and Evaluation of Mobile IPv6 Handovers over Wireless LAN. *ACM Journal on Mobile Networks and Applications (MONET)*, **8**(6).

NEEL, J. O., BUEHRER, R. M., REED, B. H. and GILLES, R. P. (2002a). Game Theoretic Analysis of a Network of Cognitive Radios. In: *45th Midwest Symposium on Circuits and Systems, MWSCAS-2002*, Tulsa, OK, August 4–7, 409–12.

NEEL, J., REED, J. H. and GILLES, R. P. (2002b). The Role of Game Theory in the Analysis of Software Radio Networks. In: *SDR Forum Technical Conference and Product Exhibition, SDR'02*, San Diego, November.

NEEL, J. O., REED, J. H. and GILLES, R. P. (2004). Convergence of Cognitive Radio Networks. In: *IEEE Wireless Communications and Networking Conference, WCNC 2004*, New Orleans, LA, March 21–5, 2250–5.

NIE, N. and COMANICIU, C. (2005). Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks. In: *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN*, Baltimore, MD, November 8–11.

NOAM, E. (1997). Beyond Spectrum Auctions: Taking the Next Step to Open Spectrum Access. *Telecommunications Policy*, **21**(5), 461–75.

NOAM, E. (1998). Spectrum Auctions: Yesterday's Heresy, Today's Orthodoxy, Tomorrow's Anachronism. Taking the Next Step to Open Spectrum Access. *Journal of Law and Economics*, **44**, 765–90.

OKUMURA, Y., OHIMORI, E., KAWANO, T. and FUKUDA, K. (1968). Field Strength and its Variability in VHF and UHF Land Mobile Radio Service. *Rev. Elec. Communication Lab.*, **16**.

ORFANOS, G., WALKE, B., JANSEN, J. and REUMERMANN, H. (2005). Interference Aware Adaptive MAC Protocol for MC-CDMA Ad Hoc Wireless LANs. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, September 11–14.

ORTHMAN, F. (2005). *WiMAX Handbook, Building 802.16 Wireless Networks*. New York, NY: McGraw-Hill.

OSBORNE, M. J. and RUBINSTEIN, A. (1994). *A Course in Game Theory*. Cambridge, MA: MIT Press.

PABST, R., ESSELING, N. and WALKE, B. (2005). Fixed Relays for Next Generation Wireless Systems – System Concept and Performance Evaluation. *Journal of Communications and Networks, Special Issue on "Towards the Next Generation Mobile Communications"*, **7**(2), 104–14.

PABST, R., WALKE, B. H., SCHULTZ, D. C., HERHOLD, P., YANIKOMEROGLU, H., MUKHERJEE, S., VISWANATHAN, H., LOTT, M., ZIRWAS, W., DOHLER, M., AGHVAMI, H., FALCONER, D. D. and FETTWEIS, G. P. (2004). Relay-Based Deployment Concepts for Wireless and Mobile Broadband Radio. *IEEE Communications Magazine*, **42**(9), 80–9.

PAHLAVAN, K., KRISHNAMURTHY, P., HATAMI, A., YLIANTTILA, M., MÄKELÄ, J., PICHNA, R. and VALLSTRÖM, J. (2000). Handoff in Hybrid Mobile Data Networks. *IEEE Personal Communications Magazine*, **7**(2), 34–47.

PARSONS, J. D. (1992). *The Mobile Radio Propagation Channel*. Chichester, uk: John Wiley & Sons, Ltd.

PEHA, J. M. (1998). Spectrum Management Policy Options. *IEEE Communication Surveys*.

PEHA, J. M. (2000). Wireless Communications and Coexistence for Smart Environments. *IEEE Personal Communications*, 66–8.

PEHA, J. M. (2005). Approaches to Spectrum Sharing. *IEEE Communications Magazine*, 10–12.

PEREIRASAMY, M. K., LUO, J., DILLINGER, M. and HARTMANN, C. (2004). An Approach for Inter-Operator Spectrum Sharing for 3G Systems and Beyond. In: *15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004*, Barcelona, Spain, September 5–8.

PEREIRASAMY, M. K., LUO, J., DILLINGER, M. and HARTMANN, C. (2005). Dynamic Inter-Operator Spectrum Sharing for UMTS FDD with Displaced Cellular Networks. In: *IEEE Wireless Communication and Networking Conference, WCNC 2005*, New Orleans, LA, March 13–17.

PERKINS, C. E. (1998). *Mobil IP: Design Principles and Practices*. Prentice-Hall.

PERKINS, C. E. (2000). *Ad Hoc Networks*. Addison-Wesley.

PERKINS, C. E. (2002). *IP Mobility Support for IPv4*. IETF Network Working Group, RFC 3344.

PERKINS, C. E., BELDING-ROYER, E. and DAS, S. (2003). *Ad Hoc On-Demand Distance Vector (AODV) Routing*. RFC 3561. IETF Network Working Group.

PERKINS, C. E. and BHAGWAT, P. (1994). Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. In: *ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications*, 234–44.

PETRAS, D. (1999). *Entwicklung und Leistungsbewertung einer ATM Funkschnittstelle*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Aachen, Germany.

PFLETSCHINGER, S., HOOLI, K., HUSCHKE, J., MOHR, W. and OJANEN, P. (2005). Possible Impact of Emissions from UWB Devices on Systems beyond 3G. In: *14th Meeting of the Wireless World Research Forum, WWRF14*, San Diego, CA, July 7–8.

POLK, J., SCHNIZLEIN, J. and LINSNER, M. (2004). *Dynamic Host Configuration Protocol Option for Coordinate-based Location Configuration Information*. RFC 3825. IETF. Available from: http://www.rfc-archive.org/getrfc.php?rfc=3825 [March 2006].

RAMJEE, R., PORTA, T. L., THUEL, S., VARADHAN, K. and WANG, S.-Y. (2002). HAWAII: A Domain-based Approach for Supporting Mobility in Wide Area Wireless Networks. *IEEE/ACM Transactions on Networking*, **10**(3), 396–410.

RAYCHAUDHURI, D. and JING, X. (2003). A Spectrum Etiquette Protocol for Efficient Coordination of Radio Devices in Unlicensed Bands. In: *14th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2003*, Beijing, China, September 7–10.

ROBERTS, L. (1975). ALOHA Packet System With and Without Slots and Capture. *Computer Communications Review*, April.

ROBINSON, J. W. and Randhawa, T. S. (2004). Saturation Throughput Analysis of IEEE 802.11e Enhanced Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications*, **22**(5), 917–28.

RODRIGUEZ, V., MOESSNER, K. and TAFAZOLLI, R. (2005). Auction Driven Dynamic Spectrum Allocation: Optimal Bidding, Pricing and Service Priorities for Multi-Rate, Multi-Class CDMA. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, September 11–14.

ROSENBERG, J., H, S., CAMARILLO, G., JOHNSTON, A., PETERSON, J., SPARKS, R., HANDLEY, M. and SCHOOLER, E. (2002). *RFC 3261 – SIP: Session Initiation Protocol*.

ROUSE, T., Band, I. and MCLAUGHLIN, S. (2002). Capacity and Power Investigation of the ODMA Networks in TDD-CDMA Based Systems. In: *IEEE International Conference on Communications, ICC 2002*, New York, NY, April 28–May 5, 3202–6.

SALBU (1978). *Adaptive Communication System*. SA Patent. Salbu Research & Development Ltd.

SAMARASINGHE, R., FRIDERIKOS, V. and AGHVAMI, A. H. (2003). Analysis of Intersystem Handover: UMTS FDD and WLAN. In: *London Communications Symposium*, September 8–9.

SCHINNENBURG, M., DEBUS, F., OTYAKMAZ, A. and BERLEMANN, L. (2005). A Framework for Reconfigurable Functions of a Multi-Mode Protocol. In: *Software Defined Radio Technical Conference, SDR'05*, Orange County, CA, November 14–18.

SCHREIBER, F. (1988). Effective Control of Simulation Runs by a New Evaluation Algorithm for Correlated Random Sequences. *AEÜ International Journal of Electronics and Communications*, **42**(6), 347–54.

SCHULTZ, D. C., WALKE, B., PABST, R. and IRNICH, T. (2003). Fixed and Planned Relay Based Radio Network Deployment Concepts. In: *10th Meeting of the Wireless World Research Forum, WWRF10*, New York, NY, October 27–8.

SDRFORUM (2005). Available from: http://www.sdrforum.org [June 2005].

SECAN-LAB (2006). *Ad-Hoc Routing Protocols* [online]. Available from: http://wiki.uni.lu/secan-lab/Ad-Hoc+Protocols+($28)Characteristics($29).html [March 2006].

SIEBERT, M., LOTT, M., SCHINNENBURG, M. and GÖBBELS, S. (2004). Hybrid Information System. In: *IEEE Semiannual Vehicular Technology Conference, VTC2004-Spring*, Milan, Italy, May 17–19.

SMITH, M., WELTY, C. and MCGUINNESS, D. L. (2003). *OWL Web Ontology Language Guide, W3C Candidate Recommendation* [online]. World Wide Web Consortium. Available from: http://www.w3.org/TR/2003/CR-owl-guide-20030818/ [June 2005].

SOBRINHO, J. L. and KRISHNAKUMAR, A. S. (1999). Quality-of-Service in Ad Hoc Carrier Sense Multiple Access Wireless Networks. *IEEE Journal on Selected Areas in Communications*, **17**(8), 1353–68.

SRINIVASAN, V., NUGGEHALLI, P., CHIASSERINI, C. F. and RAO, R. R. (2003). Cooperation in Wireless Ad Hoc Networks. In: *22nd Conference of the IEEE Communications Society, INFOCOM 2003*, San Francisco, CA, March 30–April 3.

STALLINGS, W. (1991). *Data and Computer Communications*. New York: Macmillan.

STEMM, M. and KATZ, R. H. (1999). Vertical Handoffs in Wireless Overlay Networks. *Mobile Networks and Applications*, **3**(4).

STEPPLER, M. (1998). Performance Analysis of Communication Systems Formally Specified in SDL. In: *First International Workshop on Simulation and Performance '98, WOSP'98*, Santa Fe,NM, October 12–16.

TAFAZOLLI, R. (2006). *Technologies for the Wireless Future: Wireless World Research Forum (WWRF), Volume 2*. Chichester, UK: John Wiley & Sons, Ltd.

TAKAGI, H. and KLEINROCK, L. (1985). Throughput-Delay Characteristics of Some Slotted-ALOHA Multihop Packet Radio Networks. *IEEE Transactions on Communications*, **33**(11), 1200–7.

TANENBAUM, A. S. (1988). *Computer Networks*. Prentice-Hall.

TANTRA, J. W., FOH, C. H. and MNAOUER, A. B. (2005). Throughput and Delay Analysis of the IEEE 802.11e EDCA Saturation. In: *IEEE International Conference on Communications, ICC 2005*, Seoul, Korea, May 16–20.

TIRLA, O., MUFIOZ, A. B., SCHOO, P. and TESSIER, S. (2002). Accounting Management in Heterogeneous Mobile Access Networks: The Mind Approach. In: *IEEE Personal Indoor and Mobile Radio Communications, PIMRC'02*, Lisbon, Portugal.

TSAI, T.-C. and WU, M.-J. (2005). An Analytical Model for IEEE 802.11e EDCA. In: *IEEE International Conference on Communications, ICC 2005*, Seoul, Korea, May 16–20.

TUTTLEBEE, W. (2002a). *Software Defined Radio: Enabling Technologies*. Chichester, UK: John Wiley & Sons, Ltd.

TUTTLEBEE, W. (2002b). *Software Defined Radios: Origins, Drivers and International Perspectives*. Chichester, UK: John Wiley & Sons, Ltd.

USB (2000). *Universal Bus Specification, Revision 2.0, Compaq, Hewlett-Packard, Intel, Lucent, Microsoft, NEC, Philips* [online]. Available from: http://www.intel.com/technology/usb/spec.htm [June 2005].

VALKO, A. (1999). Cellular IP: A New Approach to Internet Host Mobility. *ACM SIGCOMM Computer Communication Review*, **29**(1), 50–65.

VALLETTI, T. M. (2001). Spectrum Trading. *Telecommunications Policy*. Available from: http://www.ms.ic.ac.uk/tommaso/spectrumtrading.pdf.

VAN NEE, R. and PRASAD, R. (2000). *OFDM for Wireless Multimedia Communications*. Artech House.

VDE/ITG (2005). *Mobility in IP-based Networks*. VDE/ITG Section 5.2.4. Available from: http://www.ikr.uni-stuttgart.de/Content/itg/fg524/ [October 2005].

VORNEFELD, U., WALKE, C. and WALKE, B. (1999). SDMA Techniques for Wireless ATM. *IEEE Communications Magazine*, 52–7.

WALFISH, J. and BERTONI, H. L. (1998). A Theoretical Model of UHF Propagation in Urban Environments. *IEEE Transactions on Antennas and Propagation*, **36**(12), 1788–96.

WALKE, B. (2000). HiperLAN2 – ein weitgehend in Deutschland entwickeltes Breitbandnetz für den Internetzugang. In: *Symposium of BMBF "Mobilkommunikation von morgen"*, Neu-Ulm, Germany, June 19.

WALKE, B. H. (2001). *Mobile Radio Networks*. New York: John Wiley & Sons, Inc.

WALKE, B. H. (2002). *Mobile Radio Networks*. New York: John Wiley & Sons, Inc.

WALKE, B. H. and BRIECHLE, R. (1985). A Local Cellular Radio Network for Digital Voice and Data Transmission at 60 GHz. In: *Cellular and Mobile Communications International*, London, UK, November, 215–25.

WALKE, B. H., ESSELING, N., HABETHA, J., HETTICH, A., KADELKA, A., MANGOLD, S., PEETZ, J. and VORNEFELD, U. (2001). IP over Wireless Mobile ATM – Guaranteed Wireless QoS by HiperLAN/2. *Proceedings of the IEEE*, **89**, 21–40.

WALKE, B. H., WIJAYA, H. and SCHULTZ, D. C. (2006). Layer-2 Relays in Cellular Mobile Radio Networks. In: *IEEE 63rd Vehicular Technology Conference, VTC2006-Spring*, Melbourne, Australia, May 7–10.

WEINBERGER, D., GILL, J., HENDRICKS, D. and REED, D. P. (2005). *An Open Spectrum FAQ* [online]. Available from: http://www.greaterdemocracy.org/OpenSpectrumFAQ.html [July 2005].

WEINSTEIN, S. B. and EBERT, P. M. (1971). Data Transmission by Frequency Division Multiplexing Using the Discrete Fourier Transform. *IEEE Transactions on Communications*, **19**, 628–34.

WEISS, E., FREWEL, M. and XU, B. (2004). Improving Ad Hoc Routing for Future Wireless Multihop Networks. In: *European Wireless 2004, EW'04*, Barcelona, Spain, February 24–27.

WEISS, E., HIERTZ, G., HISCHKE, S., XU., B., WALKE, B. and GROSS, S. (2006). Improving Routing Performance in Wireless Ad Hoc Networks Using Cross-Layer Interactions. *Elsevier Ad Hoc Journal*.

WEISS, E., HIERTZ, G. R. and XU, B. (2005). Performance Analysis of Temporally Ordered Routing Algorithm Based on IEEE 802.11a. In: *61st IEEE Vehicular Technology Conference 2005, VTC Spring 2005*, Stockholm, Sweden.

WEISS, T. A. and JONDRAL, F. K. (2004). Spectrum Pooling: An Innovative Strategy for the Enhancement of Spectrum Efficiency. *IEEE Communications Magazine*, **42**(3), S8–S14.

WERBACH, K. (2002). *Open Spectrum: The New Wireless Paradigm* [online]. Available from: http://werbach.com/docs/new_wireless_paradigm.htm [July 2005].

WIFI-ALLIANCE (2005). Available from: http://www.wi-fi.org [December 2005].

WIJAYA, H. (2004). Interoperability Concept for Wireless LANs. In: *IEEE 60th Vehicular Technology Conference, 2004, VTC2004-Fall*, Los Angeles, CA, September 26–9, 919–23.

WIJAYA, H. (2005). *Broadband Multi-Hop Communication in Homogeneous and Heterogeneous Wireless Lan Networks*. Thesis (PhD), ComNets, RWTH Aachen University, Aachen, Germany: Wissenschaftsverlag Mainz.

WIJAYA, H. and ZIRWAS, W. (2003). High Speed Broadband Radio Access Multi-Hop Communication. In: *IEEE ICON 2003*, Sydney, Australia, September 28–October 01, 31–6.

WIKIPEDIA (2005). Available from: http://www.wikipedia.org [June 2005].

WIMAX-FORUM (2005). Available from: http://www.wimaxforum.org [December 2005].

WINNER (2005). Available from: http://www.ist-winner.org [June 2005].

WWRF (2005). Available from: http://www.wireless-world-research.org [June 2005].

XIAO, Y. (2004). Performance Analysis of IEEE 802.11e EDCF under Saturation Condition. In: *IEEE International Conference on Communications, ICC 2004*, Paris, France, June.

XU, B. (2002). *Self-organizing Wireless Broadband Multihop Networks with QoS Guarantee*. Thesis (PhD), Chair of Communication Networks, RWTH Aachen University, Germany: Wissenschaftsverlag Mainz.

XU, B. and WALKE, B. (2002). A New Air Interface Concept for Wireless Multimedia Communications beyond the 3rd Generation. *Wireless Personal Communications*, **23**(1), 121–35.

YANIKOMEROGLU, H. (2002). Fixed and Mobile Relaying Technologies for Cellular Networks. In: *2nd Workshop on Applications and Services – Wireless Networks, ASWN'02*, Paris, France, July 3–5, 75–81.

ZADEH, A. N. and JABBARI, B. (2001). On the Capacity Modeling of Multi-Hop Packet CDMA Networks. In: *IEEE Military Communications Conference, Milcom 2001*, Washington, DC, October 28–31, 601–4.

ZANELLA, A. and DE PELLEGRINI, F. (2005). Statistical Characterization of the Service Time in Saturated IEEE 802.11 Networks. *IEEE Communication Letters*, **9**(3), 225–7.

ZANG, Y., HIERTZ, G. R., HABETHA, J., OTAL, B., SIRIN, H. and REUMERMAN, H.-J. (2005). Towards High Speed Wireless Personal Area Network – Efficiency Analysis of MBOA MAC. In: *International Workshop on Wireless Ad-Hoc Networks 2005, IWWAN 2005*, London, UK, May 23–26.

ZHAO, Q., TONG, L. and SWAMI, A. (2005). Decentralized Cognitive MAC for Dynamic Spectrum Access. In: *1st IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, DySPAN2005*, Baltimore, MD, November 8–11.

ZHAO, R. and WALKE, B. (2004). Traffic Performance of the Wireless Channel-Oriented Ad-Hoc Multi-hop Broadband System. In: *IEEE 60th Vehicular Technology Conference, VTC2004-Fall*, Los Angeles, CA, September 26–29.

ZHAO, R., WALKE, B. H. and HIERTZ, G. R. (2006). An Efficient IEEE 802.11 ESS Mesh Network Supporting Quality of Service. In: *IEEE Journal Selected Areas in Communications (JSAC), Special Issue on Multi-Hop Wireless Mesh Networks*, 24(11), November 2006.

ZHAO, R., WALKE, B., EINHAUS, M. and ORFANOS, G. (2005). A Primary Adaptation of the W-CHAMB Protocol for Gigabit WPANs. In: *16th IEEE Conference on Personal, Indoor and Mobile Radio Communications, PIMRC 2005*, Berlin, Germany, 11–14 September.

ZHENGHUA, F., HAIYUN, L., ZERFOS, P., SONGWU, L., LIXIA, Z. and GERLA, M. (2005). The Impact of Multihop Wireless Channel on TCP Performance. *IEEE Transactions on Mobile Computing*, **4**(2), 209–21.ABRAMSON, N. (1970). The ALOHA System – Another Alternative for Computer Communications. In: *Fall Joint Computer Conference, AFIPS Press*, 295–8.

ZIMMERMANN, E., HERHOLD, P. and FETTWEIS, G. (2003). On the Performance of Cooperative Diversity Protocols in Practical Wireless Systems. In: *IEEE Vehicular Technology Conference Fall, VTC-Fall 2003*, Orlando, October 6–9.

ZIRWAS, W., GIEBEL, T., ESSELING, N., SCHULZ, E. and EICHINGER, J. (2002). Broadband Multi Hop Networks with Reduced Protocol Overhead. In: *European Wireless Conference 2002*, Florence, Italy, February 25–8, 333–9.

Index